RATERS' ACCENT-FAMILIARITY LEVELS AND THEIR EFFECTS ON PRONUNCIATION SCORES AND INTELLIGIBILITY ON HIGH-STAKES ENGLISH TESTS

Thesis submitted for the degree of Doctor of Philosophy at the University of Leicester

by

Kevin Cogswell Browne
Department of Education
University of Leicester

2016

Abstract

Raters' accent-familiarity levels and their effects on pronunciation scores and intelligibility on high-stakes English tests

Kevin C. Browne

Some current high-stakes tests of English have abandoned native-speaker models of pronunciation for scoring purposes, and instead rely largely on raters' estimations of 'listener effort' needed to cope with test-takers' speech in order to determine pronunciation scores. Recent studies within the field of language testing have revealed significant score variance occurring on such tests due to raters' differing familiarities with test-takers' accents. The studies that investigated raters' accent-familiarity differences as a threat to reliability and validity of scores on high-stakes tests have only determined significant score differences can occur, but have offered little more than speculation concerning why accent-familiarity impacts raters' score decisions. The purpose of this thesis was to investigate not only the veracity of the threat, but also attempt to provide an explanation why raters' accent-familiarity differences affect scores. A strong rationale exists supporting a hypothesis that exposure to the speech of a particular group of speakers, or accent, positively affects listeners' speech processing abilities of utterances in that accent by increasing intelligibility. In order to determine the veracity of the hypothesis two studies were conducted: a pilot study examined the pronunciation scores and intelligibility differences between raters with different levels of accent-familiarity with Japanese-English, and a larger study investigated pronunciation score and intelligibility differences with Arabic-English, Spanish-English and Dhivehi-English. Many-Facets Rasch Measurements of the data revealed significant differences in both pronunciation scores and intelligibility occurred between accent-familiarity rater groups with all accents. The findings also showed significant correlations between level of accent-familiarity and score leniency, as well as accent-familiarity level and increased intelligibility, though the measures and effect sizes were not equal with each accent. Raters' accent-familiarity differences were confirmed as a valid threat to pronunciation scores.

Acknowledgements

I would like to first thank ETS for awarding me a TOEFL Small Grant for Doctoral Research in Second or Foreign Language Assessment. It both helped me with the costs of this research and gave me confidence to see it through. I am indebted to them.

I would of course, like to thank Professor Glenn Fulcher who supervised my doctoral studies. His unwavering support and faith in me throughout every step and stage of this project is greatly appreciated. I would also like to thank my assistant-supervisor, Professor Pamela Rogerson-Revell, for sharing her expertise in pronunciation with me.

I am deeply indebted to Azlifa Ahmed for her amazing generosity of time and patience. I never could have included Dhivehi-English in this study without her. Thank you for being such a great friend and colleague. Other colleagues that assisted me and deserving of recognition here are Lesley Underwood, Shane Dick, Ahmed Nazif and Nathaniel Owen. I would also like to give special thanks to Professor Mike Linacre for his generosity with his time to answer all of my many questions about Rasch and FACETS. Dr. Linacre's online Rasch Forum is a treasure trove of wisdom and information for all things Rasch. Thank you so much.

I would like to thank Jo Singh, Douglas McCall and the crew at the F Bar in Leicester – thank you for taking me in, calling me friend and giving me memories to last a lifetime. I never would have felt like a student without you guys.

I am most appreciative to family. To my parents and brother and sister, thank you for believing in me, and your never ending enthusiasm toward my life that took me so far away from home so long ago. To my children, Hunter and Noah, I tried my best to balance my life of work, study and family, and my greatest hope is that you never realized I was so busy or how much it hurt when I thought about the time I was missing with you. Finally, thank you to my wife Yuko. You are the reason I am who I am, and I never could have accomplished this without you. I could never imagine a more supportive and caring partner to go through life with. This dissertation is dedicated to you.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACTFL/ETS	American Council on the Teaching of Foreign Languages and the Educational Testing Service
APU	Asia Pacific University (Ritsumeikan)
BKB-R	Bamford-Kowal-Bench revised
CL	Common Language effect size
EIL	English as an International Language
EFL	English as a foreign language
ELF	English as a Lingua Franca

ELPAC	English Language Proficiency for Aeronautical Communication
ESL	English as a second language
ETS	Educational Testing Service
IELTS	International English Language Testing System
IPA	International Phonetic Alphabet
L1	First language
L2	Second language
LFC	Lingua Franca Core
MFRM	Many-facets Rasch Measurement
NNS	Nonnative speaker
NS	Native speaker
OPI	Oral proficiency interview
RP	Received Pronunciation
TEFL	Teaching English as a Foreign Language
TESL	Teaching English as a Second Language
TOEFL iBT	Test of English as a Foreign Language Internet-based Test
WE	World Englishes

Chapter 1 : Introduction

Chapter 1 introduces the framework of this dissertation placing the research in the context of current trends and theories in the field of language testing. It begins by introducing some of the difficulties associated with high-stakes English pronunciation testing. This is followed by a discussion of how raters perform a contributory role in pronunciation scoring that is often mistakenly considered as 'passive' in high-stakes tests of spoken English. The term *high-stakes test* refers to any norm-referenced test that is designed and employed for purposes where the scores determine qualification for enrollment in higher education, study abroad programs, employment or other factors of high-level importance to the stakeholders. It will be revealed that raters' linguistic experiences, or 'accent-familiarities', should be included as a facet of any high-stakes test's pronunciation construct. Accents and accent-familiarity are then defined and discussed. From this background the problem statement is proposed that raters' differing accent familiarities pose a threat to the validity of pronunciation scores on high-stakes tests of spoken English. The aim of this research is then presented to address the problem statement followed by the research questions and hypotheses. A discussion of how the hypotheses were tested is then presented before clarifying what the scope of the study included. Though traditionally included in the conclusion or discussions of most research papers, the chapter continues with the introduction and discussion of two potential limitations of the study. They are included in this chapter to in order to better clarify the intent and design of this research. Some unique aspects of the research instrument are then introduced, followed by a brief explanation of the original contributions to the field of language testing this research provides. The chapter concludes first with some important terms defined, followed by outlines of the chapters included in this dissertation.

1.1 The challenges to high-stakes testing of English pronunciation

Assessing performance aspects of language is a difficult endeavor, and assessing speaking is the most difficult to do reliably (Alderson & Bachman, 2004, p. ix). Pronunciation is possibly the most salient aspect of spoken language, and accents are an easily recognized feature of pronunciation (Scovel, 1988) both by native speakers and non-native speakers alike (Derwing, Rossiter & Munro, 2002). Test developers aiming to produce reliable English speaking tests have long faced the difficulty of attempting to choose or define a single, appropriate model of English pronunciation that can be agreed upon. The difficulties are largely due to the problematic nature surrounding any attempts to describe, standardize or even determine what the parameters of 'correct' English pronunciation should be. If pronunciation is to be assessed, logic suggests it would be most beneficial if there could be a single, agreed upon model of performance to be employed. While a single model of English pronunciation is a logical ideal in theory, any attempts to define or agree on such a model have proven problematic and controversial.

Traditionally, many high-stakes, norm referenced tests of English have applied the model of "the educated native speaker" as the ideal of English pronunciation (Fulcher, 2003, p. 93). The educated native speaker model influenced benchmarks and score descriptors used to assess pronunciation. However, researchers have long discussed the difficulties related to defining and determining precisely what the model of the 'educated native speaker of English' is. Savignon (1997) argued that the educated native speaker is problematic as the model of communicative competence because for the second language learner, the competence is tested, but for the native speaker it is assumed. This suggests a kind of infallibility and universal intelligibility of speech based on first language (L1) status. The model allows for pronunciation variation from English L1 speakers, but expects a kind of homogeneity of pronunciation from English L2 speakers reflective of an ambiguous and undefined set of standards. There is, in fact, no precise agreement of what a 'native speaker' is (Bachman & Savignon, 1986; Lantolf & Frawley, 1985; Jarvis, 1986; Jenkins, 2000; Barnwell, 1987; Davies, 1990, 2003; Fulcher, 2003). The educated native speaker model, it seems, is too controversial to prove practical or appropriate as the model for high-stakes English pronunciation testing. Other factors besides the lack of a definitive description of native speaker (NS) pronunciation have emerged to further complicate attempts at reliable, valid tests of spoken English.

English has become a global language. The spread of English use around the world brings additional complications to performance assessment. Native speakers of English no longer dominate or even participate in the majority of English exchanges (Kachru, 1997). Kachru's (1985) three circles of English (see Figure 1) is perhaps the best illustration of the spread of English, and how the number of English nonnative speakers (NNSs) greatly outnumbers NSs. From this situation, a lot of research and attention has shifted toward a non-native-centric approach to criteria (among other areas of language interest). Fields of research like World Englishes (e.g. Smith, 1987; Kachru, 1997), English as an International Language (e.g. Canagarajah, 2006) and English as a Lingua Franca (e.g. Dauer, 2005; Jenkins, 2006) have emerged that propose different ideas and theories concerning the use and direction of the language. These new fields reject and challenge the use of the educated native speaker of English to serve as the model of ideal proficiency (including for pronunciation). They take a strong philosophical stance that native-centric approaches to both pedagogy and testing are culturally biased, and lack validity in testing contexts. Lowenberg (2002) argues that native speaker standards employed in tests outside of Kachru's 'inner circle' do not function appropriately for measuring ability in the language in the contexts the language is actually being used. He argues against the necessity of insisting on native norms when communication rarely if ever includes native speakers. Smith (1983) further argues that nonnative varieties of English are no less intelligible than those of native speakers. Nonnative patterns of pronunciation have begun to be considered as equally valid as native norms. Harding (2011) questioned the exclusion of nonnative speakers of English in the recordings used in listening tests of English for academic purposes. His findings suggest it is feasible to include second language (L2) accented speech in high-stakes listening tests without threatening the validity of such tests.

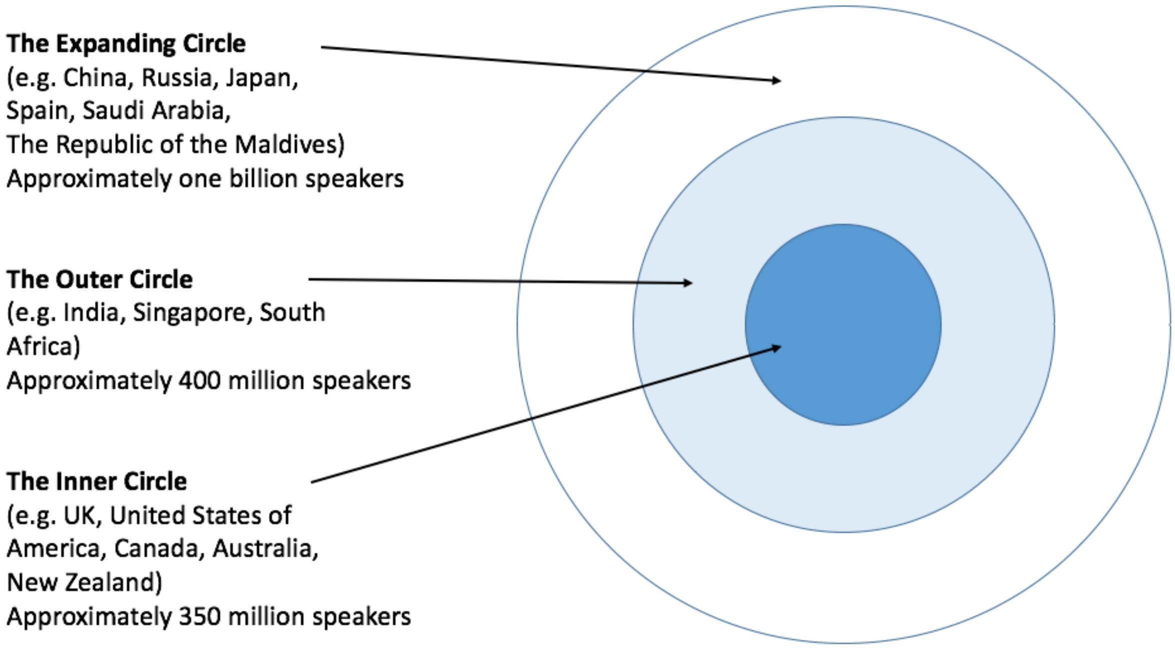


Figure 1.1: Illustration of Kachru's (1985) concentric circles of English

The spread of English and the emergence of the nonnative speaker as the dominating force of English usage have changed the way performance aspects of English speaking are assessed. Pressure and/or influence from fields like World Englishes and English as a Lingua Franca (Jenkins 2002), it can be argued, have resulted in the reduction of the use of descriptive terms such as 'native speaker' or 'native like' on tests like the International English Language Testing System (IELTS). Some high-stakes tests completely avoid using references to native speakers entirely like the Test of English as a Foreign Language Internet-based test (TOEFL iBT). Test developers that have adopted this notion that nonnative patterns (at least for pronunciation) can be no less intelligible than native patterns face difficult challenges. As Davies (2003) astutely stated, "defining language proficiency is just as elusive as defining the native speaker" (Davies, 2003, p. 173).

While tests like the TOEFL iBT and IELTS have largely abandoned the native speaker model of performance, another problem with reliable pronunciation scoring has emerged. Rather than constructing pronunciation proficiency descriptors that clearly define what the tests' standards of pronunciation performance are, current

TOEFL iBT and IELTS scoring rubrics for pronunciation instead rely heavily on raters' individual estimations of 'listener effort' to determine pronunciation scores. The greater the listener effort the rater experiences, the lower the score the rater is instructed to deliver. Intelligibility also contributes to raters score-making decisions (see Tables 1.1-1.3). What is problematic, and central to this study, is that this approach includes the presumption that either inherently or as a result of rater training, all raters experience the same measure of listener effort processing the same speaker or groups of speakers. Likewise, it presumes that intelligibility is speaker-dependent only. The rater is not included as a contributing factor to intelligibility.

Table 1.1: TOEFL iBT Independent Speaking Rubrics for Delivery

Score	Descriptors
4	Generally well-paced flow (fluid expression). Speech is clear. It may include minor lapses, or minor difficulties with pronunciation or intonation patters, which do not affect overall intelligibility.
3	Speech is generally clear, with some fluidity of expression, though minor difficulties with pronunciation, intonation, or pacing are noticeable and may require listener effort at times (though overall intelligibility is not significantly affected).
2	Speech is basically intelligible, though listener effort is needed because of unclear articulation, awkward intonation, or choppy rhythm/pace; meaning may be obscured in places.
1	Consistent pronunciation, stress, and intonation difficulties cause considerable listener effort; delivery is choppy, fragmented, or telegraphic; frequent pauses and hesitations.
0	Speaker makes no attempt to respond OR response is unrelated to the topic.

Note. Descriptive note. Adapted from *TOEFL iBT Tips How to prepare for the TOEFL iBT.* ETS TOEFL. 2008, p. 44.

Table 1.2: TOEFL iBT Integrated Speaking Rubrics for Delivery

Score	Descriptors
4	Speech is generally clear, fluid, and sustained. It may include minor lapses or minor difficulties with pronunciation or intonation. Pace may vary at times as the speaker attempts to recall information. Overall intelligibility remains high.
3	Speech is generally clear, with some fluidity of expression, but it exhibits minor difficulties with pronunciation, intonation, or pacing and may require some listener effort at times. Overall intelligibility remains good, however.
2	Speech is clear at times, although it exhibits problems with pronunciation, intonation, or pacing and so may requires significant listener effort. Speech may not be sustained at a consistent level throughout. Problems with intelligibility may obscure meaning in places (but not throughout).
1	Consistent pronunciation and intonation problems cause considerable listener effort and frequently obscure meaning. Delivery is choppy, fragmented, or telegraphic. Speech contains frequent pauses and hesitations.
0	Speaker makes no attempt to respond OR response is unrelated to the topic.

Note. Descriptive note. Adapted from *TOEFL iBT Tips How to prepare for the TOEFL iBT.* ETS TOEFL. 2008, p. 45.

Table 1.3: IELTS Pronunciation score bands and descriptors (public version)

Band (score)	Pronunciation Descriptors
9	• uses a full range of pronunciation features with precision and subtlety • sustains flexible use of features throughout • is effortless to understand
8	• uses a wide range of pronunciation features • sustains flexible use of features, with only occasional lapses • is easy to understand throughout; L1 accent has minimal effect on intelligibility
7	• shows all the positive features of Band 6 and some, but not all, of the positive features of Band 8
6	• uses a range of pronunciation features with mixed control • shows some effective use of features but this is not sustained • can generally be understood throughout, though mispronunciation of individual words or sounds reduces clarity at times

5	• produces basic sentence forms with reasonable accuracy • uses a limited range of more complex structures, but these usually contain errors and may cause some comprehension problems • shows all the positive features of Band 4 and some, but not all, of the positive features of Band 6
4	• uses a limited range of pronunciation features • attempts to control features but lapses are frequent • mispronunciations are frequent and cause some difficulty for the listener
3	• shows all the positive features of Band 3 and some, but not all, of the positive features of Band 4
2	• Speech is often unintelligible(sic)
0-1	(no descriptions given)

Note. Descriptive note. Adapted from *IELTS Speaking: Band Descriptors (public version).* IELTS. < http://www.ielts.org/pdf/SpeakingBanddescriptors.pdf>

A common characteristic of the score descriptors in Tables 1~3 are estimations of the effort raters experience coping with test candidates' pronunciation. Expressions like, "effortless to understand"; "easy to understand"; "some difficulty for the listener"; "may require some listener effort"; "requires significant listener effort"; "cause considerable listener effort," suggest that pronunciation scoring includes raters determining their personal measures of difficulty coping with test candidates' phonetic production, not to mention raters' ability to accurately distinguish their listener effort as requiring 'significant' or 'considerable' effort. These skill-level descriptors serve more to describe listener reactions than actual test candidate production. This method of relying on the raters' speech processing difficulty and ability to find test-takers' speech intelligible to determine scores builds individual rater variations into the scores. It is precisely this point of pronunciation scoring that is of greatest interest to this research, and theorized to be where the greatest construct-irrelevant threats to tests utilizing such measurements lie.

One argument from the field of World Englishes may explain why current test developers have written pronunciation score descriptors the way they have. It is possible that test developers have incorporated Smith's (1983) argument that nonnative varieties of English are no less intelligible than those of native speakers, and therefore raters' differences in familiarity with nonnative speech is inconsequential. While nonnative speech in general may be no less intelligible than native speech it does not suggest that all listeners, whether native or nonnative, familiar or unfamiliar with a particular accent, will find that particular accent equally intelligible. This study will demonstrate that intelligibility is a collaboration of speaker and listener, and that raters' familiarity with the accent of the speaker positively affects intelligibility. What will be shown is that perhaps what Smith meant was that nonnative accents can be as intelligible as any other accent if the listener becomes familiar with that accent. Empirical research is needed to determine how homogeneous raters' actually are when measuring test-takers for pronunciation. Empirical research is also needed to determine whether or not all raters experience similar measures of 'listener effort' processing all speakers' speech, and if the intelligibility of test-takers' speech is not rater-dependent. This study endeavors to provide this evidence.

1.2 The role of the rater within pronunciation constructs

Raters perform a participatory role in any evaluation of a speaker's pronunciation. Browne and Fulcher (2016) discussed the "janus-faced" nature of the construct of fluency that entails both the performance of the speaker and the perception and interpretation of the listener. Pronunciation constructs, like fluency constructs also include the same dynamic of speaker performance with rater perception and interpretation to determine scores. It will be revealed that current high-stakes tests, like the TOEFL iBT, do not adequately consider the raters' accent familiarities as a facet of the construct of pronunciation. Pronunciation, and intelligibility as a component of the pronunciation construct, cannot, or should not, merely be perceived as one-sided mechanical processes that are solely speaker dependent. Pronunciation and intelligibility will be presented in this study as components of this duality. Together pronunciation and intelligibility share a codependence for determining success of second language (L2) speech processing. These include both the speaker's production of linguistic features of utterances that can be referred to as 'pronunciation', and the listener's perception and interpretation of the utterances that determines 'intelligibility'. Smith and Nelson (1985) stated, "intelligibility is not speaker- or listener-centered but is interactional between the speaker and listener" (p.333). While the mechanics of producing utterances associated with pronunciation may be speaker dependent the intelligibility of the resulting utterances are largely listener dependent.

Intelligibility in this study is defined according to Field's (2005) interpretation that intelligibility is determined through the successful transcription of a speaker's utterances, and is limited to only features of the speech signal. This definition excludes all notions of locutionary (pertaining to the meaning of an utterance) or illocutionary force (pertaining to the intended meaning of an utterance by the speaker). Intelligibility, when applying this interpretation, is where a speaker's pronunciation and the listener's familiarity with their accent interact to determine speech-processing success. It is a participatory role that raters perform that determines outcomes, or scores, for pronunciation, and that role should be included as a facet of such constructs by test developers.

1.3 Conceptualizing accents and accent-familiarity

Exactly what an accent or variety of English is, how they are derived and how they are perceived by native speakers and non-native speakers are not easily agreed upon by researchers. When attempting to define accents, Zuengler (1988) describes them as "phonetic variants" that socially mark speakers as members of particular social or psychological states. Wells (1982) distinguishes accents as patterns of pronunciation used by native speakers of shared communities or social groups to which they belong. Patterns of spoken English seen as 'typical' by speakers for whom English is not their native language are labeled as 'foreign accents', and that they possibly reflect phonetic and phonological traits of their respective first languages (L1s). Riney, Takagi and Inutsuka (2005) define accents as, "the perceived degree of native or foreign accent in someone's speech, a characteristic style of pronunciation determined by (or at least associated with) the speaker's regional, social, or linguistic background" (p.442). They provide additional clarity and insight into what the characteristics of accents are with their five facts we can assume about accents:

> 1. Everyone has an accent. 2. All accents are linguistically equal. 3. Any two accents can in certain circumstances be socially unequal. 4. Every accent is speaker-listener dependent. 5. The intelligibility (recognition) of words spoken with any accent is also speaker-listener dependent.
>
> (Riney, Takagi & Inutsuka, 2005, p.442)

These facts are important to this research for a few reasons. First, Riney et al.'s facts about accents recognize that all speakers, both native and nonnative, have accents. It is not necessary to distinguish native accents from nonnative accents; all speech is accented. That point is important because it eliminates the distance often presumed between test-taker and rater that might view the test-taker as an 'accented speaker' (flawed) and the rater as 'non-accented' (infallible). Secondly, they make it clear that both the recognition of an accent and its intelligibility are speaker-listener dependent; both perception and processing of speech are collaborations between speaker and listener. The saliency of accents is so intense, Derwing and Munro (2009) claim that accents can be detected by listeners, even when listening to voice recordings played backward (p.477). Major (2007) asserts that listeners can even distinguish native speakers from nonnative speakers when listening to samples of spoken language in languages not known, previously heard or studied by the listeners. While the listener can detect accents in almost any spoken exchange it is not clear how accent-familiarity affects intelligibility in all circumstances. In order to describe the context of this study further, it is necessary to first define and discuss what accent-familiarity is and how it poses a threat to high-stakes speaking tests.

1.4 Raters accent familiarities as a construct-irrelevant threat to pronunciation scores

As previously stated, this research supports a theory that raters' differing measures of familiarity with the L2 accented speech of different groups of test-takers cause construct irrelevant threats to the reliability and validity of pronunciation scores on high-stakes tests of spoken English. Browne and Fulcher (forthcoming) define accent-familiarity as, “a speech perception benefit developed through exposure and linguistic experience” (p. 4). Accent-familiarity and its potential effects on test scores have been a recent point of interest in language testing (e.g. Carey, Mannell & Dunn, 2011; Winke, Gass & Myford, 2012, 2013; Xi & Mollaun, 2009, 2011; Zhang & Elder, 2011). Research concerning raters’ accent-familiarity has focused on several means of acquisition such as raters’ shared L1 with test-takers (Xi & Mollaun, 2009; Kim, 2009), raters’ prior L2 study of test-taker’s L1 (Winke et al., 2013), raters living in the country or region where test-takers’ L1 is spoken (Carey et al., 2011).

Accent-familiarity has been found to cause a speech perception benefit (Carey et al., 2011; Browne and Fulcher, forthcoming). The benefits of interlanguage familiarity were first thought to be limited to only L2 speakers (see Bent & Bradlow, 2003). It was thought that this kind of familiarity only affected speech perception between speakers with a shared L1. Benefits were extended also in some cases between L2 speakers of different L1 backgrounds that Bent and Bradlow referred to as a “mismatched interlanguage speech intelligibility benefit” (p.1606). Such interlanguage and mismatched interlanguage speech familiarity benefits were thought to not affect native speakers. The notion that interlanguage speech processing benefits could occur among native English listeners was first introduced by Carey et al. (2011), and coined the phrase, “interlanguage phonology familiarity” (p.204). Browne and Fulcher (forthcoming) also confirmed that native speakers can experience not only an interlanguage phonology familiarity benefit, but also revealed an intelligibility benefit, at least for raters familiar with Japanese-English. Additionally, Browne and Fulcher determined that the familiarity speech intelligibility benefit is incremental based on the amount and types of linguistic exposure the listener has experienced.

This study theorizes that raters who are familiar with the accented speech of certain test-takers and groups of test-takers experience a speech intelligibility benefit when scoring those speakers. This theory is grounded and based on empirical evidence (Browne & Fulcher, forthcoming). If the implications of this theory are found to be valid it would mean that raters’ accent-familiarity differences are causing a construct-irrelevant threat to pronunciation scores. Gass and Varonis (1984) determined that listeners familiar with a particular nonnative accent were more successful completing transcription tasks than listeners unfamiliar with that accent. Familiar accents are more intelligible, but language testers have not given Gass and Varonis' findings adequate consideration as a potential threat to pronunciation scores.

This study is the first of its kind to theorize raters experience a speech intelligibility benefit because of their personal familiarities with the accents of groups of speakers. Previous studies have determined raters' familiarities with nonnative English accents in high-stakes tests of spoken English affect scores (Winke et al., 2012, 2013; Carey et al., 2011; Xi 2007; Xi & Mollaun 2009, 2011; Zhang & Elder, 2011); however, none have considered or measured intelligibility differences due to accent-familiarity. In all of these earlier studies (with the exception of Browne & Fulcher, forthcoming), the method of scoring spoken performance based on raters' listener effort was largely ignored and not discussed. Carey (et al., 2011) theorized a speech perception benefit but did not make any direct mention that the use of listener effort in scoring aspects of the IELTS oral proficiency interview could be at the heart of the score differences. It may have been implied, but this clarity is needed if test developers are to attempt to address the problem.

The previous research others have done investigating whether or not raters' accent familiarities affect scores have been greatly beneficial to attracting awareness of the threat to tests, but there is much that has not been determined yet. With the exception of Carey et al. (2011), all of the previous studies others have conducted on this topic have focused on how raters' accent-familiarity differences affect holistic scores for speaking. The scope of the potential, or possible impact, raters' differing accent familiarities may cause was made too broad in these studies, and included too many aspects of speech performance. This I feel was a major problem. There is no logical reason or empirical evidence I am aware of that suggests familiarity with a particular accent contributes to how raters determine the grammatical accuracy of speakers, their ability to successfully complete a test-task, or determine the appropriateness of vocabulary to context, or any of the other aspects related to scoring speech performance beyond measuring the pronunciation or intelligibility of test candidates' speech. The holistic scores they examined included too many aspects of speech performance unrelated to what can be considered 'threatened' by raters' differing familiarities with test-takers' L2 accents. Only Carey et al. (2011) closely examined pronunciation scores, and determined significant differences in scores occurred, yet they provided no concrete evidence suggesting why raters' accent-familiarity differences causes inter-rater score variation. None of the previous studies concerned with raters' accent-familiarities affecting scores accomplished more than determining that accent-familiarity, as a rater trait, is a potential threat to scores. This study does. This research attempts to both confirm that raters' accent-familiarity differences cause construct irrelevant threats to scores, and attempts to determine why this occurs.

The argument is presented that current high-stakes tests' pronunciation constructs do not adequately consider the rater or their differing accent familiarities as contributing factors to test candidates' scores. Current test designers are implying that all raters are either equal in their abilities or can be trained to process the phonological content of test-takers' speech equally. This study contends that this is not possible, or at least not likely, and that test designers must include raters' accent-familiarity and level of familiarity with the test-taker's L1 accent as a facet of the pronunciation construct.

1.5 The problem statement, aim and implications of this study

The previous sections have identified the difficulties pertaining to reliably testing English pronunciation in high-stakes tests of spoken English. From this background the problem this research aims to address emerges. In this section the problem statement, aim of the research and hypotheses are presented.

1.5.1 The problem statement

Raters' differing familiarities with the L2 accented speech produced by test candidates from specific L1 backgrounds cause a rater-dependent, construct-irrelevant threat to the reliability and validity of pronunciation scores on current high-stakes tests of spoken English.

1.5.2 The aim of this research

This research aims to determine what effect raters' differing levels of familiarity with Japanese-English, Spanish-English, Arabic-English and Divehi-English have on the pronunciation scores awarded on high-stakes, semi-direct tests of spoken English. This study theorizes that raters' differing levels of familiarity with the L2 accented speech of test candidates' specific L1 backgrounds affects the measures of listener effort raters experience processing test candidates' speech. Accent-familiarity, this study argues, not only causes what Carey et al. (2011) called an 'interlanguage phonology benefit', but that it causes what I call a *secondary interlanguage speech intelligibility benefit.* This benefit is not dissimilar to Bent and Bradlow's (2003) 'matched' and 'mismatched' interlanguage speech intelligibility benefits. This theory, however, suggests native speakers are also capable of developing the speech perception benefits of increased intelligibility when listening to familiar L2 accents. Intelligibility and pronunciation are further theorized to share a kind of symbiotic relationship between test candidate and rater when determining pronunciation scores. Pronunciation is comprised of the speaker's production of sounds, and intelligibility is determined by the rater's ability to decode the speaker's sounds into words. Together, pronunciation and intelligibility inform and influence the pronunciation score awarded by the rater. It is proposed that even within the framework of semi-direct speaking tests, pronunciation must include the rater as an active contributor to the communicative act. Additionally, any pronunciation construct must include the rater's level of familiarity with the test candidate's L2 accent and L1 background as a facet of the construct. If, in fact, this hypothesis is found to be accurate it would add weight to the view that listening and rating are not passive activities.

This is a strong interactionist theoretical stance concerning the nature of test constructs. From this standpoint, predictions can be made and tested empirically. If it can be demonstrated that raters' accent-familiarity influences or affects the construct, two main findings will be revealed. First, raters' level of accent-familiarity positively affects intelligibility success rates. Second, with greater intelligibility comes reduced 'listener effort' and higher scores for pronunciation delivered to speakers whose accents are more familiar to the rater. These findings would confirm that a speech perception benefit exists due to accent-familiarity, and serve to confirm the hypotheses of this study.

The link between raters' accent familiarities as a speech perception benefit that both affects intelligibility success rates and causes pronunciation score differences has not previously been considered by other researchers (excluding Browne & Fulcher, in press). The earlier studies concerned with raters' accent-familiarity as a threat to test scores (Winke et al., 2012, 2013; Xi & Mollaun, 2009, 2011; Zhang & Elder, 2011) only established that raters' accent familiarities can cause score differences. With the exception of Browne and Fulcher (forthcoming), which describes the pilot study from this the research described in this thesis, there have been no other studies that have attempted to also determine why score differences occur.

1.5.3 The potential implications and importance of this research

Some researchers may argue that since only some of the previous studies investigating this potential threat to test reliability and validity found significant score differences that it is not necessary that this problem be corrected. The same researchers may argue that the implications of not addressing this possible threat are minimal. However, the implications of this research have the potential to affect numerous stakeholders. Tests like the TOEFL iBT and IELTS are designed to be valid, unbiased and with objective scoring (ETS, 2008, p.4). McNamara and Roever (2006), using gender as an example suggest, "if scores can be shown to systematically vary by gender when the variable being measured is required to be insensitive to gender, then construct-irrelevant variance has been detected and has to be eliminated by redesigning the aspects of the test" (p. 18). If the tests in question have not considered raters' familiarity or lack of familiarity to be an element of the pronunciation or intelligibility constructs, but significant variance is detected as a result of such familiarity then the issue(s) should be addressed. Furthermore, if accent-familiarity is determined to cause score variance it is possible that some widely known and more familiar accents, like Spanish-English for example, may possess an unfair advantage to other less-well-known accents (e.g. Dhivehi-English) because the likelihood of having a rater familiar with their accent assigned to score parts of their test is higher. These high-stakes tests like the IELTS and TOEFL iBT are used primarily for entrance requirements for higher education in several countries around the world. Clearly, the implications for test-takers can be no higher; if any construct-irrelevant threats to scores are determined they must be confirmed, and if confirmed addressed and eradicated. The implications of this research likewise could affect the institutions of higher education that employ these tests, as it is their goal to determine the best candidates for their degree programs.

1.6 The research questions and hypotheses

1.6.1 Main research questions

- Does accent-familiarity as a rater characteristic cause significant differences in pronunciation scores?
- Does accent-familiarity as a rater characteristic cause significant differences in intelligibility success-rates?

1.6.2 Sub-questions

In order to address the aim, the following sub-questions are considered:

- Do the raters behave like independent experts?
- Do the raters have the same leniency/severity?
- Does accent-familiarity cause a positive bias in pronunciation scores?
- Do raters 'very familiar' with a particular accent tend to show the greatest leniency

scoring pronunciation of that accent?

- Do raters 'very familiar' with a particular accent tend to show the greatest intelligibility success coping with speech in that accent?
- Is there a correlation between levels of accent-familiarity, pronunciation scores and intelligibility success-rates?
- Is the rater accent-familiarity effect more prevalent with some accents than others?
- Do the population sizes of test candidates' L1 affect the likelihood of raters having familiarity with their accents?

1.6.3 The hypotheses

1. A rater's familiarity with the L2 accented speech produced by speakers from a specific L1 background positively affects the rater's 'listener effort' expended processing utterances from those speakers.
2. Raters familiar with the L2 accented speech of speakers from a specific L1 background will tend to score those speakers higher than raters less familiar or with no familiarity with the accented speech from the same speakers.
3. A rater's familiarity with the L2 accented speech produced by speakers from a specific L1 background positively affects the intelligibility of utterances produced by those speakers.
4. Raters familiar with the L2 accented speech of speakers from a specific L1 background will find utterances produced by those speakers' more intelligible than raters less familiar or unfamiliar with the accented speech from the same speakers.

1.7 Testing the hypotheses

While the problem of raters' accent familiarities as a threat to test scores is not entirely new, there is still much that is unknown. The previous work other researchers have conducted concerning the threat of raters' accent familiarities have addressed it from various different theoretical positions (Carey et al., 2011; Winke et al., 2012, 2013; Xi & Mollaun, 2009). The scope of this research is more focused than any of the previous studies, and attempts to isolate and focus on where raters' accent-familiarity poses the greatest threat to score variance – pronunciation scores. This section of this chapter provides details pertaining to the research instrument created to test the hypotheses, and explains what the scope and limitations of the design are.

1.7.1 Focusing on pronunciation scores

The scope of this research is limited to determining how raters' accent-familiarity differences affect pronunciation scores and intelligibility as a component of the pronunciation construct of high-stakes tests of spoken English. As stated previously, earlier related studies have been valuable toward developing a broad perspective of the potential threat raters' accent familiarities pose to test scores, but they have failed to determine precisely what aspect of test candidates' performance is most influenced by raters' accent-familiarity. These studies focused only on the final, holistic scores for speaking. The studies that focused on actual scores from the TOEFL iBT (Winke et al., 2012, 2013; Xi & Mollaun, 2009), for instance, examined the full range of oral proficiency assessment. It is simply not practical to suggest that accent-familiarity affects raters' measuring test-takers' task-completion ability, fluidity of expressions produced, the effective use of grammar and vocabulary, cohesion of expressions, topic development or relationships between ideas. Carey et al.'s (2011) study focused on the final scores from the IELTS oral proficiency interviews that included raters considering not only the pronunciation of test candidates' speech, but also their fluency and coherence, which included how often repetition or self-correction was needed, the candidates' "flexibility and precise use" of vocabulary in all contexts, sustained use and accuracy of idiomatic language and candidates' grammar accuracy "apart from 'slips' characteristic of native speaker speech". Fortunately, though, Carey et al. examined all of the different scores raters awarded on the IELTS interview test, and they were able to identify pronunciation scores to be most affected. They even suspected intelligibility to be a factor of the interlanguage phonology benefit they determined accent-familiarity caused, and suggested further research be conducted to examine the extent to which accent-familiarity contributes to intelligibility.

This research suggests that raters' accent familiarities affect a more limited range of speech perception aspects that can affect scores. It considers pronunciation scores to be the most likely to be affected in large part due to differences in intelligibility of speech in familiar accents. This study does not question the potential that accent-familiarity, as a rater characteristic, can affect holistic scores - the findings from earlier studies confirm, at the least, it does impact holistic scores, but the differences are not always significant. This study is different in this respect. Pronunciation, in isolation, has not been the primary focus of any previous research investigating the affects raters' accent familiarities have on outcomes.

1.7.2 Scoring pronunciation in this study

This study examined how different raters scored the pronunciation of speakers of Japanese-English, Spanish-English, Arabic-English and Dhivehi-English. This section introduces details pertaining to how pronunciation was scored in this study. Rater participants scored speaker participants' pronunciation according to the pronunciation scoring rubrics designed specifically for this study (see table 4.25 in Chapter 4). Scoring was limited in its scope to only include raters' judgments of segmental, subsegmental and suprasegmental elements of production. Rater participants were instructed that scoring should not include any judgments or consideration of elements related to locutionary or illocutionary force of speakers' utterances. To be clear, it was not necessary for the rater participants to understand the context or 'meanings' of the speakers' utterances, or what the speakers may have intended the utterances to mean; all scores were to only pertain to the raters' estimations of the speakers' pronunciation according to the provided pronunciation score band descriptors. The pronunciation scoring rubrics were designed based on the TOEFL iBT score band descriptors for 'Delivery'. This methodological choice included raters making scoring decisions based on their personal measures of difficulty, or 'listener effort', they experienced coping with speaker participants' pronunciation. This choice of methodology was made because it duplicates the prime aspect of pronunciation scoring employed in some current high-stakes tests that most likely cause score variance. This type of scoring that relies on raters' listener effort to determine outcomes would help to establish if the hypotheses that raters familiar with a particular accent experience less listener effort coping with that accent than raters unfamiliar with the same accent. The next section introduces how raters' accent familiarities were ascertained, and includes details pertaining to the content of the speaker participants' utterances.

1.7.3 Isolating accent-familiarity from other forms of familiarity that contribute to speech perception

Gass and Varonis (1984) determined there are four types of familiarity that contribute to comprehension: familiarity of topic (or context), familiarity with a particular speaker, familiarity with a particular nonnative accent and familiarity with nonnative speech in general. One of the hypotheses proposed in this research is that familiarity with a particular accent positively affects the intelligibility of utterances produced by speakers of that accent. In order to test this hypothesis, it was deemed necessary to attempt to isolate accent-familiarity from other the other types of familiarity known to affect speech processing. If accent-familiarity could be isolated from the other types of familiarity, a greater insight into its impact on scores and intelligibility could be gained. This is a unique aspect of the focus and methodology of this research. No other study has attempted to measure accent-familiarity while also eliminating the other types of familiarity.

Gass and Varonis found that among the four types of familiarity they investigated, familiarity of context had the greatest effect on intelligibility (see also Kennedy & Trofimovich, 2008). While it is true that they claimed their research was investigating comprehension and not intelligibility, it can be argued (and is argued in Chapter 2) that their findings had more to do with intelligibility than any notion of comprehension since their findings were based entirely on transcription task accuracy. Nevertheless, since part of this study aimed to measure how accent-familiarity affects the intelligibility of test-takers' speech by raters it was deemed necessary to attempt to isolate accent-familiarity from context-familiarity. In short, if context-familiarity is eliminated from affecting intelligibility, it is possible to presume that all intelligibility success in the study would be due to raters successfully deciphering the phonetic content of the speakers' pronunciation. To accomplish this design aim, the speaker participants read prepared sentences constructed specifically to have unpredictable contexts. This is explained in detail in Chapter 4.

1.7.4 Testing raters not speakers

It is important at this point to clarify that this study and its research instrument were designed specifically to determine if and how raters' differing accent familiarities affect pronunciation scores. The research instrument was designed as a test for raters - not speakers. It should be considered only as a research tool for addressing the particular research questions posed herein. It was not designed to be, and is not recommended, as an ideal test for measuring L2 learners' pronunciation or for testing other more holistic aspects of speech proficiency. As previously mentioned, this research attempted to reduce or eliminate context familiarity from interfering with scoring pronunciation. This is a unique methodological approach to addressing the research questions, and it is not recommended that high-stakes tests eliminate familiar topics or predictable contexts from their tasks or tests.

The reason this point is raised here in the introduction is that it has been argued that language testing research should be designed to duplicate actual test rating experiences (Winke et al., 2013). This study did include aspects of actual test rating experiences, like employing the use of pronunciation score descriptors based on the TOEFL iBT. However, it is wrong to suggest that language-testing research be limited to only duplicating actual tests. After all, language testing research is not language testing. Research must be flexible and dynamic if it aims to uncover and address any and all testing related questions, problems and concerns. This study examines a specific rater-dependent threat to test reliability and validity. If the full impact of raters' differing accent familiarities on scores is to be revealed it requires more than only replicating the actual tests. This is especially true for a test like the TOEFL iBT that employs a holistic rating scale that relies on subscale 'Delivery' measures that are arguably most affected by accent-familiarity (discussed in Chapter 4). As such, limiting research to only observing raters' performance in a test whose validity and reliability is in question will not provide the greatest insight into both the potentiality of the threat and why the threat occurs. If research is limited in such a way, the advancement of language testing will be adversely affected. It is for this reason that the design of the research instrument described in this thesis should not be considered a limitation for its lacking in complete duplication of actual test experiences, but should be considered an innovation in research design. The next section addresses another potential limitation to this study – the absence of rater training.

1.7.5 The problematic nature of rater training in validity studies

Rater training is often considered vital to the validity of any assessment (Bachman & Palmer, 1996; Luoma, 2004; McNamara, 1996, 2000); however, there is a question of whether it is vital to validity before the validity of a test is established. Flucher (2003, p. 145-146) argues that training should not take place before validation if any of the evidence presented is in the form of rater agreement. The validity of a rating scale is codependent with the ability of the rater to adequately apply its use in the task of scaling (Gescheider, 1976); thus, rater training in how to interpret the chosen scales is often vital (Fulcher, 2003, p. 143). What is not clear is if the positive effects of raters' familiarities with different accents can be normalized. The argument that the subtle differences between terms like 'significant' and 'considerable' listener effort in the IELTS and TOEFL iBT scoring rubrics may be operationalised in rater training has its difficulties. Training generally includes providing rater-trainees with several examples of each level to 'normalize' the rater population. Any attempt to force agreement amongst raters has been criticized as attempting to make 'clones' of raters (Alderson, Clapham & Wall, 1995, p.108).

Attempting to pre-determine what kind of pronunciation causes all raters 'significant' or 'considerable' effort is precisely contrary to what this research theorizes; namely, that accent-familiarity affects listener effort.

This research will provide evidence that raters' differing accent familiarities affect how speakers in familiar and unfamiliar accents are perceived. This evidence raises the concern that test scores are, at least partially rater-dependent. Bachman and Palmer (1984) astutely observed that, "…many educators and researchers have lost sight of the fact that communication involves two parties, and success in communicative performance will always be dependent upon the abilities of two people" (p. 42). Raters are participating (possibly equally) with test-takers in the success or failure of test-related communicative acts. It is questionable if the results of rater training can eliminate raters' accent familiarities as a threat to the validity and reliability of high-stakes speaking tests. Fulcher (2010) suggests that, "training judges to agree and then presenting the level of agreement as a part of a validity argument is problematic" (p.244). It is for this reason this study did not include any formal rater training as part of its methodology, and does not consider it to be regarded as a negative aspect or limitation of the study.

1.7.5 Contributing concepts from outside the language testing literature

Discussed in this thesis will also be theories from the social sciences and cognitive sciences that offer some possible explanations why raters behave differently due to their varying accent familiarities. These include the Exemplar theory (Johnson, 2005) and Perceptual Magnet theory (Kuhl, 1991; Iverson & Kuhl, 1995) that Carey et.al. (2011) briefly introduced. These two theories provide explanations of how familiarity with accents and speakers occur, and describe how accent-familiarity functions at a cognitive level. The Exemplar theory claims that listeners amass memories of sound structures they have experienced, referred to as 'exemplars.' These exemplars are activated and aid listeners when engaged in the task of speech perception. The Perceptual Magnet (Kuhl, 1991; Iverson & Kuhl, 1995) also provides insight suggesting that our experiences with different speakers, accents and phonologically similar sounds (like exemplars) that are different from speaker to speaker, accent to accent, can be perceptually assimilated like a magnet, attracting to other sounds in that category resulting in comprehension. Examining these theories and others serves to establish how accent familiarities affect the way speech is processed. By reviewing and understanding how these theories clarify the mechanics of speech processing, a better understanding can be achieved of how and why raters with differing levels of accent-familiarity interpret rating scales differently when scoring the same speaker or groups of speakers.

1.8 Original contribution to the literature

There are gaps in the current literature concerning the effect raters' accent familiarities have on pronunciation scores. The existing literature demonstrates only that raters' accent familiarities can pose a threat to the reliability and validity of holistic scores on high-stakes speaking tests. No other study has examined how raters' accent-familiarity differences function to threaten test reliability and validity, or why score differences occur. This study presents new insight into this problem in language testing. Firstly, this study attempts to determine not only if familiarity as a rater characteristic significantly affects pronunciation scores, but also examines how familiarity affects intelligibility. Previous studies examining whether or not raters' accent familiarities affected scores only determined if the effect was significant or not, and provided little or no explanation or evidence to suggest why score differences occur. By examining intelligibility success rate differences as well as the pronunciation scores raters deliver we might be able to reveal a definitive reason accent-familiarity causes score differences. Namely, familiar accents are more intelligible. It is suspected that pronunciation scores are where the greatest inter-rater reliability concerns exist. Accent-familiarity causes what I call a *secondary interlanguage speech intelligibility benefit*, but this benefit is not likely to include improved or enhanced perception abilities beyond word/utterance recognition. Another way this study is different from previous studies is that it considers accent-familiarity in terms of four levels rather than an 'either/or' distinction. Carey et al. (2011) did consider levels of familiarity during data collection, but determined during analyses that collapsing their data to only 'familiar' or 'unfamiliar' was all that was necessary. This study is also the first to consider accent-familiarity acquisition from multiple sources of exposure. Previous studies were more one-dimensional focusing only on raters acquiring accent-familiarity from a single means of acquisition such as: shared L1 (Xi & Mollaun, 2009); native and nonnative raters (Kim, 2009; Xi & Mollaun, 2009) L2 study (Winke et al., 2012) and familiarity of the local accent where raters live (Carey et al., 2011). Additionally, no previous study has questioned or challenged the wording of pronunciation score rating scale descriptors currently in use in semi-direct, high-stakes tests of spoken English. Tests like the TOEFL iBT and IELTS that rely on raters to assign pronunciation scores (or 'delivery' scores) based on their personal measures of 'listener effort' processing test-takers' speech, it will be argued, cause scores to be rater-dependent. Finally, this study is the first to specifically select four L2 accents that are all different in their language family and population sizes. This study investigated Spanish-English, Arabic-English and Divehi-English accents in the main study, and Japanese-English was researched in the pilot study that was featured in Browne and Fulcher (forthcoming)(discussed in full in Chapter 3). Divehi-English is spoken only by people in or from the small atoll nation, the Republic of the Maldives, and whose population experiences little to no media exposure. Both Arabic-English and Spanish-English are accents that are spoken by people from numerous countries and whose cultures and countries are widely featured in both entertainment and the media. If it is determined that raters accent-familiarity differences cause score variances, it is plausible that test candidates from L1 backgrounds like Spanish and Arabic are more likely to have a rater or raters assigned to score their speaking tests that are familiar with their accents than test candidates from the Maldives. This would not only determine that raters' accent familiarities cause a rater-dependent, construct-irrelevant threat to test scores, but that it also causes negative score bias for test candidates whose L1s are spoken by smaller populations. For these reasons, this study has the potential to increase our collective knowledge and understanding of the issue of raters' accent familiarities causing construct-irrelevant threats to the validity and reliability of high-stakes tests of spoken English.

1.9 Important terms defined

Smith and Nelson (1985) recommended that future researchers clarify the use of terms *intelligibility, comprehensibility, comprehension,* and *interpretability* since many researchers have used them interchangeably. Though Chapter 2 includes discussions of how these terms have been interpreted both differently and interchangeably by researchers, it was considered best to briefly clarify how these terms are defined and applied in this study in this chapter in order to hopefully avoid possible confusion. *Accent* is also defined here.

Intelligibility – This study adheres to Field's (2005) interpretation. The term is restricted,

> to features of the speech signal . . . it refers to the extent to which the acoustic-phonetic content of the message is recognizable by a listener. On this analysis, intelligibility forms a part of a wider construct of comprehensibility . . . It also serves to separate perceptual evidence at phoneme, word, and tone-group levels from higher level evidence such as world knowledge, which originates outside the signal.
>
> (Field, 2005, p. 401)

Intelligibility is a term frequently discussed in this thesis. The scope of intelligibility according to Field is limited to the listener, or in the case of this study, the rater's ability to successfully transcribe the lexical content of a speaker's utterance. In this interpretation the intelligibility of a speaker participant's utterance is not decided based on a rater's opinion of the quality of a speaker's pronunciation. Rather, intelligibility is determined only through accurate transcription of speakers' utterances. It is, therefore, possible for a rater to believe he or she has found a speaker's utterance intelligible even if the transcription task results demonstrate intelligibility did not occur.

Comprehensibility – is interpreted in this study according to Derwing and Munro's (2009) application where the term is limited to an overall measure of how difficult or easy it is for a listener to understand a given speaker; it "is a judgment of difficulty and not a measure of how much actually gets understood."(p.478) Their interpretation was similar to that of Field's (2005) that determined comprehensibility "by an overall rating of how easy it is to understand a given speaker" (p. 401). Derwing and Munro's (2009) definition was selected because theirs restricts comprehensibility to only aspects of the speech signal, whereas Field's (2005) included the rater's 'understanding'. In both interpretations comprehensibility is determined similarly to how pronunciation is scored in some high-stakes tests of spoken English, as well as in this study by relying on raters' measures of 'listener effort'. Since the TOEFL iBT and IELTS both largely rely on listener effort for pronunciation scoring this interpretation of the term was well suited to this study.

Comprehension - is interpreted here as the positive or negative result of Derwing and Munro's (2009) listener's comprehensibility measurement. Since this study is not concerned with whether or not the rater participants perceive the locutionary and/or illocutionary force of the speaker participants' utterances, this interpretation was selected. This study is concerned primarily with how raters cope with and score the pronunciation of test-takers, so discussions of comprehension are limited in this thesis.

Interpretability – This study applies Smith and Nelson's (1985) definition. According to their interpretation the term refers to "meaning behind word/utterance (illocutionary force)" (p. 334).

1.10 Outline of the dissertation

This dissertation is organized into six chapters. What follows are explanations of the structure and content of each chapter:

Chapter 1 – Introduction

This chapter placed this study within the context of current theories and debates in the fields of language testing and applied linguistics. The rationale was introduced that raters' differing familiarities with the L2 accented speech produced by speakers from specific L1 backgrounds causes a construct-irrelevant threat to the reliability of high-stakes tests of spoken English. A theory was presented that accent-familiarity causes a speech intelligibility benefit that makes familiar accents more intelligible and easier to process than unfamiliar accents. As a direct result of this benefit, raters tend to score test-takers' speaking in an accent they are familiar with higher than raters not familiar or less familiar with that accent. The aim of the research and hypotheses were provided, as well as details pertaining to how the hypotheses were tested, and an outline of the dissertation was presented.

Chapter 2 – Literature review

Chapter 2 delves deeper into the concerns and concepts introduced in the first chapter through a review of the relevant literature. First discussed are research concerning the different types of high-stakes tests of spoken English with particular attention to pronunciation. The chapter continues by examining the impact of emerging fields of research that question with whom the 'ownership' of English pronunciation is most valid, and explores findings related to accent-familiarity's effect on speech perception that could affect pronunciation scores. Two theories from psychology: The Perceptual Magnet Effect, and The Exemplar Theory are then presented that provide valuable insight and potential reasoning why raters' accent-familiarity differences can affect pronunciation scores. Five recent studies concerned with rating and accent-familiarity are then reviewed. The chapter concludes by clarifying the interpretations of the terms intelligibility, comprehensibility and interpretability applied in this study.

Chapter 3 – The pilot study

The details of the pilot study conducted to test the hypotheses and inform the design of the final, larger study are presented in Chapter 3. This study examined how raters' familiarity with Japanese-English affects pronunciation scores on high-stakes, semi-direct tests of spoken English. The results from many faceted Rasch analyses and other analyses revealed significant differences of pronunciation scores, as well as significant differences in intelligibility between groups of raters divided according to their reported levels of familiarity with Japanese-English. Significant correlations were also determined between raters' level of accent-familiarity with Japanese-English and both higher pronunciation scores and intelligibility. It was discovered that raters 'very familiar' with Japanese-English demonstrated the greatest measures of leniency scoring pronunciation, and also experienced the highest measures of intelligibility coping with the Japanese-English accent.

Chapter 4 – Methodology

This chapter describes the rigorous processes employed to design the test made to address the research questions of this study, and the analyses choices made to examine the resulting data. The test was the first of its kind to attempt to collect both data concerning how rater-participants score speakers of three different accents (Spanish-English, Arabic-English and Dhivehi-English) for pronunciation and also measure intelligibility between speaker and rater. The chapter includes discussions of what many facets Rasch measurement (MFRM) is and how it can be utilized to determine performance aspects of test-candidates, test items and raters – including rater-severity and rater bias.

Chapter 5 – Findings and discussion

Chapter 5 presents the findings from the analyses of the test data, and includes discussions of the findings in order to answer each of the research questions posited in this study. Among the findings included and discussed in the chapter is that significant variance was observed among both pronunciation scores and intelligibility success-rates between rater-groups divided according to raters' levels of accent-familiarity with of the three included accents, though the measures of impact to pronunciation scores and intelligibility were not equal between the three accents. Furthermore, significant positive correlations were determined between higher accent-familiarity levels with all three accents and both higher pronunciation scores and greater intelligibility success. Significant empirical evidence is also presented showing that the 'very familiar' level of accent-familiarity for all three accents included the greatest measures of leniency and bias scoring pronunciation, as well as the highest levels of intelligibility.

Chapter 6 - Conclusion and implications

Chapter 6 begins by first providing brief summaries of the research approach and methodology employed in the study. These are followed by presentations of the key findings from both studies. Conclusions are drawn based on the findings, and implications to test design and management are suggested. Recommendations for future research are also included followed by final concluding remarks.

Chapter 2 : Literature Review

In this chapter, discussions of the relevant literature are presented to provide supporting evidence to the claims made in the introduction. The introductory chapter established that there is evidence and theoretical justification to support the claim that raters' differing accent familiarities represent a construct-irrelevant threat to pronunciation scores. This threat jeopardizes both the reliability and validity of high-stakes tests of spoken English. The introduction also determined just cause to suggest that accent-familiarity increases intelligibility when processing speech in familiar accents, and that intelligibility may be a key factor to understanding why accent-familiarity affects raters' judgments. Since the scope of this study is confined to addressing only pronunciation testing, this review of the literature is likewise mainly focused on this aspect of English testing; however, at times the chapter includes discussions of important theories and findings from outside of pronunciation testing and language testing in general to present a broader perspective and understanding to the claims presented in this research.

The chapter is divided into seven main sections. Section 2.1 discusses the most common designs of current high-stakes tests of spoken English, and the two major types of rating scales employed in such tests. From this point, the focus shifts from discussing speaking tests in general to a more detailed discussion of English pronunciation testing and how the foci and standards have evolved in section 2.2. The section continues with discussions of different theories that challenge native speakers' 'ownership' of the English language, including its pronunciation. This discussion reveals difficult challenges test developers face designing high-stakes tests of spoken English pronunciation. In particular, what is revealed is the problematic nature of determining the appropriateness of accented speech. The chapter continues with section 2.3 that presents how accented speech is addressed in current high-stakes tests of spoken English that include pronunciation. The chapter continues in section 2.4 with a discussion of accent-familiarity and its effects on speech perception. This section includes discussions of Gass and Varonis' (1984) landmark paper, the works of Bent and Bradlow, and evidence suggesting all accents are not equally intelligible to all listeners. From there the chapter continues in section 2.5 with introductions of the Perceptual Magnet Effect and Exemplar theories from outside the language testing literature offering explanations of how familiarity affects speech perception, and why speaker normalization is problematic. Section 2.6 concentrates on the most recent language testing research concerned with the concern that raters' different types of linguistic exposure may be affecting scores on high-stakes tests of spoken English. Finally, the chapter concludes with section 2.7 by answering Smith's (1983) suggestion that researchers define and clarify their interpretations of the terms intelligibility, comprehensibility and interpretability to reduce or eliminate any potential confusion of their use or meaning here in.

2.1 High-stakes tests of spoken English

In this section the two leading types of high-stakes speaking tests, direct and semi-direct tests, are introduced and discussed followed by explanations of the two main types of rating scales employed in such tests: holistic scales and analytic scales. The speaking components of two current high-stakes of English, the Test of English as a foreign language Internet based test (TOEFL iBT) and the International English Language Testing System (IELTS) are discussed as key examples of both types of tests and rating methods. This section provides some necessary background information needed to understand how high-stakes tests of spoken English are designed and scored before focusing more closely on pronunciation testing and scoring.

2.1.1 Direct and semi-direct tests

The two most common types of high-stakes tests of spoken English are *direct* and *semi-direct* tests. Clark (1979) defined these two test types, and distinguished them from *indirect* tests that did not include speaking components. *Direct* speaking tests are, according to Clark, "procedures in which the examinee is asked to engage in a face-to-face communicative exchange with one or more human interlocutors" (p.36). *Semi-direct* tests, Clark explained, were tests that elicited spoken responses from test-takers, “by means of tape recordings, printed test booklets, or other ‘non-human’ elicitation procedures, rather than through face-to-face conversation with a live interlocutor” (p.36)[1].

Perhaps the most common type of direct test of spoken English is the Oral Proficiency Interview (OPI). First designed and implemented in the 1950s by the U.S. Foreign Services Institute, it was designed to be an unstructured and flexible interview that would be conducted by a trained interviewer that would also score the interviewee according to a global scale. Despite criticisms concerning the validity and reliability of the original OPI, the basic test style has continued to be implemented, though it has evolved to be more standardized. Stated plainly, the OPI, “has been widely adopted around the world since the 1970s as the most appropriate method for measuring general speaking proficiency in a second language” (O’Loughlin, 2001, p. 4).

The IELTS, a test primarily used for determining English proficiency for entry into higher education in the UK, is arguably one of the most popular current high-stakes test of English in use that includes an OPI speaking test. The IELTS OPI, however, is much more structured than the original OPI. It is conducted with a live interlocutor, though the interlocutor will not necessarily score the test, and the interviewees are scored based on analytic scales, not a global scale (discussed later in this chapter). The IELTS OPI consists of three sections that are each designed to elicit different types of speech and challenges from the test candidate[2].

On the other hand, semi-direct tests do not include live interlocutors like OPI tests, but instead rely on other means to elicit spoken responses from test-takers. Generally, these tests involve either paper and/or recorded prompts delivered via computer or other means, and test-takers’ responses are recorded and then scored at a later time and/or location. These tests first appeared in the 1970s, and were

---

[1] Fulcher (2003, p.190) used the term “indirect” as a substitute for *semi-direct*, but should not be confused with Clark’s use of indirect tests.

[2] The details pertaining to the IELTS were retrieved from the British Council’s IELTS information webpage on September 21, 2015 from http://takeielts.britishcouncil.org/prepare-test/understand-test-format/speaking-test

designed to increase the reliability of the delivery of an OPI style test (Hughes, 2003). Sometimes called “simulated oral proficiency interviews (SOPI)” (Fulcher, 2003, p.190), these tests are less expensive to administer than face-to-face tests, and can be given in locations where it is not practical, either logistically or financially, to administer direct tests (O’Loughlin, 2001).

The Test of English as a Foreign Language Internet-based test (TOEFL iBT) produced for nonnative English speakers seeking entry into universities in the United States (and other countries) is one of the most widely known high-stakes tests of English that includes a semi-direct speaking test. The TOEFL iBT speaking test includes six tasks (ETS 2008). The first two tasks are *independent tasks* that include the test-takers speaking for 45 seconds each. These include the test-taker expressing their personal preferences concerning a particular topic specified through a written prompt, and the test-taker making and defending an opinion between two opposing choices of action or behavior. The remaining four tasks are considered *integrated tasks.* These tasks involve first reading, listening and then speaking in the first two tasks, and then listening and speaking in the remaining two tasks. These tasks include summarizing what the speaker said within the context of the passage (task 3); expressing cohesion between a textbook passage and a lecture (task 4); expressing problem solving and comprehension skills (task 5); and summarizing and expressing the connectedness of a lecture and textbook passage (task 6). These tasks are scored according to two global, or ‘holistic’ rating scales for independent tasks and integrated tasks.

O’Loughlin (2001) conducted what is perhaps the most thorough investigation into the equivalence of direct and semi-direct tests. The study, his PhD research, investigated the equivalence of semi-direct and direct versions of the *Access:* oral interaction subtest. O’Loughlin investigated various types of data from both tests administered in two separate trials. His findings supported Shohamy’s (1994) conclusions that the two types of tests were not equivalent. At the heart of the lack of equivalence were the interlocutors themselves. O’Loughlin echoes Lazaraton (1996) and McNamara (1997) stating that:

> performance in any direct test of speaking is jointly achieved by the participants (typically, a single candidate and interlocutor) in the interaction, or co-constructed. In semi-direct tests the candidate's performance is also jointly achieved but in this instance with an 'unresponsive' interactional partner.
> (O'Loughlin, 2001, p. 169)

The two test types employ different language skills. It is not a safe assumption to suggest that semi-direct and direct tests measure speaking proficiency equally or alike. Direct speaking tests are arguably more valid because they include two-way communication, but have reliability concerns due to the variation different interlocutors bring to the test. Semi-direct tests are considered to be more reliable due to the control over the delivery of the test, but are arguably an inauthentic and less valid method of testing speaking. Both test types have their advantages and disadvantages, and both types are currently in wide circulation.

There is a problem, however, related to how pronunciation is scored in both direct and semi-direct tests. In the next section, the two types of rating scales employed in high-stakes tests of spoken English are discussed. From this background the problem concerning pronunciation testing that this study aims to address is revealed.

2.1.2 Rating scales for measuring speaking in high-stakes tests

The two most common types of assessment criteria, or rating scales, used to measure speaking proficiency fall into two categories: *analytic* and *holistic* rating scales (Taylor & Galaczi, 2011, p. 177). Raters applying a global, or holistic scale, score speakers' proficiency according to a single scale delivering an "impressionistic assessment" (Davies et al., 1999, p. 75). Tests that employ analytic, or profile, scoring scales measure distinct aspects of test-takers' speaking proficiency, or sub-scales. It will be shown that there are advantages and disadvantages to both approaches of developing rating scales. Like in the previous section, the IELTS and TOEFL iBT are provided as examples of the two types of scales, and a common problem concerning pronunciation scoring in both tests is revealed.

Analytic scales employed in high-stakes speaking tests are used for *multiple-trait scoring* (Alderson, 1981, Hamp-Lyons, 1991). Bachman and Palmer (1996) state, "analytic scales tend to reflect what raters actually do when rating samples of language use" (p. 211). This is a major benefit of analytic scales because scoring is based on multiple observations that increase score reliability (Weir, 1990). However, as Taylor and Galaczi (2011, p. 179) argue, "it assumes raters can reliably distinguish between specific sub-skills or performance features", and suggest that raters may have difficulty distinguishing between criteria especially if constructs overlap. Luoma (2004) suggests that if raters are required to cope with more than five constructs they experience cognitive load problems. Apart from cognitive load difficulties reliability could be adversely affected by a 'halo effect' (Alderson, 1981) where raters tend to score test candidates with the same score in all categories, or that by focusing on multiple aspects of speech that raters will fail to determine the overall success or failure of communication (Hughes, 2003). Employing analytic scales requires a great deal of effort from raters.

Holistic rating scales do not provide specific details concerning every aspect of speech proficiency like analytic scales do. Cooper (1977) provides what Fulcher (2003, p. 89) describes as "the classic definition of holistic assessment":

> Any procedure which stops short of enumerating linguistic, rhetorical, or informational features of a piece of writing. Some holistic procedures may specify a number of particular features and even require that each feature be scored separately, bit the reader is never required to stop and count or tally incidents of the feature.
>
> (Cooper, 1977, p. 4)

These 'features' that are scored separately generally include, "pronunciation or intelligibility, fluency, accuracy and appropriateness" (Taylor & Galaczi, 2011, p.177).

Holistic scales have both advantages and disadvantages, and have been discussed widely in the literature (e.g. Davies et al., 1999; Fulcher, 2003; Luoma, 2004; O'Loughlin, 2001). The primary advantage of holistic scales is their economic advantage especially in large-scale testing situations. Raters can score larger numbers of test candidates' speech more quickly than using multiple skill-specific scales for each speaker for each task. Taylor and Galaczi (2011) discussing the potential merits of holistic scales suggest:

> Holistic descriptors can also offer an intuitively accessible summary of skill levels, an approach that may be especially appropriate in the context of self-assessment (e.g. the English Language Portfolio) or by 'lay' or 'naïve' assessors, i.e. those who may need to make more informal judgments about a second language user's proficiency level but who are not themselves linguistic specialists.
>
> (Taylor & Galaczi, 2011, p. 178)

Holistic scales, however, are not without their critics. Fulcher (2003) argues that holistic scoring is "problematic" because it fails to address the many constructs included in speaking, and instead addresses speaking as a single construct. "A single score may not do justice to the complexity of speaking" (p.90). Taylor and Galaczi (2011) describe holistic scoring as a, "blunt instrument, unable to credit or penalise the relative strengths and weaknesses that invariably characterise a performance" (p. 187). Holistic scales have been criticized for being designed more on intuition than empirical evidence (Weir, 1993), and that raters may focus on different subscales differently to determine scores (Bachman & Palmer, 1996; Davies et al., 1999). As a result, test developers have created holistic scales, like those used for scoring the TOEFL iBT speaking test, that could be considered hybridized scales.

The TOEFL iBT holistic rating scales

The scales used to guide raters in scoring speaking on the TOEFL iBT are holistic, or global scales, but have elements of analytic scales. Table 2.1 shows the rubrics for the 'general description' of the holistic scales for measuring independent speaking. Rating, with this scale is determined primarily according to how a test candidate's speech matches these 'general descriptions', though the scale is argued to be a hybrid because these general descriptions are largely dependent on how well a test candidate's speech matches at least two of three sub-skills' levels. The sub-skills included in the TOEFL iBT scoring rubrics for both independent and integrated speaking are: 'delivery' (that includes pronunciation within its parameters), 'language use' and 'topic development.' The TOEFL iBT speaking test rating scales can be labeled as a hybrid between a single holistic scale and those of analytic scales is because of language included in the 'general description' rubrics. The phrasing in the descriptors, "A response at this level is characterized by at least two of the following" suggests raters determine holistic scores according to a combination of scores from the sub-skill categories.

Table 2.1: The TOEFL iBT Independent speaking rubrics "General Description" category

Score	General Description
4	The response fulfills the demands of the task, with at most, minor lapses in completeness. It is highly intelligible and exhibits sustained, coherent discourse. A response at this level is characterized by all of the following:
3	The response addresses the task appropriately, but may fall short of being fully developed. It is generally intelligibility and coherent, with some fluidity of expression, though it exhibits some noticeable lapses in the expression of ideas. A response at this level is characterized by at least two of the following:
2	The response addresses the task, but development of the topic is limited. It contains intelligible speech, although problems with delivery and/or overall coherence occur; meaning may be obscured in places. A response at this level is characterized by at least two of the following:
1	The response is very limited in content and/or coherence or is only minimally connected to the task, or speech is largely unintelligible. A response at this level is characterized by at least two of the following:

ETS, 2007, p. 44.

The TOEFL iBT hybrid rating scales for scoring speaking attempt to encompass multiple aspects of speech performance into one score. It is arguable though that it is a blunt instrument, and largely reflective of an "armchair" approach to scoring speaking (Fulcher, 1993, p.25) that leaves problems and questions of rater inference (Bachman & Palmer 1996). It is blunt because scores from these scales provide little specific information to the test-taker. The scores reflect an indefinite description of how the rater based their score decisions because it is not possible to know precisely which sub-skills were met by the speaker for that level. Of course, whether or not the rating scales provide useful information to test candidates is not in question in this research, but how raters base their decisions, and in particular – score pronunciation (or delivery) is in question. Since scores are at least in part determined by how raters score test candidates for pronunciation and intelligibility, any potential threats to the validity or reliability of such sub-scores deserves attention.

The subtopics, 'Language use' and 'topic development' are arguably speaker-dependent constructs; however, the phrasing of the 'delivery' descriptors (see Table 1.1 in Chapter 1) indicate a clear rater-dependent aspect to determining proficiency. Language like the ambiguity between a score of 'three' that states, *"may require listener effort at times (though overall intelligibility is not significantly affected)"* and a score of 'two' stating, *"speech is basically intelligible, though listener effort is needed"* imply the scores are dependent on the rater's personal opinion. The language of these 'delivery' descriptors lacks any reference to a model of pronunciation performance, and instead suggests that a kind of homogeneity among raters' abilities to equally process any given speaker's pronunciation exists. Intelligibility, in such a situation is not a shared enterprise between speaker and listener (Smith & Rafiqzad, 1979), but completely speaker-dependent. While the inclusion of sub-scales in a holistic rating scale can provide additional clarity concerning test-taker proficiency, the descriptors of the TOEFL iBT subscale for 'delivery' are problematic.

The IELTS OPI analytic scales

The IELTS OPI employs four analytic scales: fluency and coherence; lexical resource; grammatical range and accuracy; and pronunciation. Using analytic scales rather than a global scale is a departure from the original OPI from the 1950s, and with only four constructs it is arguably within the cognitive limits of Luoma's (2004) cognitive load concerns. The scales provide mostly clear descriptions that can both assist the raters in their tasks and offer useful details to test candidates concerning their level and abilities, but there are problems concerning pronunciation scoring.

The phrasing of the pronunciation descriptors in the IELTS OPI analytic scales indicates scores are, like the TOEFL iBT descriptors for 'Delivery', also rater-dependent (see again Table 1.3). Though the descriptors include illustrative language, there is no indication of, or reference to, a specific model or models of English pronunciation considered 'ideal'. The phrasing includes language like, "*effortless to understand*"; "*accent has minimal effect on intelligibility*"; and "*mispronunciations are frequent and cause some difficulty for the listener*". The use of *understand* as another term for 'intelligibility' is excusable, as it pertains to a pronunciation construct, but expressions like "effortless", "minimal effect" and "some difficulty" suggest or imply there is inter-rater homogeneity concerning how accented speech is processed. This is a problem.

Tests like the TOEFL iBT and IELTS rely on raters' listener effort to determine pronunciation scores. These tests fail to determine or describe a specific model or models of appropriate English pronunciation. Scoring pronunciation without a model implies that either inherently or through rater training all raters can, or do, experience the same measure of difficulty coping with the pronunciation of a given speaker. This implication is the very problem this study is concerned with. In the next section of this chapter how the story of how measuring English pronunciation proficiency has evolved, and explains the reasoning that current high-stakes tests of spoken English lack descriptive models of performance.

2.2 Measuring pronunciation: traditions and trouble

This section of the chapter provides descriptions and discussions of how English pronunciation came to be assessed the way it is in current high-stakes tests. The previous section of this chapter made it clear that whether or not the test is a direct or semi-direct test, or employs holistic or analytic rating scales, pronunciation scores are determined without benefit of a clear model of idealized pronunciation. It also established that pronunciation scores on tests like the TOEFL iBT and IELTS are largely based on the raters' personal measures of 'listener effort' and how intelligible the rater found the test candidate's speech. The purpose of this section of the chapter is to provide the background to this unusual method of measuring performance without the aid of a detailed description of ideal pronunciation. The section starts by introducing the model of the 'educated native speaker of English' that once served to measure pronunciation proficiency. It is revealed that this model is lacking in both validity and clarity. Ultimately, it is determined difficult, if not impossible, to finalize what or who a native speaker is, or what their pronunciation sounds like. This is followed by a discussion of how the spread of English usage and study has raised into question whether or not native speakers of English hold any real proprietary claims over how English should be used (including pronunciation). From this argument opposing native-centric views of English emerge fields of research and pedagogy like World Englishes, English as an International Language and English as a Lingua Franca, which are then discussed. It is from this background of contention that test developers adopted the current method of assessing pronunciation that lacks any clear descriptive model to aid raters in their task.

2.2.1 The educated native speaker model

A problem within the assessment of spoken English, particularly concerning pronunciation testing, is assigning a single model to any test construct of how an English learner should speak. For many tests the measuring tool used for the assessment of speaking English has been the concept of "the educated native speaker" (Fulcher, 2003, p. 93), and the primary aspect of speaking that people assign the label "native speaker" (NS) or "nonnative speaker" (NNS) is pronunciation (Luoma, 2004, p. 10). While perhaps it seems logical to look to native speakers for the best way to model spoken performance, defining the native speaker is just as difficult as defining language proficiency (Davies, 2003, p. 173).

This section illustrates how test developers progressed from a model for speaking assessment that on the surface appeared to be a logical and appropriate attempt to model performance, but resulted in a method that could not be clearly defined and raised concerns not only pertaining to test validity challenges but those of social and cultural bias. Fifty years ago Lado (1961) argued that the native speaker was the criterion of a test's validity, and that test items should be trialed using native speakers of the language. He suggested that test item validity could be determined when, "items eliciting the desired response from native speakers 95% of the time or better should probably be kept" (p.94). Frawley and Lantolf (1985) contradicted Lado's assertion stating, "there is no such thing as a native speaker of a language in absolute terms" (p. 147), and suggested that language performance was both task-dependent and could vary from speaker to speaker without dependency on L1 or L2 status of the speaker. They criticized the use of the native speaker model for testing speaking in the American Council on the Teaching of Foreign Languages and the Educational Testing Service (ACTFL/ETS) (Lantolf & Frawley, 1985). At the heart of their argument was Ballmer's (1981) interpretation that the native speaker does not exist, but that there are four 'types' of native speakers: idiolectal- classified as informants; statistical- representing what Ballmer considered 'typical'; normative, or 'expert speakers'; and former- a type of speaker based on historical records. Tests like the ACTFL/ETS were "not concerned with native speakers; they were concerned with *the* native speaker" (p.343). Lantolf and Frawley discussed the inherent difficulties of attempting to make normative claims concerning native speakers as all speaking in exactly the same manner in every task, every context and with any difficulty level. '*The* educated native speaker' is a normative model of the 'perfect speaker' based on statistical abstractions never thoroughly identified or clarified. Lado, as mentioned earlier did provide some manner of guideline when trialing potential items suggesting that a ninety-five percent agreement among native speakers "should probably be kept", but it became clear that even if agreement were high based on native speaker consensus, the standard was losing ground as the most reliable means of determining test validity.

Researchers have long discussed the difficulties related to defining or determining precisely what the model of a 'Native Speaker' is (Jarvis, 1986; Barnwell, 1987; Fulcher, 2003; Aliakbari, 2001; Davies, 1990, 2003[3]). Bachman and Savignon

[3] Alan Davies' (2003), *The Native Speaker: Myth and Reality* provides the most complete exploration into the many issues and debates related to the subject of the native speaker.

(1986) also criticized the notion of the native speaker as "the criterion of absolute language ability" (p.383) due to native speakers being widely varied in ability. Bachmann (1990) continued to distance his position on the appropriateness of the use of the model in testing stating:

> There are serious problems in determining what kind of language use to consider as the 'native speaker' norm, while the question of what constitutes a native speaker, or whether we can even speak of individuals who are native speakers, is the subject of much debate.
>
> (Bachman, 1990, p. 248)

And with finality Paikeday (1985) even declared the native speaker as "dead"[4].

The educated native speaker model is still embraced however, and recommended by some researchers. Lowenberg (1993, 2002) argued that educated native English speakers determine and follow the standards that can be considered normative for English usage. His arguments include that those standards should determine the design of English proficiency tests. However, defining what the educated native speaker is precisely and what the norms and standards are for English language usage that raters should apply have never been easily or unanimously agreed upon. What may have been a factor for researchers and test developers that support the use of the model to not find cause for definitive descriptors for measuring test candidates' pronunciation (or lexical choices or grammar) was due to attitudes such as those proposed by Ambercrombie (1951) concerning Standard English stating:

> Standard English has nothing to do with the way people pronounce. Standard English is a language, not an accent, and it is as easily recognizable as Standard English when it is written down as when it is spoken . . . There is, in Standard English, a certain amount of regional variation, perhaps, but not very much - it is spoken, and even more written, with remarkable uniformity considering the area which it covers.
>
> (Ambercrombie ,1951, p. 219)

---

[4] Paikeday's (1985) book is titled *The native speaker is dead!*

This attitude is in direct opposition to Riney, Takagi & Inutsuka's (2005) claim that "everyone has an accent" (p. 442). If the Educated Native Speaker has 'no accent' then it is free from definitive definitions of performance standards, and therefore, speakers 'with accents' are speaking incorrectly. Accents will be defined and discussed later in this chapter.

The topic of the educated native speaker model has many facets, and perhaps all can be regarded as contentious. The 'nativeness' of a speaker's linguistic heritage leaves many unanswered questions and concerns when attempting to assess pronunciation proficiency. Firstly, there is no assurance that native speaker status guarantees proficient pronunciation. Widdowson (1994) stated that most native speakers speak nonstandard English and require instruction in the standard form, though he was not solely concerned with pronunciation or accent, it is reasonable that pronunciation variation was included. As stated in chapter one, pronunciation is possibly the most salient aspect of spoken language, and that accents are easily recognized features of pronunciation (Scovel, 1988). Though the educated speaker model for testing purposes is an idealized, normative version of performance and proficiency, the spectrum of pronunciation patterns even when limiting the discussion to only accents of educated speakers from countries and regions where English has official language status, the 'norms' are drastically dissimilar. Ambercrombie's notion of speakers of Standard English having 'no accent' is simply untrue.

The concern amongst many researchers is that for testing purposes if the educated native speaker model is to be employed by raters, and expected to encapsulate English proficiency standards, which and whose norms should be included? (Spolsky, 1993; Clapham, 1996; Criper & Davies, 1988; Davies, 2003) In short, it is not clearly stated which native accent or accents are included or preferred in this idealized norm, though it is plausible that the British 'Received Pronunciation' (RP) and General American are the leading varieties. Perhaps the reason testing research has largely avoided the topic of precisely how the educated native speaker model has been applied to measure English pronunciation has been due to this point of which accent or accents should be employed in the model. According to the British

Council list of countries where English is an official language [5] there are 56 sovereign states where English is one of the official languages with an additional 25 non-sovereign entities where English is one of the official languages. Even if the standards were to only include the seventeen sovereign states where English is both the primary and official language, or more restrictive to only include the six countries where English is the de facto language[6], it would have social ramifications test developers would rather avoid. Kramsch (1993) argues that the concept of a native speaker has lost all meaning, as it implies a level of heterogeneity not possible among the world's many cultures where English is used as the primary language. The Educated Native Speaker model employed in high-stakes tests of spoken English would suggest that some accents are acceptable, and by so doing suggest that others are not (Nelson, 1995). British RP and General American English are not the only two "globally useful or appropriate versions" (Jenkins, 2006, p.42). English is quite simply everywhere.

2.2.2 The "ownership" of English

English has become a global language, and as a result the ownership of English has come into question (Widdowson, 1994). Graddol stated, "Native speakers may feel the language 'belongs' to them, but it will be those who speak English as a foreign language who will determine its world future" (1997, p. 10). With an estimated 1.2 to 1.5 billion English speakers in the world (Crystal, 1998. p.61), the diversity of use of this common language not only unites people but also separates them (White, 1997). As a result of the global spread and varied use of English, areas of research have emerged such as World Englishes (WE)(e.g. Kachru, 1985; Smith, 1987), English as an International Language (EIL) (e.g. Jenkins, 2002, 2006; Matsuda, 2003) and English as a Lingua Franca (ELF) (e.g. Jenkins, 2009; Seidlhofer, 2000, 2001) to name only a few.

---

[5] Available online at http://www.britishcouncil.org/map_of_countries_where_english_is_an_official_language.pdf. The page sites Wikipedia as its source of data. Retrieved August 10, 2014.
[6] According to Wikipedia – List of countries where English is an official language, Australia, Bahamas, Canada, New Zealand, United Kingdom and the United States are the six countries where English is the de facto language and primary language.

These fields of study have not only altered the awareness of different ‘varieties of English’ amongst linguists, but have affected change in English pedagogy and assessment (see Lowenberg, 1993; Quirk, 1990). The following is an introduction to these three fields and philosophies associated with the global spread of English, and how these ideologies have affected, and continue to affect, the efforts of test developers to reliably measure pronunciation. Similar to the concerns of whether or not tests should measure speakers against the educated native speaker model, the fields of World Englishes, EIL and ELF raise arguments concerning definitive models of performance for rating purposes. These concerns further complicate the issue not only of how, and to what standard test developers should employ for measuring pronunciation, but also from what perspective - native speaker or nonnative speaker, is most important and valid.

World Englishes

The spread of English has resulted in numerous nativized forms as well as uses of English as a lingua franca. As a result, many applied linguists and sociolinguists no longer think of the language as a singular ‘English’, but have begun to refer to the language in terms of ‘Englishes’ (Kubota, 2001; Jenkins, 2006). From this mode of thought comes the research field of ‘World Englishes’ (WE). Perhaps the best known concept in WE is Kachru’s (1985) method of stratifying WE into three concentric circles (illustrated in Figure 1.1 in Chapter 1):

1. The Inner Circle - where English is spoken as a first language (L1), or mother tongue; is comprised of an estimated 375 million people, and includes the United States, the UK, Australia, New Zealand, Canada and Ireland. Kachru refers to these varieties of English as ‘Norm Providing.’
2. The Outer Circle where English is not the dominating L1, but functions as an institutionalized or ‘official’ language; is comprised of 350-450 million people, and includes India, Singapore, Malawi and at least 50 other nations and territories. In Kachru’s scheme these varieties are ‘Norm Developing’ where

conflicts of linguistic behavior and linguistic norms collide. These WE varieties are both exo- and endonormative.

3. The Expanding Circle where English is used as a second or foreign language but has no official status; is comprised of an estimated 750 million to possibly one billion speakers, and includes China, Russia, Japan, Korea, Indonesia, Egypt, much of Europe and others. These varieties according to Kachru are exonormative and 'norm-dependent.'

The question of the ownership of English based on a WE perspective arises due to a shift of focus from Inner Circle varieties out toward the periphery (Outer and Expanding Circles) where the numbers of non-native speakers of English is increasing (Schnitzer, 1995). Kachru (1997) noted nearly twenty years ago that the number of English speakers in Asia was greater than the total number of speakers in the US, UK and Canada combined. Graddol (2006) suggests there could be as many as two billion leaners of English in the next decade (p. 100).

The discussion of ownership, as it applies to the English language as a result of WE, includes the question of whose responsibility it is to make accommodations for communicative success. The answer to this question affects how performance can be measured. English is no longer used exclusively between NSs or NSs with NNSs, but also between NNSs (Graddol, 1997; Kachru & Smith 2009; Widdowson, 1994). Lippi-Green (1997) argued that speakers from dominating language groups defer or reject their role as a listener and place the full burden of communicative success on speakers using non-mainstream varieties when communicating with them. WE adherents encourage Inner Circle speakers to empathize with NNSs, and should be held at least partly to blame for communication 'problems' NNSs face (Crittenden, 1994). WE research has criticized test developers for being overly conservative and not embracing and reflecting the kind of diversity of use of English in the world. Hill and Parry (1994) suggest that tests be designed to measure the local variety of English actually in use in the country where candidates live rather than more standard varieties. Lowenberg (1993) claims that some items on the TOEIC are discriminatory, and rejects the notion that native speakers alone determine the norms and standards of English for testing purposes.

The philosophy of WE suggests that all varieties are of equal value and condemn the monocentric, potentially ethnocentric, notion that the standard form(s) are "correct" and non-mainstream forms are not (Nelson 1995). Seidlhofer (2009) describes varieties, "in sociolinguistic thinking it is primarily identification with a particular community that makes a variety a variety" (p. 238). WE has, of course, not been embraced by all linguists and applied linguists. Quirk (1990) finds the effects WE has had on pedagogy to be so wildly liberal as to refer to it as "half-baked quackery" (p.9)(see also Kachru, 1991 for his rebuttal). And yet WE has also been accused of not reflecting adequate diversity. Tripathi (1998) claimed that although Kachru's stratification of World Englishes is separated into three concentric circles, the model lacks the necessary complexity needed to include the many varieties of English that often exist within a Circle, country or region.

WE supporters criticize tests and test developers' reliance on Inner Circle norms for measuring all speakers of English, yet the field of WE does not offer any solutions to test developers to produce unbiased assessments. Applying a WE philosophy for testing speaking and/or pronunciation purposes has similar problems as the educated native speaker model. Both WE and the educated native speaker model are lacking in adequate clarity or in their ability to be defined in any agreed upon manner that could be reflected in score descriptors; one appears to be overly conservative and prescriptive, and the other too accommodating to any deviations from NS norms simply because the different variations are deemed 'acceptable' in their local contexts. Both fields raise test validity concerns when applied for modeling performance. The fields of English as a Lingua Franca (ELF) and English as an International Language (EIL) make attempts to clarify English use as a global language, though as will be shown, present their own validity concerns and difficulties to reliably testing speaking and pronunciation.

English as an International Language

With the advent of the recognition and influence of World Englishes, the fields of English as a Lingua Franca (ELF) (e.g. Gnutzman, 2000; Seidlhofer, 2001) and

English as an International Language (EIL) (e.g. Jenkins, 2000; McKay, 2002) (among others) have emerged. These fields each attempt to provide greater insight and some measure of description to reflect English use and purposes beyond the Inner Circle perspective. Both are mainly concerned with NNSs communicating with other NNSs, rather than English communication including NSs. This is because it is between NNSs where the majority of English use is occurring (see e.g. Jenkins, 2006). What will be shown in this section is that while these two fields attempt to provide clarity concerning the use and pronunciation of English, they present additional challenges to test developers than those previously discussed related to the Educated Native Speaker model or World Englishes.

The field of EIL is primarily concerned with the Englishes of Kachru's Outer Circle, but is increasingly including Englishes from the Expanding Circle (Seidlhofer, 2000, 2004; Jenkins, 2006). Precisely defining EIL is problematic, but Smith's (1987) conceptualization of EIL can be summarized to include three basic tenets: (1) people from numerous linguistic and cultural backgrounds use English as a means of communication; (2) native and non-natives alike own the language; (3) it is not imperative that non-native speakers use native speaker norms or standards. EIL is purported to be 'culture-free' (McKay, 2002) and in stark contrast to Quirk's (1981) 'Nuclear English', yet Quirk claimed his model to be as, "culture-free as calculus, with no literary, aesthetic or emotional aspirations" (p.155). Nevertheless, EIL, like World Englishes includes a multitude of varieties of English each claiming their own validity based on purpose and effectiveness in their own contexts.

English as a Lingua Franca (ELF) and The Lingua Franca Core (LFC): implications for testing pronunciation

English as a Lingua Franca concentrates on the use of English in the Expanding Circle, and intentionally excludes native speakers from data collection. Jenkins states that ELF (also referred to as English as an International language (EIL)), researchers:

> seek to identify frequently and systematically used forms that differ from inner circle forms without causing communication problems. . . their purpose is not

to describe and codify a single ELF variety. The existence of ELF is not intended to imply that learners should aim for an English that is identical in all respects. ELF researchers do not believe any such monolithic variety of English does or ever will exist.

(Jenkins, 2006; p. 160)

ELF can occur between any members of the three circles (Seidlhofer, 2009, Seidlhofer, Breiteneder & Pitzl, 2006; Jenkins, 2009), and allows for a wide range of variation of use. While on the surface it seems an *anything goes* approach to English usage, it is different from World Englishes in that research has resulted in an attempt to define what the requirements are for intelligible, effective English usage in contexts both inclusive and exclusive of native speaker participation - the development of the Lingua Franca Core (LFC).

Perhaps the most detailed attempt to clarify the requirements of intelligible English in a global context is Jenkins' (2000) 'Lingua Franca Core'. The LFC is an attempt to provide testers and teachers with a comprehensive description of appropriate or acceptable parameters of pronunciation that focus on intelligibility. Intelligibility, rather than 'native like' pronunciation would be the ultimate goal for L2 learners. The LFC is not only intended as a syllabus for EIL/ELF pedagogy, but is intended to be a guide to test developers and raters in how to effectively measure pronunciation performance. Field (2005) states the LFC is, "suggesting that a new international form of English may evolve that retains those features most critical to intelligibility between nonnative speakers but suppresses others that are peripheral" (p.401).

Gimson (1980) was among the first to suggest that English learners' nonnative pronunciation could still be intelligible. He attempted to establish what he called the *rudimentary international phonology*. This system suggested that some aspects of native English pronunciation that were known to cause difficulty for nonnative speakers, such as voiced consonants and diphthongs, were not detrimental to what he determined was *minimum general intelligibility*. Brown (1991) described Gimson's simplification of vowels as, "all vowels may be reduced to a central long and short pair /əː, ə/, and the following sentence is still reasonably intelligible. /ˈwən də jə ˈθəŋk ðə ˈtʃəldrən wəl gət ˈhəːm frəm ˈskəːl/"(p. 50). Gimson's model makes provisions for nonnative speech patterns of pronunciation, though it is still largely focused on the learner achieving native speaker pronunciation in what he labeled *high acceptability.* He described this ultimate goal as, "a form of speech which the native listener may not identify as non-native, which conveys information as readily as would a native's and which arrives at this result through precision in the phonetic (allophonic) realization of phonemes and by confident handling of accentual and intonational patterns" (Gimson 1980, p. 303).

Jenner (1989) developed the *common core* of phonological features that determine intelligibility regardless of accentedness. Like the LFC that would follow, Jenner considered what the requirements for intelligible speech were without the need of native pronunciation patterns as the goal. Like Field (2005) noted, attempts to teach native pronunciation patterns were unrealistic and too time consuming. Both Jenner and Jenkins would attempt to provide both teachers and language testers a description of pronunciation performance that was focused on international intelligibility rather than any similarity to native patterns. Jenner argued that suprasegmentals, such as intonation, word stress, pitch and rhythm were of greater importance than segmentals (phonemes) to intelligibility. Many researchers have continued to promote this idea (Avery & Ehrlich, 1992; Morley 1991; Smith & Rafiqzad, 1979); however, Jenkins does not.

Jenkins (2000) suggested that the key to intelligibility of English speech between nonnative speakers relied more on segmentals than suprasegmentals. She argued that the evidence suggesting suprasegmentals' importance was insufficient (see also Levis, 1999). Concerning word stress, she argued, "word stress rules are so complex as to be unteachable" (p.20). In her theory only nuclear stress is learnable, and suggested that most of the other aspects of intonation and other weak forms (e.g. connected speech, stress-timing, reductions) were not. By focusing on phonemes, she developed the LFC and outlined what she deemed were the necessary characteristics of intelligible pronunciation.

The LFC core is presented concisely in Jenkins' (2002) article titled *A Sociolinguistically Based, Empirically Researched Pronunciation Syllabus for English as* *an International Language.* The LFC identifies the phonological and phonetic features deemed most "crucial as safeguards of mutual intelligibility" (p.96). It focuses on the needs of NNSs and dismisses the comprehension needs of NSs since, "in EIL the listener is more likely to be an NNS". It is described as:

1. The consonant inventory with the following provisos:

- some substitutions of /θ/ and /ð/ are acceptable (because they are intelligible in EIL);
- rhotic 'r' rather than non-rhotic varieties of 'r';
- British English /t/ between vowels in words such as 'latter', 'water' rather than American English flapped [r];
- allophonic variation within phonemes permissible as long as the pronunciation does not overlap onto another phoneme, for example Spanish pronunciation of /v/ as [β] leads in word-initial positions to its being heard as /b/ (so 'vowels' is heard as 'bowels' etc.).

2. Additional phonetic requirements

- aspiration following word initial voiceless stops /p/ /t/ and /k/ e.g. in [p^{h}ɪn] ('pin') as compared with /spɪn/ ('spin'), otherwise these stops sound like their voiced counterparts /b/ /d/ and /g/;
- shortening of vowel sounds before fortis (voiceless) consonants and maintenance of length before lenis (voiced) consonants, for example the shorter /ñ/ in 'sat' as contrasted with the longer /æ/ in 'sad', or the /iː/ in 'seat' as contrasted with that in 'seed'.

3. Consonant clusters

- no omission of sounds in word-initial clusters, e.g. in promise, string;
- omission in middle and final clusters only permissible according to L1 English rules of syllable structure, e.g. 'factsheet' can be pronounced 'facsheet' but not 'fatsheet' or 'facteet';
- /nt/ between vowels as in British English 'winter' pronounced /wɪntər/ rather than American English where, by deletion of /t/, it becomes /wɪnər/;
- addition is acceptable, for example 'product' pronounced[pərˈɒdʌk ʊtɔ] was intelligible to NNS interlocutors, whereas omission was not, for example 'product' pronounced /ˈpɒdʌk/.

4. Vowel sounds

- maintenance of contrast between long and short vowels for example. between 'live' and 'leave';
- L2 regional qualities acceptable if they are consistent, except substitutions for the sound /ɜː/ as in 'bird', which regularly cause problems.

5. Production and placement of tonic (nuclear) stress

- appropriate use of contrastive stress to signal meaning. For example, the difference in meaning in the utterances 'I came by TAXi' and `I CAME by taxi' in which nuclear stress is shown in upper case. The former is a

> neutral statement of fact, whereas the latter includes an additional meaning such as 'but I'm going home by bus'.
>
> (Jenkins 2002; p. 96-97)

There are concerns both pedagogically and for testing purposes with the LFC because it does not reflect any single variety of English actually spoken anywhere. The literature does not reflect or suggest that Jenkins considers the LFC or EIL to be considered an individual variety of English that any particular group or groups of English speakers use. Nevertheless, it is questionable if the LFC is a practical or possible goal for pedagogy (Dauer, 2005) or testing (Taylor, 2006). Jenkins (2002; p.100) points out pedagogic problems of LFC and EIL in classrooms comprised of mainly the same L1 speakers echoing Bygate's (1988; p.76-77) concern that such classes permit and possibly encourage fossilization of "deviant L2 forms." In short, she recognizes that "the vast majority of English teaching takes place in same-L1 classrooms in the learners' own countries" (p.100), and that these situations result in what can be described as different 'accents' of English. Jenkins makes this point clear stating:

> In effect, what I am claiming is that the items which are excluded from the LFC are not crucial to intelligibility in EIL contexts, and that they can therefore be considered areas in which L1 transfer indicates not 'error' but (NNS) regional accent. In other words, what we have here is a redefinition of phonological and phonetic error for EIL: one which incorporates the sociolinguistic facts of regional variation instead of regarding any deviation from NS pronunciation as a potentially harmful error (the EFL perspective).
>
> (Jenkins, 2002, p.97)

Jenkins makes an attempt to clarify what features of accented speech are acceptable and contribute to different accents but do not negatively affect intelligibility. However, the fact remains that the LFC is an attempt to model English pronunciation ideals for pedagogy and testing based on descriptions of speech production not actually represented in the already innumerable varieties of English. While its ideals are defined it can still be argued as no less complicated or unrepresentative a form than the Educated Native Speaker Model.

The concern related to how the LFC affects rater selection is due to how it was determined. Jenkins adopted the perspective that since previous research concerning mutual intelligibility of English had been based entirely on native speaker listeners' perspectives, the conclusions may not be the same for nonnative speakers. As stated earlier, the LFC justifies its relevance based on the fact that English communication is currently dominated by exchanges between NNSs. What Jenkins found was that NSs and NNSs employ polar opposite processing strategies to determine intelligibility. Jenkins found that NSs primarily use top-down processing strategies whereas NNSs employ bottom-up strategies. Nuclear stress and segmental errors affect NNS listeners, whereas NS listeners find suprasegmental errors detrimental to intelligibility. NSs utilize contextual and syntactic information to compensate for segmental errors, whereas NNSs seem unable to employ such strategies (Dauer, 2005). The LFC is a syllabus and guide to determining what can be described as "errors" in pronunciation, but they are based only on the NNS perspective and ignore aspects of pronunciation that negatively affect intelligibility for NS listeners, yet most raters in high-stakes tests are NSs (Xi & Mollaun, 2009).

Both ELF and World Englishes have difficulty determining notions of 'correctness' and 'error' in regards to not only pronunciation, but other aspects of production (e.g. Bamgbose, 1998; Bhatia, 1997; Kachru, 1992). 'Appropriateness' has supplanted what qualifies as 'correct' or connotations of 'error.' Seidlhofer (2000) argues that the orientation of Teaching English as a Foreign Language (TEFL) should shift "from correctness to appropriateness, from parochial domesticity and exclusive native-speaker norms to global inclusiveness and egalitarian license to speak in ways that meet diverse local needs" (p.52). These instances of correctness appear situational and outside the normative standards of the Inner Circle. Jenkins (2000) explains:

> There is really no justification for doggedly persisting in referring to an item as 'an error' if the vast majority of the world's L2 English speakers produce and understand it. Instead, it is for L1 speakers to move their own receptive goal posts and adjust their own expectations as far as international (but not intranational) uses of English are concerned.

(Jenkins, 2000, p. 160)

Some researchers argue that the normative standards established over centuries have been deemed of less importance as a result of ELF and the like (e.g. Prodromou, 2006). Concerning pronunciation, Sobkowiak (2005, p.141) argues that the ELF approach will lower the standards of Received Pronunciation “down into the gutter with no checkpoint along the way.” The discussion of ownership of English in EIL and EFL, like WE, appears to have more to do with numbers than with anything else. Pickering (2006) when discussing the matter of the global number of NNSs states, “Once we acknowledge this revolutionary change, it becomes clear that models and practices that privilege native varieties of English are no longer serviceable” (p.1). The mere matter that NNSs outnumber NSs seems to justify a shift of perspective from Kachru’s interpretation that Expanding Circle speakers are norm-dependent. It is as though due to their sheer numbers of speakers their status is deserving of being ‘norm developing’, or even ‘norm providing.’ Quirk (1990) argued against such mob-rule notions, for example when he criticized a Japanese textbook[7] for suggesting that it is ‘good enough’ if the speaker can make themselves understood, as such attempts represent ‘a respectable variety of English’ (p.9). Estimations of error or correctness based on notions of appropriateness to a multitude of situational factors certainly raises concerns and problems for test developers aiming to embrace this shift in attitude toward English or Englishes. And Seidlhofer (2009), one of ELF’s strongest supporters concedes, “Unless we know a good deal more about how much speakers of the Expanding Circle really use English in their communities of practice, what their shared repertoires look like, and which communication processes characterize ELF as it is used in jointly negotiated enterprises, the significant contribution that ELF speakers make to norm development will remain invisible” (p.239).

Research has shown negative attitudes toward World Englishes, ELF and the LFC and other non-native varieties of English from both teachers (e.g. Llurda, 2007; Young & Walsh, 2010) and students (e.g. Timmis, 2002; Groom, 2012). Jenkins (2006) implies that perhaps English L2 learners do not actually aim to have native like

[7] *Four Seasons Composition Book* – Pereira & O’Reilly (1988).

pronunciation suggesting that, “the use of more target-like forms in the different L1 pairings was born of a desire to be understood rather than a desire to be ‘native like” (p. 47), and even advises to testers that, “examinations also have to provide for those students whose preferred goal remain, despite EIL developments, a near-native variety of English” (p.48). The notion that ELF or EIL approach is preferred to native like norms is arguable. Taylor (2006) argues directly with Jenkins on this point stating, “we should not assume that such students are in the minority, nor regard them as ‘unenlightened’” (p.52). Groom (2012) questioned 127 English learners from 22 different European L1 backgrounds, and found 79.53% of learners preferred a NS model for pronunciation and only 3.1% preferred a NNS model. 79% of her respondents stated that they speak with both NNSs and NSs, and that they aim toward speaking in a manner that is intelligible and respected by both NSs and NNSs. Sung (2013) suggests NS norms remain the primary focus of pedagogy, but that ELF perform a complimentary role in pedagogy to heighten learners’ awareness of English use in the world. Jenkins herself opposes forcing people to learn linguistic forms against their will (1998, p.120), but there is evidence that in her attempt to replace minority NS norms with EIL, ELF and the LFC she is equally guilty of imposing upon learners, teachers and test developers another unwanted model. Unfortunately, there are no studies the researcher is aware of that examines other stakeholders’ (university administrators and educators for tests like the TOEFL iBT, or business owners and managers for tests like the TOEIC) feelings and opinions of shifting focus away from NS pronunciation norms to ELF, EIL or LFC models.

It is not the intention of this researcher to discredit Jenkins or her work. She raises very valid arguments for pronunciation testing validity (among other points of speaking proficiency and performance), yet there are problems. The concern of whether or not the standards and targets of pronunciation pedagogy and assessment should favor NSs when it is irrefutable that the majority of spoken exchange in English is currently occurring between NNSs is certainly valid. It is also a reasonable concern to question the fact that convergence in speaking tests is often rewarded when NS variants are employed to bring about a successful communication outcome, but when NNS variants are used they are often penalized. Jenkins’ work is based on sound, empirical evidence, but it is this researcher's opinion that the NS, too, must play some role in the determination of what elements of pronunciation determine and maximize comprehension and intelligibility. It would be a greater benefit to establish a more universal understanding of pronunciation intelligibility that could be more easily applied in testing situations. It seems reasonable to suspect that NSs probably will continue to maintain a majority as raters in high-stakes tests, so their input should be considered. Further, it is not suggested that the supporters of ELF do not recognize that tests specifically designed to measure or predict NNS communicative success in NS dominated situations are valid, but are in fact concerned with tests that claim to measure comprehensive English proficiency. Finally, it is arguable that ELF, EIL and the LFC present compelling challenges to valid, reliable testing of pronunciation. And it is arguable that raters of pronunciation are included among those Canagarajah (2006) addressed when stating, "We need to be aware that all speakers of English are affected by the recent geopolitical changes, compelling us to rethink the meaning of norms and proficiency in English for everyone" (p.241). Raters, whether NS or NNS, contribute to the success or failure of communication, and test developers should consider the impact of the global spread of English beyond the NS norms when attempting to describe or define pronunciation standards for determining scores.

2.3 Accented speech in modern English pronunciation assessment – scoring pronunciation without a model

The previous section of this chapter offered reasoning why the model of ideal English pronunciation for both pedagogy and testing purposes should no longer be limited to that of educated native speakers. As a result, the validity-concerns of imposing inner-circle standards to outer- and expanding-circle contexts are real challenges in modern testing (see Lowenberg, 2002). While native speaker pronunciation has long been seen as "the yardstick for intelligibility" (Golombek & Jordan, 2005, p. 520), it is currently too controversial to even include the words 'native' or 'native like' in score descriptors of most current high-stakes tests. The avoidance of such language in the 'pronunciation' and 'delivery' score descriptors of the IELTS OPI and TOEFL iBT (respectively) suggest that the influence of current trends like World Englishes, English as an International Language and English as a Lingua Franca have also influenced tests designed for inner-circle contexts.

The outer-circle influence of inner-circle tests has not been limited to only speaking. Inclusion of nonnative accented speech in the listening section has been called for. Harding (2011) researched the validity of including highly intelligible nonnative speakers in the University Test of English as a Second Language listening subtest. His findings determined that including nonnative accents in the listening subtest did not threaten the validity of the test, and test-takers' attitudes toward the inclusion of nonnative accented speakers in the test were demonstrably positive.

It appears that at least among high-stakes English test developers and researchers the equality of accents Riney et al. (2005) described has been largely accepted: "all accents are linguistically equal" (p. 442). The descriptors raters use to determine pronunciation proficiency in the IELTS and TOEFL iBT may be equally accepting of all accents, but whether or not raters can actually employ them equally between speakers of all accents remains uncertain. In this section of the chapter how pronunciation proficiency is scored in current high-stakes tests of spoken English without a definitive model is described.

If raters are to determine test candidates' pronunciation proficiency without the benefit of a definitive model for comparison, other means must be employed. Both the Cambridge and Educational Testing Services' (ETS) published materials about their tests never mentions directly what raters should primarily focus on to determine pronunciation scores. Xi and Mollaun (2009) when investigating how raters from India perform scoring the TOEFL iBT speaking section for ETS determined rater training lacked specific enough instruction concerning how raters should score pronunciation. Their rater participants that had been trained both in the standard way actual TOEFL iBT raters are trained and with specialized training for Indian raters, but reported, "the raters expressed the need for more detailed guidelines for evaluating pronunciation" (p.27). It is arguable, however, based on the score band descriptors in both the TOEFL iBT and IELTS OPI that pronunciation and delivery scores are mainly based on the intelligibility of test candidates' speech, and the amount of ‘listener effort’ raters experience processing test candidates’ speech. As stated previously, scores based on such determinations are rater-dependent and not speaker-dependent. This argument is based on the hypothesis that all raters, whether trained or untrained, do not process all accents equally. Accents and raters’ differences in familiarity with accents affects the intelligibility of speech and the ‘listener effort’ needed to process speech.

2.4 Accent-familiarity and speech perception

Basing pronunciation proficiency scores on intelligibility and raters’ ‘listener effort’ results in outcomes that are arguably rater-dependent. At the heart of the argument of this thesis is the hypothesis that accent-familiarity affects speech perception and intelligibility. In this section accent-familiarity is first defined, followed by examples from the literature related to how familiarity affects speech processing. In particular, Gass and Varonis’ paper (1984) is discussed at length, as it is the most widely referenced example of familiarity-based research concerning comprehension and intelligibility. This section concludes with a discussion of why it is not possible or practical for test developers to suggest that speakers of any accent will be equally judged for pronunciation by all raters.

2.4.1 Accent-familiarity

Accent-familiarity, as a rater trait, has in recent years become a matter of importance and focus in English language testing research (e.g. Carey et al., 2011; Winke et al., 2013; Xi & Mollaun, 2009). The meaning of accent-familiarity, interestingly, has not been clearly defined in the literature. Accent-familiarity has instead been ‘determined’ based on different characteristics or experiences acquired by the subjects involved. For example, English L2 speakers sharing the same L1 are considered to be familiar with with the patterns of their particular L1 accented English (Kim, 2009; Xi & Mollaun, 2009). Other examples from the literature include test raters being regarded as familiar with the accents of test candidates from the same country or countries where the raters have lived (Carey et al., 2011), and that raters could become familiar with the English accents of speakers whose native languages the raters have studied as an L2 (Winke et al., 2013). Browne and Fulcher (forthcoming) were the first to attempt to define accent-familiarity by suggesting, "accent-familiarity is a speech perception benefit developed through exposure and linguistic experience"(p.3). Though accent-familiarity has not long been defined, nor is there any reason to presume it will not continue to be interpreted in various ways, nearly all studies concerned with how familiarity affects speech perception include in their discussions Gass and Varonis' (1984) study.

2.4.2 Gass and Varonis, 1984

Perhaps the most cited research concerning how familiarity affects the intelligibility of nonnative speech is Gass and Varonis' (1984) article *The effect of familiarity on the comprehensibility of nonnative speech.* Gass and Varonis hypothesized the construct of comprehensibility in a manner more commonly attributed to intelligibility – that of word/utterance identification. Two of their findings are of particular importance to this study; namely, that familiarity with a particular nonnative accent aids intelligibility, and that familiarity of topic (or context) is the leading contributor to intelligibility. Another point of theirs that is important to this study is how they conceptualized (though did not define) comprehensibility. The following is a critical review of this important piece of research. A summary of the study and its main findings are presented first followed by a discussion of some concerns that have not previously been addressed in the literature that are of importance to this study.

Gass and Varonis investigated the effects four types of familiarity have on native speakers' intelligibility of nonnative speech. The four types of familiarity they investigated and determined contribute to intelligibility were:

1. Familiarity with topic of discourse - both with a specific topic and based on "real world" knowledge familiarity
2. Familiarity with nonnative speech in general

3. Familiarity with a particular nonnative accent
4. Familiarity with a particular nonnative speaker

One hundred forty-two NS students at the University of Michigan in the United States participated as listeners, and four male NNS speaking participants (Arabic L1 n=2; Japanese L1 n=2) were selected from a pool of fifteen advanced level ESL students. Each speaker participant was deemed to be "equally comprehensible" (p.67) by review of ten ESL teachers. The speaking participants were recorded completing three tasks: (1) reading a story[8]; (2) reading a set of five 'related sentences' that pertained to the story though themselves were not included in the reading; (3) a set of 'unrelated sentences' with contexts or topics that could be considered pertaining to 'real world knowledge' were read. The recordings were used to create 24 different 'tapes'; each tape included first either a reading of the 'related' or 'unrelated' sentences, followed by a reading of the story and followed by the set of sentences not included prior to the story read by different combinations of speakers. For example, tape one was comprised of 'Japanese speaker one' reading the unrelated sentences followed by 'Arabic speaker one' reading the story and then also the related sentences; tape seven was comprised of 'Japanese speaker one' reading the 'related' sentences, followed by 'Arabic speaker one' reading the story and 'Arabic speaker two' reading the 'unrelated' sentences. Each tape included a different variation of speakers and orderings of sentence lists read prior to the story or after[9]. This design provided four independent variables:

1. The four speakers
2. The two possible positions of the speakers (pre-story or post-story)
3. Three potential possibilities for the post-story position (1. different language; 2. different speaker; 3. same speaker)
4. Two conditions ('related' or 'unrelated') of the sentences to the story

[8] The North Wind story
[9] Two Spanish speakers' recordings of the two sets of sentences who were also determined to be 'equally comprehensible' were included on seven of the tapes as controls in the prereading position, but their resulting data was not included in the analyses.

Each of the 142 NS participants listened to one tape each and completed two tasks: transcribing each sentence from both lists and writing a short summary of the story to determine comprehension. Mistakes in the sentence transcriptions were counted as 'errors', and the mean number of errors were calculated for each speaker.

Comprehensibility was never clearly defined in their article. Instead, comprehensibility of nonnative speech to native speakers was schematized as:

$$C = p_{\alpha} + g_{\beta} + f1_{\gamma} + f2_{\delta} + f3_{\varepsilon} \ldots fl_{\zeta} + s_{\eta} \ldots$$

> C=comprehensibility. p=pronunciation, g=grammar, f1=familiarity with topic, f2=familiarity with person, F3=familiarity with speaker's native language, fl=fluency, s=social factors
>
> (Gass & Varonis, 1984, p. 67)

This schematization takes into account various aspects of speech, but the measures of the effects of different familiarities on comprehensibility were determined through transcription error ratios alone. Interestingly, transcription tasks have traditionally been used to determine intelligibility (e.g.: Catford, 1950; Smith & Rafiqzad, 1979; Smith & Nelson, 1985) and not comprehension. It is true that their research included having the listening participants write brief summaries of the story they heard "so that we could determine whether they had indeed understood the story" (p.69), but beyond that mentioning in the methodology, there is no further discussion of these summaries or results of any analyses anywhere in the paper. It appears the reader is to assume that all summaries from the data reflected accurate accounts of the story, but this is only speculation. Equally interesting is that the terms 'intelligible' and 'intelligibility' fail to appear at all in the study. It is, arguable to suggest that the findings of Gass and Varoins (1984) are more associated with the affects different types of familiarity have on intelligibility and not comprehensibility. The foundations for all of their conclusions are based on the results of transcription exercises (intelligibility tasks). Of course, without a clear understanding of the interpretations of the terms 'comprehension', 'comprehensibility', intelligibility' and 'intelligible' they adhered to it is not possible to truly confirm or refute their conclusions.

Gass and Varonis' findings concluded that 'familiarity of topic' is the greatest contributor to intelligibility of the four familiarity types researched. This was determined by one-tailed t-tests comparing the pre- and post-text positions of the related sentences. The results revealed a significant difference of means of errors ($p<.05$) for three of the four speakers (p.72). More errors were reported in the pre-story transcriptions of the 'related' sentences than in the post-story position suggesting that NSs are more capable of determining the content of NNSs' speech if they know the specific topic. Likewise, the 'unrelated' sentences determined to be comprised of 'real world knowledge' resulted in a significantly lower instance of errors ($F=19.64$, $p=.0001$) when compared to the 'related' sentences when they occurred in the pre-story position on the tapes. These findings reflect Dauer's (2005) finding that suggest native speakers employ primarily top-down strategies for comprehension and intelligibility. The differences in error ratios suggest that native speakers utilize contextual and syntactic information to compensate for segmental errors.

Familiarity of speaker, familiarity of accent and familiarity of nonnative speech in general were found to contribute to the intelligibility of nonnative speakers though these findings were not based on any significant differences in the data. Familiarity of speaker and familiarity of accent were determined contributors to intelligibility by observing speaker error instances in the pre and post story positions. Listeners did tend to find the speakers of the same accent more intelligible in the post story sentences when they had encountered the other speaker of the same accent reading either the pre story sentences or the story, as well as for individual speakers but not in all cases. 'Japanese speaker two' was not found to be more intelligible reading the related sentences in the post position when he had provided some example of his speech prior when compared to his error reports having heard 'Japanese speaker one' prior. 'Arabic speaker two' had the same lack of improved error reports for the unrelated sentences. In short, no significant results were reported to substantiate the claim that familiarity of speaker contributes to intelligibility. Additionally, the data only revealed a 'tendency' for familiarity of nonnative speech in general to contribute to intelligibility. The evidence they provided for this claim was the, "small difference between pre-text and post-text unrelated sentence" (p.77). Additionally, no data was collected or included concerning the 142 listener participants' overall familiarity with nonnative speech to demonstrate comparative differences, so it can only be speculated what their actual familiarity with particular accents, the included accents or nonnative speech in general, were.

As stated earlier, the theory that familiarity with a particular accent contributes to intelligibility is central to this research. Gass and Varonis claim accent-familiarity is a contributing factor to intelligibility, but they did so without significant or compelling evidence. In fact, there is very little discussion in their paper dedicated to this claim. It appears that perhaps the inclusion of accent-familiarity in the study was a post hoc theory that occurred during the analyses. The belief that familiarity with a particular nonnative accent may have been an afterthought in their research is based on the following passage:

> A final point to mention is that there is a tendency for the Arabic speakers to elicit more comprehension errors than the Japanese speakers. There are a number of possible explanations for this: (1) It may be that many of our subjects were more familiar with a Japanese accent than an Arabic one, as opposed to the experienced teachers who were very familiar with both (in fact, a number of native-speaking judges mentioned having Oriental TAs)
> (Gass & Varonis, 1984, p. 74)

It is clear that Gass and Varonis did not collect accent-familiarity data for either Arabic or Japanese accents from the listening participants. It is surprising that they made a point to include that perhaps one reason some judges might be familiar with Japanese accented English is because some of the participants had "Oriental TAs". Such a statement implies that, at least in some manner, all 'Orientals', or Asians, share the same or similar accent, which is patently not true. What is clear is that the claim that familiarity with a particular nonnative accent contributes to NSs' intelligibility of NNSs' speech was not substantiated in their study. Equally, familiarity with nonnative speech in general as a contributing factor to comprehensibility was never substantiated, but only speculated.

It is not logical to suppose that Gass and Varonis' findings suggest that accent-familiarity would also include familiarity of the entire lexical inventory and/or knowledge of topics all speakers of an accent possess. Unfortunately, no mention of this limitation of the meaning of 'comprehension' was included concerning accent-familiarity's ability to facilitate intelligibility. This thesis argues that any definition of intelligibility that includes both locutionary and illocutionary force cannot be applicable to accent-familiarity benefits. Accent-familiarity primarily facilitates, or increases intelligibility success-rates, and must be limited to only advantages determining the phonological content of speech. Context familiarity facilitates understanding or implied meaning - aspects more associated with locutionary force, or 'comprehension' and illocutionary force, or 'interpretability' (as they are defined by Smith and Nelson, 1985). It is unfortunate that comprehension is not clearly defined by Gass and Varonis, or that the differences between content familiarity and accent-familiarity are not differentiated. Nevertheless, the benefits of accent-familiarity facilitating comprehension are grouped in their conclusions alongside context-familiarity, and these fundamental differences in how different types of familiarity affect speech processing (intelligibility, comprehension and interpretability) are not better explained.

Gass and Varonis (1984) have been cited in several papers related to, but not limited to, familiarity (e.g. Anderson-Hsieh, Johnson & Koehler 1992; Isaacs 2008, Winke et al., 2012) and comprehension of nonnative speakers (e.g. Pica et al. 1989; Munro & Derwing, 1995, 1999; Derwing & Munro, 1997). An interesting point is that their paper has never been questioned or criticized (to this researcher's knowledge) for its lack of clearly defining comprehensibility or the lack of adequate evidence to support all of the claims they made. It would seem that since the conclusions they made are perfectly logical that it was not necessary to doubt or question them. But it is necessary now, for though their findings reflect theories shared in this research, their analyses provided little evidence to support all of the claims they made.

Gass and Varonis may not have fully substantiated the claim that accent-familiarity contributes to non-native speech processing with significant statistical evidence, but the research presented in this thesis contributes toward the confirmation that they were correct nonetheless. Gass and Varonis did provide reliable evidence to suggest that familiarity of topic (or context) significantly increases comprehension, or intelligibility (depending on how the terms are defined) of NNSs' speech by NSs. Bent and Bradlow (2003) also determined contextual clues benefit the intelligibility of NNS's by NS's. By recognizing what has been established through empirical research concerning the effect of context familiarity, this study aimed to better examine and better determine how accent-familiarity functions as a contributor of intelligibility of nonnative speech by first reducing and controlling the content of the speaking participants' utterances. If context familiarity can be reduced or eliminated, accent-familiarity can be more accurately measured.

This research is primarily concerned with how raters' varying familiarities and levels of familiarities with different English accents affect the scores they deliver. If context-familiarity is the greatest aid to comprehension, it is reasonable to believe that context-familiarity may interfere with measuring how accent-familiarity affects comprehension and intelligibility. Jenkins (2000) determined that NSs employ primarily top-down strategies and find suprasegmental errors detrimental to intelligibility. NSs utilize contextual and syntactic information to compensate for segmental errors, which was also confirmed by Neely and Keefe (1989) in psychology and psycholinguists research with what they coined the *relatedness proportion effect*. Similar to the Exemplar Theory that will be discussed later in this chapter, Neely and Keefe's theory suggests that when subjects are engaged in lexical decision tasks, an expectancy set of items related to the known or understood content is facilitated. If the context is not clear the lexical access to the target is reduced. With this contextual mechanism inhibited, pronunciation can only be accurately perceived through successful pronunciation processing. It is suspected that accent-familiarity affects the processing of pronunciation.

Concerning pronunciation and its importance to speech perception related to familiarity, it is revealing that 'pronunciation' was given the alpha (α) position in Gass and Varonis' schematization of comprehensibility (or intelligibility). This choice seems to suggest a hypothesis that intelligibility of nonnative speech for native speakers begins with pronunciation. It is unfortunate that they did not explain or discuss this choice in their paper. It is unfortunate because pronunciation is often not given highest (or even high) priority in some high-stakes tests – possibly for the reasons already discussed concerning the difficulties of determining a single model of performance. The rating scales of speaking tests like the TOEFL iBT and IELTS do not reflect this level of 'alpha-importance' to pronunciation. As stated earlier, the TOEFL iBT includes pronunciation, as only one aspect of a larger 'delivery' construct. If pronunciation is the starting point for all intelligibility, it should warrant pronunciation a more prominent role in determining scores on tests. Additional discussion concerning this schematization of intelligibility, or comprehensibility according to their application may have proved beneficial toward increasing the importance of pronunciation in high-stakes tests.

2.4.3 The contributions from Bent and Bradlow and Munro, Derwing and Morton

Bent and Bradlow (2003; Bradlow and Bent, 2003, 2008) have contributed greatly to our recent understanding of familiarity and its effect on speech perception. They provided some of the most important evidence to support the theory that accent-familiarity positively affects intelligibility. They coined the phrase, "interlanguage speech intelligibility benefit" (Bent & Bradlow, 2003, p. 1600) that occurs between L2 speakers of the same L1. Interlanguage benefits to intelligibility have everything to do with speech patterns, or what could be argued as 'accents'. They investigated whether or not native language background influences the intelligibility of English speech when speaker and listener share a common L1. This notion was also theorized by Smith, Bradlow and Bent (2003) (also Van Wijngaarden, 2001; Van Wijngaarden, Steeneken & Houtgast, 2002). The participants included two Korean, two Chinese and one native English speaker reading sentences from the revised Bamford-Kowal-Bench Standard Sentence Test (BKB-R). The BKB-R sentence lists were designed to measure sensorineural hearing loss in children, and include gap-fill intelligibility tasks. Chinese (n=21), Korean (n=10), English native speakers (n=21) and a mixed group of non-native talkers of English from various native language backgrounds" (p. 1603) (n=12) participated as listeners, and completed the intelligibility tasks. Bent and Bradlow (2003) confirmed the theory that L2 learners from the same L1 share "linguistic and phonetic knowledge that facilitates speech communication in the non-native language" (p. 1602). Their findings determined that speech from relatively high proficiency level nonnative speakers is as intelligible as native speech to speakers of the same L1.

'Mismatched interlanguage speech intelligibility benefits' were also determined possible in their study. These 'mismatched' intelligibility benefits were observed occurring between Chinese and Korean English speakers, and suggested the cause may be "similarities in the sound structure of the two languages" (p. 1607). However, they later suggested that such mismatched interlanguage speech intelligibility benefits are likely less to do with similarities between speakers' L1s and more to do with "certain tendencies in foreign-accented English regardless of native language background" (p. 1608). This conclusion was made from observing instances of potential intelligibility benefits occurring between speakers and listeners from diverse L1 backgrounds. This finding is important because it suggests that familiarity with nonnative speech in general might lead to speech perception benefits. They confirmed that listeners can adapt to accented speech through exposure, and initiated the understanding that accent-familiarity can be developed, and that it positively affects speech perception and intelligibility. Interestingly, they determined no such intelligibility benefits occurring among the native English speaker participants in their 2003 interlanguage study (Bent & Bradlow, 2003). In 2003 they could not confirm that mismatched interlanguage speech intelligibility benefits were likely for native speakers, but their later work would establish the possibility was likely.

Bradlow and Bent (2008) further investigated speech intelligibility benefits resulting from familiarity with what they labeled, "foreign-accented English" (p. 707). They examined how native speakers coped with first a single Chinese-English speaker, and then with four Chinese-English speakers reading BKB-R sentences. This study expanded on their earlier study (Bradlow & Bent, 2003) that determined listeners could be trained to cope with Chinese-English. Their 2008 findings confirmed their earlier findings, and additionally concluded that accent-familiarity intelligibility benefits were possible both at the single speaker level, and possible at the accent level. Additionally, they determined that native speakers are also capable of developing accent-familiarity benefits. Their findings concerning accent-familiarity benefits are of great importance to this research.

It is necessary to include in this discussion of the contributions of Bent and Bradlow, the outcomes of Munro, Derwing and Morton's (2006) research of how shared and differing L2 English accents impact intelligibility, comprehensibility and measures of accentedness, as they reveal both contributory evidence and challenges to the notions of both interlanguage speech intelligibility benefits and mismatched interlanguage speech intelligibility benefits. In their study 40 native listeners of Cantonese, Mandarin, Japanese and English (ten listeners from each L1) listened to, transcribed and scored the same group of 48 native Cantonese, Japanese, Polish and Spanish (12 from each L1) speakers' utterances in English for accentedness and comprehension. Their findings did not reveal the same measures of interlanguage speech intelligibility across all shared L1 to L2 groupings that Bent and Bradlow (2003) had determined. Although the native Japanese listeners did find the Japanese-English speakers' utterances to be more intelligible than all the other accents, the Cantonese-English speakers were no more intelligible to the native Cantonese listeners than the other groups; however, the native Cantonese listeners found the Japanese-English speakers' utterances more intelligible than either the Polish or Spanish speakers. The native Mandarin listeners also demonstrated potentially mismatched interlanguage speech intelligibility benefits by being more successful transcribing both the Cantonese and Japanese speakers' utterances better than they did the Spanish speakers. And the native English listeners demonstrated no intelligibility benefits transcribing the Polish-English utterances even though they had reported having more familiarity with Polish-English than any of the other listener groups had, and showed no group to be more or less intelligible.

It is possible some aspects of Munro, Derwing and Morton's study may have affected the intelligibility tasks. It is not possible to do more than speculate potential limitations to their study; however, it is possible that since the utterances shared a common context that transcription task success may have been affected. All the utterances between 4.5 to 10.5 seconds and extracted from spontaneous speech samples of the speakers describing the same cartoon. The resulting *clips* may have had adequate context similarity that allowed the listener participants to accurately guess any words they might have had difficulty deciphering due to pronunciation quality. It is also possible that the samples did not adequately include characteristic elements of each of the different L1 inspired accents to allow for interlanguage speech intelligibility benefits to be witnessed. It would be beneficial to further investigate the question of interlanguage speech intelligibility, mismatched interlanguage speech intelligibility and whether or not native English listeners can also demonstrate accent-familiarity based intelligibility benefits in a study that controls both the context and the content of the stimuli for transcription and scoring to confirm or reject the findings of Bent and Bradlow (2003; Bradlow and Bent, 2003, 2008), as well as those of Munro, Derwing and Morton (2006).

2.4.4 Why all accents cannot be considered equally intelligible to all raters

There is no evidence that I am aware of to suggest that all listeners, whether native or nonnative, trained raters or untrained novices, process and perceive all accented speech equally. There is, however, evidence to suggest that the opposite is true. In this section, the argument is made that familiarity with an accent affects speech perception and intelligibility of speech in that accent. Current high-stakes tests like the TOEFL iBT and IELTS rely on raters to determine speakers' proficiency scores, at least in part, on the intelligibility of test-takers' speech, and the difficulty the raters experience coping with their speech. Accents are, after all, closely related to identity and social membership (see Golombek & Jordan, 2005; Levis, 2005), so suggesting inferiority of any accent raises concerns of xenophobia. The TOEFL iBT 'delivery' score band descriptors, for example, include more attention to suprasegmentals than segmentals, which is reflective of Jenkins' (2002) LFC. Intelligibility is a key factor. The evidence from the literature suggests that it is neither practical nor prudent to suggest that all raters will find the speech of any given test-taker or accent equally intelligible.

2.5 Speech processing and speaker normalization: examples from outside the fields of language testing and linguistics

This section describes and discusses two theories from outside the language testing literature: the *Perceptual Magnet Effect* and the *Exemplar Theory.* Both were first suggested as potentially in connection with raters' accent familiarities affecting test scores by Carey et al. (2011). Unfortunately, Carey et al. could not include a lengthy discussion pertaining to how these theories may offer important insight into both how speech is processed and how accent-familiarity affects speech processing. I believe these theories provide reasonable explanations of how accent-familiarity functions to facilitate speech perception. Raters are, in the most basic sense, professional listeners, and at the heart of effective listening is effective speech processing. The present study investigates how raters process the pronunciation of test candidates, and questions whether or not test developers can claim that all raters either do, or can be trained to, process all speakers and groups of speakers equally.

Speech is highly variable. Speech variation is not limited to differences between nonnative or native accents, or the differences between the genders or between the young and old. Speech is extremely diverse even between talkers sharing the same backgrounds (Goggin, Thompson, Strube & Simental, 1991). 'Speaker normalization' according to Johnson (2008) is, "a line of research centering on the fact that phonologically identical utterances show a great deal of acoustic variation across talkers, and that listeners are able to recognize words spoken by different talkers despite this variation" (p.363). The Exemplar Theory and Perceptual Magnet Effect provide explanations that offer both clarity concerning how speech is varied, and how listeners (and raters) decipher speakers' acoustic codes into meaningful utterances. These theories also provide evidence that our linguistic experiences and familiarities shape and facilitate speech processing.

2.5.1 The Perceptual Magnet Effect

The perceptual magnet effect is a theory first introduced by Kuhl (1991). The theory suggests that phonetic categories are internally structured, and that prototypes of speech categories affect perception. These ‘prototypes’ used in speech categorization are idealized speech memories the listener has stored in their long term memory, and are utilized for perception. These prototypes, it has been determined, develop from as early as infanthood (see also Kuhl et al. 1992).

The theory includes the matter that the typicality of input is also determined to strongly affect perception. Especially concerning stimuli that is close in proximity to the acoustic space of the prototype, or stimuli that is similar to the prototype, is considered to strongly affect perception. Kuhl (1991) measured how the typicality of stimuli affects perception by presenting two groups of listeners (n=8 per group) with two different referent vowel samples. One group was given the prototype American /i/ vowel sound, and the other group a nonprototype /i/; sometimes also labeled as /e/ or “not /i/” (see Frieda et al., 1999; Iverson & Kuhl, 1995; Lively & Pisoni, 1997; Sussman & Lauchner-Morano, 1995). Each listener group completed 64 discrimination tasks that amounted to a total of 128 trials. Each task required the participants to listen to first their group’s referent vowel sound for one second followed by either one of 32 “surrounding variants” (p. 97), or the same referent vowel sound for an additional 4.5 seconds. The participants were instructed to press a button if the vowel sound changed from the referent, and not press the button if the sound did not change. The tasks included 32 instances where the vowel changed, and 32 with no change. The results determined that the group whose referent vowel was the prototype American /i/ were significantly less capable of correctly distinguishing the referent from the surrounding variants than the group given the nonprototype referent. Even though the phonetic differences of the two groups’ stimuli were equal, the results demonstrated significant differences in the groups’ abilities to distinguish the differences [$t(14)= 6.89$, $p< .001$]. What this established is that the listeners of the prototype group were perceiving the surrounding variants, or variants within close proximity to the acoustic space of the prototype, as the prototype. Perception, it appears, was magnetized to the prototype. This suggests that speech perception is at times warped, altering perception to match an internal prototype, or acoustic memory. The typicality of speech strongly affects this warping effect. Speech that is more typical is more difficult to distinguish than non typical speech, at least concerning vowels.

The theory further suggests that, “phonetic perception is altered as a function of exposure to language” (Kuhl & Iverson, 1995, p. 122-123). This is an interesting aspect of the theory because it attempts to explain not only perceptual difficulties encountered by language learners distinguishing different phonemes in noise, but also why speech production of those same phonemes often include characteristics that can be considered ‘typical’ of speakers of an accent. For example, it is well known that Japanese learners of English have difficulties with /r/ and /l/ distinction and production (Carruthers, 2006). Japanese-English speakers lack the phonemes /r/ and /l/ in their native language, and commonly replace them with what Aoyama et al. (2004) describe as an “apicoalveolar tap /ɾ/” (p. 234) that is present in Japanese phonology. According to the perceptual magnet effect theory, it is possible that Japanese learners of English do not actually perceive the phonemes /r/ or /l/ any differently than they do /ɾ/ because the stimuli is magnetized to that prototype. Of course, this is debatable considering the perceptual magnet research focused only on vowel perception and not consonants, but it is possible.

The perceptual magnet effect theory can be applied to present an interesting potential line of reasoning why raters with contrasting levels of familiarity with different test candidates’ accents score those candidates’ pronunciation differently when using rating scales that rely on measures of listener effort to determine scores. The argument could be made by applying the theory concerning how the typicality of speech in familiar accents affects raters’ ability to distinguish phonetic variants. While Kuhl (1991; also Kuhl et al. 1992) determined that prototypes can be established in infanthood, there is no evidence to suggest that exposure to new speaker groups, or accents, cannot also lead to new prototypes later in life. Speech in familiar accents may be perceived differently according to accent-specific categorical prototypes raters familiar with the accent have. These prototypes might cause difficulty attempting to distinguish phonetic differences as sensitively as unfamiliar raters. It has been argued that prototypes are not “one-size-fits-all”, but rather can be different amongst listeners even from the same language or dialect background (Freida et al., 1999;

Lively & Pisoni, 1997). The internal structure of phonetic categories and their effect on speech perception is not universal among all native or nonnative speakers of English, so it is possible that detecting phonetic variants will not be equal among all raters if the perceptual magnet effect is, in fact, real.

2.5.2 The Exemplar Theory

The Exemplar Theory, or exemplar-based models of speech perception and processing, is a concept from cognitive psychology that is a memory-based framework. This framework of memories allows for detailed representations of various types of stimuli, including linguistic, to be stored and utilized for categorizing, identification and determining meaning (Boomershire, 2006). These memories, called 'exemplars', are like the 'prototypes' described in the explanations of the perceptual magnet effect, except they are not limited to only vowels. Johnson (1996) and Pierrehumbert (2001) were among the first linguists to suggest the theory could serve to explain aspects of speech perception. According to the theory, exemplar memories represent instances of linguistic experiences related to words, sounds, particular people, particular accents or dialects, gender and languages that are stored for a period of time in what are referred to as 'exemplar clouds' (Pierrehumbert, 2001, p.3). These exemplar clouds are also similarly categorized, and are utilized to identify not only the content of speech, but other aspects of the speech itself, as well as the speaker and/or speaker group. It has been found that the retention of very fine details of speech can be retained as exemplars for long periods of time in long-term memory (Pisoni, 1993). "Input is stored as detailed exemplars, which activate, and in turn are activated by, other categories such as stereotypes (e.g., age, gender, dialect, etc.) and phonological generalizations (e.g., final-word nasals are velar)" (Boomershine, 2006, p. 58).

The exemplar theory offers a logical explanation of how speech perception and speaker identification occurs. Exemplar clouds are activated and deactivated during speech perception and normalization according to the similarity and dissimilarity newly encountered speech has with other stored exemplars. Figure 2.1 presents an illustrated schematization of the exemplar model. Incoming speech is compared with numerous exemplars, and as Johnson (2006) explains, “activation from the exemplars feeds up to linguistic categories and to gender categories” (p. 493). The strength of the exemplar is reflected in the width of the arrows in figure 2.1, and the strongest connections serve to identify that the newly encountered input is, in fact, the word “saw” being uttered by a male speaker. This is a very simplified illustration, but could be expanded to include additional clouds pertaining to accent, and individual speaker. After all, a word like ‘saw’ spoken by different speakers of both genders and from various accents can still be identified as the word “saw”, though spectrograms can be used to demonstrate that men and women, and different accents will result in very different vowel formant frequencies (see Johnson, 2005). In the simplest of examples, the exemplar theory explains how it is possible for a person answering the telephone to identify their own mother as the caller before she has completed uttering her first full word.

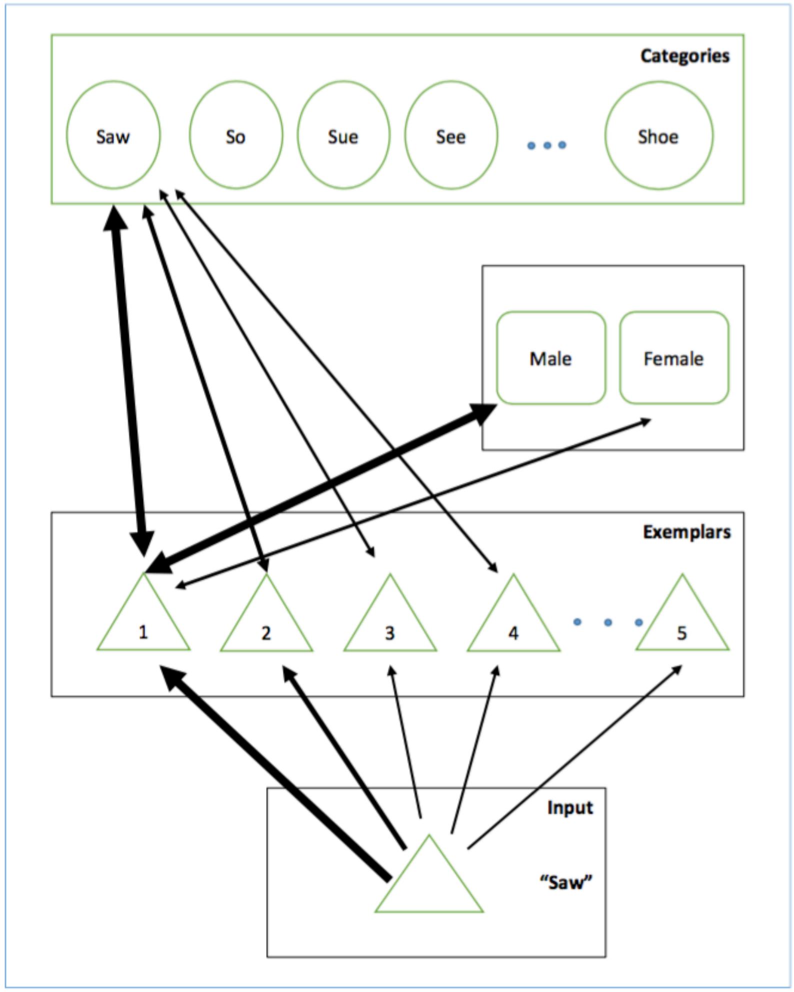


Figure 2.1: A schematic illustration of an exemplar model of speech perception, based on the model illustrated in Johnson, 2006, p. 493.

An important aspect of the exemplar theory that pertains to this study is how frequency of exposure affects the strength and duration of exemplars as they contribute to speech perception. Frequency of exposure leads to familiarity. Mullennix et al. (1989) determined that frequency affects both word identification success, and the difficulty listeners experience completing word identification tasks. In one experiment they had a group of eleven listeners transcribe words from one speaker, and another group of eleven listeners transcribe words from fifteen different speakers. The success rates and times needed to complete the transcriptions were calculated. Significant differences of transcription success rates were observed

[F(1,20) = 7.9, $p < 0.02$], and noted that listeners responded more slowly when confronted with multiple speakers than with only one speaker. This suggests that familiarity decreases processing time; speech processing is less taxing for the listener the more familiar the listener is with the speaker or their accent. As stated earlier, exemplar memories can be stored for relatively long periods of time. This was determined through experiments where listeners were found capable of retaining specific details of speakers' voices when tasked with recalling lists of spoken words one month after hearing them (see also Goldinger, 1992; Goldinger et al., 1991; Martin et al. 1989). The fact that frequency affects the strength of exemplars, and exemplar memories are long lasting suggest that prolonged frequent exposure to a particular accent and speakers of that accent will create both very strong and long lasting specific accent related exemplar clouds.

2.5.3 Speech processing and the problematic nature of speaker normalization

Though speech is extremely varied, successful communication still occurs because "natural language is redundant, there is usually more than enough information to transmit a message" (Goggin et al., 1991, p. 449). The variation in speech resulting from accent, according to Francis (1983) can be considered as noise. The noise affects the message, but the underlying message is still preserved. This idea of noise in the signal suggests that listeners utilize a kind of internal template of speech production that they use to warp each speaker's speech, just like the perceptual magnet effect postulates. The Exemplar theory suggests that speaker normalization does not occur due to the listener warping the sound to match an internal template of production. Instead normalization occurs through the activation of exemplar clouds that represent the identity of the speaker, and deactivates exemplar clouds for other speakers or speaker groups (Nosofsky, 1988; Johnson, 1990). When presented with a new speaker, exemplar cloud activation can still occur, as the listener will activate the exemplar cloud that most closely matches the new input. Activation of exemplar clouds occurs from various cues including recognition of a particular voice, gender or accent. Also affecting activation are the expectations of the identity of the speaker (this can include expectations of speech production from a particular accent or gender), acoustic cues and also visual cues.

In the context of semi-direct tests of spoken English, the exemplar theory presents some explanation of why intelligibility can be affected when raters encounter test candidates speaking in unfamiliar accents. As stated, speech processing occurs by the activation of existing exemplar clouds by new input (Nosofsky, 1988; Johnson, 1990). These activated exemplar clouds include those for speaker group identification. In semi-direct tests raters are not provided with any personal information concerning the test candidate in order to eliminate possible bias; however, according to the exemplar theory the listener still attempts to identify aspects of the speaker, including speaker group or accent, in order to accurately and efficiently decipher their linguistic messages. When encountering an unfamiliar accent, it is possible that the listener, or rater, will incorrectly associate the test candidate as a member of an accent group they do not belong to. This could result in the rater expecting to encounter certain accent related speech characteristics that the incorrectly identified test candidate's accent will not deliver. As a result, intelligibility may be negatively affected, or the measure of difficulty the rater has processing the candidate's utterances increased.

The exemplar theory does not conflict with usage-based frameworks that suggest phonetic targets and patterns of pronunciation are the result of learned behavior developed gradually through repeated use during language acquisition (Hooper, 1976), but instead share an important aspect with them. Namely, the exemplar theory suggests retention and frequency are connected (see Bybee, 2006). The frequency of exposure to the same words, sounds and characteristics of a speaker, or group of speakers, increases the strength of those linguistic memories, or exemplars, to be used to process speech. This is an aspect of the theory that may explain why inter-rater reliability issues could occur due to raters' differing accent familiarities. New, unfamiliar or peculiar examples of speech require greater effort to decode than higher frequency exemplars. They are more difficult because of the necessary effort required to perceive them and associate them with an existing exemplar cloud, or to possibly require the creation of a new exemplar cloud. New linguistic experiences are coded according to their similarity to already existing exemplars. The measure of similarity a speaker, utterance or word has to an already existing exemplar determines the perceptual distance– close or distant – to identify that speaker, utterance or word. The distance in the exemplar perceptual space determines the strength of the new exemplar (Johnson, 2008), and as a result determines the effort needed. Stronger exemplars are processed more easily than weaker ones. As a result, raters require more effort when encountering new accents or speakers with accents they rarely chance upon. The exemplar theory suggests that if a rater recognizes a test candidate's accent the decoding of that speaker's utterances will be easier than speakers whose pronunciation, or accent, is either new to them or infrequently encountered.

The exemplar theory opposes the notion that speech processing and perception are actions listeners' experience equally, but are in fact shaped by our linguistic experiences. These linguistic experiences with speakers from an accent group can be rationalized as accent-familiarity. Familiarity characteristics are included with the exemplar and serve to normalize speech (Johnson, 2005). It is likely that if transcription tasks were coupled with pronunciation scoring, as in the present study, that the exemplar theory could provide some explanation why pronunciation score differences occur between familiar and unfamiliar raters.

The exemplar theory and perceptual magnet effect conflict in some aspects concerning their explanations of how speech is processed, but both offer similar rationalizations concerning the impact familiarity has on perception. This research does not seek to endorse or favor one theory from the other, but are included here as credible examples and explanations from the literature of how the notion that linguistic familiarities possibly, if not probably, affect speech processing and perception. These theories offer supporting rationalizations of the hypotheses set forth in this study that raters' accent familiarities affect the listener effort needed to process test candidates' speech for pronunciation scoring purposes.

2.6 Investigations into raters' accent familiarities as a threat to test scores

This section of the chapter examines and discusses five studies that are most closely related to the focus of this research. At the time of first proposing this topic for doctoral research there had been no published studies that had investigated whether or not raters' familiarity with test-takers' accents affected test scores. The first studies discussed in this section were not directly focused on accent-familiarity (Xi & Mollaun, 2009; Kim, 2009), but included suspicions in their conclusions that raters' familiarity with the test-takers' L1s may have had an effect on scores. These two studies suggested future researchers investigate this concern. This research was a response to their call. The section continues with discussions of Huang, (2013), Carey et al. (2011), and the works of Winke et al. (2011, 2012; Winke & Gass, 2013). These were the first studies to directly address the potential threat raters' differing accent familiarities pose to test reliability and validity, and are most closely related to this study. All of these studies have contributed to the main hypothesis of this study that raters' accent-familiarity differences represent a valid threat to test scores; however, each of these studies approached accent-familiarity differently, and investigated different tests and aspects of tests. By examining these important related studies, it is possible to reveal what has been determined about the potential threat raters' differing accent familiarities may have on high-stakes tests of spoken English, and what has not yet been determined. It is also from these studies that the decision to focus on pronunciation scores and intelligibility was determined.

2.6.1 Xi and Molluan 2009

One of the earliest studies to consider the familiarity of raters with the speech of test-takers impacting scores was that of Xi and Molluan (2009). They investigated whether or not Native Indian TOEFL-iBT raters could be as reliable as native English speaker raters when scoring test candidates from India. Their study determined that native Indians could be trained to be as reliable as native speaking operational TOEFL iBT raters. They were mostly interested in bilingual, or multilingual, Indian raters that spoke English and one or more Indian languages. Their study is worth describing and discussing here because it investigated score differences between groups of raters divided according to not only their native speaking status, but how Indian raters rate speakers whose accents they are familiar with.

The study, published as an official TOEFL iBT report, included 26 rater participants with at least a master's degree and prior experience teaching English to Indian students. The Indian raters, "were tested for their speaking ability, went through an extensive online training program, and . . . completed a rater certification test" (p. vi). Divided into two equal groups, all rater participants completed standard training sessions similar to what operational raters receive. All raters scored the same 300 samples from both Indian and non-Indian test-takers of the TOEFL iBT speaking test. Following the first session one group continued with the standard training, and the other group received additional special training aimed specifically toward calibrating the raters for scoring Indian speakers. The special training included listening to benchmark samples of Indian speakers. Both groups then scored another set of 100 responses. The rater agreement estimations were determined by perfect, adjacent and nonadjacent percentage agreements, Pearson correlations and quadratically weighted kappa. The findings determined not only that the Indian raters scored similarly to actual operational TOEFL iBT raters when scoring both Indian and non-Indian test-takers, but the special training did improve agreement statistics between that group of Indian raters and the operational raters. Higher phi coefficient (.90) for single scores combined across six tasks from the scoring session following the special training than the first session (.84) when scoring the Indian test-takers only, and kappa estimated rose from .83 to .86 following the special training.

It is interesting that Xi and Mollaun (2009) found that training might eliminate any threats to holistic scores on the TOEFL iBT due to familiarity with the speech of test-takers. As mentioned earlier in this chapter, the TOEFL iBT employs a holistic rating scale that could also be described as an analytic scale because it relies on sub scores of various aspects of speech to determine a final score. It is unclear if the Indian raters, including those that received specialized training, would have had similar agreement statistics with the operational raters if they had been asked to only score 'delivery' from the TOEFL iBT proficiency score band descriptors. Questions also remain concerning the Indian raters' actual measures of familiarity with the four Indian languages and their distinctive English accents the test-takers reported as their co-L1 (Hindi, Telugu, Tamil and Punjabi). The raters were all recruited in Mumbai, which is a large city with speakers from most Indian languages, but precisely how familiar each of the raters were with each language is unknown. Accent-familiarity, however, was not a focus of their study. Their findings suggest that standard TOEFL iBT rater training eliminates significant differences of holistic scores, though studies like Carey et al. (2011) and Winke et al. (2013) (discussed later in this chapter) suggest otherwise.

2.6.2 Kim 2009

Another study of importance is Kim's (2009), *An investigation into native and nonnative teachers' judgments of oral English performance: A mixed methods approach.* Like Xi and Mollaun (2009), Kim determined that nonnative teachers are capable of delivering ratings similar to those of native speaker teachers on semi-direct oral English tests of overall oral communicative language ability when scoring test-takers they share a common L1 with; however, the study did reveal interesting differences between the rater groups concerning how the test candidates' pronunciation was received.

The study examined how native speaking teachers of English compared with Korean English teachers when scoring ten Korean students studying English in Canada on a semi-direct interview test of spoken English. Twelve Canadian NS English teachers and twelve Korean NNS English teachers scored the interviews. The teacher participants were all qualified professionals with at least one year's experience teaching college-level English conversation to nonnative English speakers, and had completed at least one graduate degree in either a field of linguistics or in language education. All of the Korean teachers self-scored their spoken English proficiency as either advanced (n=6) or 'near native' (n=6).

The semi-direct interview test was conducted via computer, and was designed to determine the students' overall holistic language competence. The test included three tasks: 1) a picture-based task that involved test-takers providing descriptions or narrations of visual information, “such as describing the layout of a library” (p.193); 2) a situation-based task that required students to “perform the appropriate pragmatic function in a hypothetical situation, such as congratulating a friend on being admitted to school” (p.193); 3) and a topic-based task that involved test-takers sharing their opinions on a specific topic, such as, “suggesting reasons for an increase in human life expectancy” (p.194).

The teacher participants were required to score the test-takers’ performance in each task by selecting from a four-point proficiency scale, and to provide justification for the scores they administered by writing comments. The scale, similar to those of the TOEFL iBT and IELTS relied heavily on the teachers’ personal measures of listener effort to determine scores (see table 2.2). Kim explains the purpose of including comments as well as the proficiency scores was that, “they would supply not only the evaluation criteria they drew on to infer students’ oral proficiency, but that it would help to identify the construct being measured.” The teachers’ scores and comments “were compared with regard to internal consistency, severity, and evaluation criteria” (p.187). The quantitative data was analysed using the Facets MFRM program to investigate four facets: student, teacher, teacher group and task. The FACETS software allows for simultaneous analyses of various facets providing details concerning how each facet functions and contributes to the assessment setting (see Chapter 4 for a full discussion of FACETS and MFRM). The qualitative data comprised of written comments were coded into nineteen evaluative criteria and analysed for frequency comparisons across the two teacher groups.

Table 2.2: Kim’s (2009) rating scale for the oral English test

Score	Description
4	Overall communication is almost always successful; little or no listener effort is required.
3	Overall communication is generally successful; some listener

	effort is required.
2	Overall communication is less successful; more listener effort is required.
1	Overall communication is generally unsuccessful; a great deal of listener effort is required.

The findings from the MFRM analyses revealed that both NS teachers and NNS teachers maintained acceptable levels of inter-rater consistency; however, differences emerged from the qualitative data related to the two groups' attention to pronunciation. The NSs tended to be more sensitive or strict in terms of phonological accuracy than the NNSs, a finding inconsistent with some previous studies that have claimed the opposite (e.g. Brown, 1995; Fayer & Krasinsky, 1987). The NNS teachers focused more on global comprehensibility and whether or not the language was intelligible or comprehensible. This is important to this study, as it raises the connection between pronunciation and intelligibility.

What is most relevant to this study is Kim's reasoning why the focus and details of the two groups of teachers' comments differed. She suggests that the differences may have been that the Korean teachers were more familiar with the students' pronunciation due to sharing the same L1. This supports the notion of an 'interlanguage speech familiarity benefit' (Bent & Bradlow, 2003). It is unfortunate though that it is never mentioned how familiar or unfamiliar the NS teachers were with Korean-English. If the NS teachers were, in fact, familiar with Korean accented English, it may contribute to the acceptable inter-rater reliability results.

The analyses Kim conducted contributed greatly toward many of the analyses decisions for the study this thesis describes. Kim's rating scale data was analysed using Many-facets Rasch Measurement (MFRM), and the analyses of the teachers' internal consistency included examining fit statistics and expected scores and correlations of individual raters to rater population. Additionally, the teachers' severity/leniency measures were analysed according to task difficulty, and bias analyses conducted between teacher groups and tasks, and between individual teachers and tasks. MFRM is discussed and described in Chapter 4; however, this multiple analyses approach provided converging evidence that "strengthen the validity of inferences drawn about raters' internal consistency" (Kim, 2009, p.196). Kim demonstrated how MFRM can provide a wide range of insight into individual rater behavior and those of groups of raters from simple rating scale data.

Kim's study left some unanswered questions, however. It is unclear whether or not the potential bias on the part of the Korean teachers occurred due to familiarity with the students' L1 background. Without knowing how familiar with Korean-English the Canadian teachers were, it is not possible to determine how accent-familiarity might have contributed to the way they rated. The study also lacked nonnative English-speaking teachers from other L1 backgrounds, as well as students from other L1 backgrounds, so it is difficult to say whether or not the same results would have occurred concerning the Korean teachers' consistency. Such information could have strengthened or threatened Xi and Moullaun's (2009) findings. All of these remaining questions could be determined if a single study examined how multiple nonnative accents are rated by different groups of raters - both native and nonnative speakers - whose measures of familiarity with each included accent are known.

2.6.3 Huang 2013

Huang's (2013) study investigated how accent-familiarity and ESL/EFL teaching experience affects untrained raters' judgments of nonnative speech. The study included three equal sized groups of rater participants (*n*=22 per group). The groups included non-teachers familiar with Chinese, non-teachers unfamiliar with Chinese and teachers familiar with Chinese. It is not clear why teachers unfamiliar with Chinese were not included in the study. A sample of TOEFL iBT test-taker responses from 26 Chinese L1 speakers from Task Two were used for rating purposes. Task Two from the TOEFL iBT speaking test elicits spontaneous, uninterrupted speech by means of asking test-takers to state their opinions in regard to a specific question, and include their reasoning to support their answers. All raters first scored each speaker for overall proficiency using a seven-point holistic scale designed by the researcher that was later collapsed to four points during analyses, followed by scoring the speakers for three analytic dimensions of speech performance using seven point scales: foreign accent; grammar/vocabulary; and content. All of the rating scales can be described as 'native-speaker idealized', as they all include the highest rating description as "Native English Speaker" (p.774). The raters were also asked to attempt to identify each speaker's foreign accent while rating the level of foreign accent each speaker had.

The findings were mixed. Accent-familiarity was found to enable familiar raters to identify the foreign accent of the speakers better than unfamiliar raters, though the speaker participants were all Chinese. Huang states that ETS provided no additional information concerning the speakers' actual native language (e.g. Mandarin or Cantonese). Huang wanted the raters to accurately identify which Chinese language the speakers' L1 was, but there can never be more than speculation and guess concerning whether or not the raters' guesses of the speakers' accents were accurate since there was no such information provided by ETS. No significance was observed between any of the groups concerning their holistic scoring of the speakers, nor any of the analytic ratings either. These findings support the conclusions of Xi and Mollaun (2011; see also Kennedy & Trofimovich, 2008) that raters' accent-familiarity had no significant effect on scores of accentedness or comprehensibility. Additionally, Huang's conclusions suggest that EFL/ESL teaching experience has no significant effect on how untrained raters score test-takers; however, did find that raters with ESL/EFL teaching experience were, "better able to separate the analytical dimensions and were less biased by speakers' foreign accents when judging the overall proficiency or the content of their speech" (p. 770).

There are some issues worth mentioning. The raters were not completely untrained. Her descriptions are vague. She states, "In the beginning of each session, participants rated three speech files with varying degrees of proficiency as practice. Participants were given minimal guidance for the rating sessions. It is not clear how 'proficiency' was determined or what precisely the 'minimal guidance' entailed. It is also arguable that how raters' accent-familiarity might affect scores was not focused in the most logical direction – namely, pronunciation or intelligibility. I suspect this was due largely to the study attempting to examine two rater characteristics simultaneously. It is clearly logical to examine holistic measures of speech proficiency, as well as grammar/vocabulary and content when examining prior ESL/EFL experience, but such foci seem misdirected when examining the effect of accent-familiarity. It is not clear, however, why original, 'native speaker centric' rating scales were constructed rather than employ the existing TOEFL iBT holistic scales and sub topic scales. Choosing to focus on the speakers' foreign accent measure rather than more modern applications of pronunciation or 'delivery' seemed inappropriate and outdated. Finally, it was not surprising that raters familiar with Chinese accents were more capable of identifying speakers of Chinese accents, but then again all of the speakers' L1 were Chinese languages. Huang never explained her reasoning for measuring raters' ability to identify the accent of the speaker participants. If the reasoning was to determine if accurate accent identification by raters has a bias connection, MFRM analyses might have yielded that evidence, but such analyses were not included.

Huang's study is interesting though because it raises additional awareness to the notion that accent-familiarity as a rater trait is worth examining. It was Huang's hypotheses that familiarity would affect scores. It seems that her findings might have been different had she focused more on aspects of speech performance that might affect raters' speech perception due to familiarity.

2.6.4 Carey, Mannell and Dunn 2011

Perhaps the study that most influenced this present study's hypothesis of how accent-familiarity might be functioning to affect pronunciation scores on high-stakes tests of spoken English is Carey, Mannell and Dunn's (2011) *Does a rater's familiarity with a candidate's pronunciation affect the rating in oral proficiency interviews?* Their study investigated how prolonged exposure to Chinese-English, Korean-English and Indian-English affects pronunciation scores on IELTS oral proficiency interviews. The hypotheses they tested included first, that raters' interlanguage phonology familiarity differences should not cause pronunciation scores differences, and that whether a test candidate sits the test in their country of origin or elsewhere should not result in pronunciation score differences. Both hypotheses were disproved. Their results determined raters scored test-takers' pronunciation significantly higher who had prolonged exposure to speakers of that accent than raters with less or no previous exposure; interlanguage familiarity did affect pronunciation scores. Additionally, they found that NNS raters scored test-takers from their same home country higher than test-takers from other countries; where a test candidate completes their test can affect pronunciation scores due to interlanguage familiarity. Their findings support Bent and Bradlow's (2003) interlanguage speech intelligibility benefit, and provide evidence that native speakers, too are capable of developing what they call, "interlanguage phonology familiarity" (p.204). Such familiarity is acquired through prolonged exposure to the speech of English learners from a particular L1.

An important point is that their study included pronunciation scores and not only holistic scores. This research argues pronunciation is the most logical aspect of speech production assessment to likely be affected by raters' familiarity differences with particular accents. The participants in Carey et al. included 99 IELTS examiners from five different test centers: India (*n*=20); Hong Kong (*n*=20); Australia (*n* =19); New Zealand (*n* =21); Korea (*n* =19) to serve as raters. Rater characteristics collected included their age group, nationality, L1, "how many languages they spoke" (p. 206), their parents L1, and number of years' experience teaching English. The test raters, with the exception of those at the India test center, were "predominately British, Australian and New Zealander raters. A small number of North American raters were working in Hong Kong. The remainder were born in European countries" (p.206). The raters from the Indian test center were all Indian born, and 90% of the Indian raters were English L2 speakers. Most of the other raters were native English speakers (Korea 100%; Hong Kong 95%; Australia 95%; New Zealand 91%), and all of the raters were experienced English teachers with a mean time of 15.8 years' teaching experience. Bilingualism was "common for raters in all test centres, with trilingualism featuring in 10% of Indian raters and 5% of raters in New Zealand" (p. 206).

Three recordings from actual IELTS OPIs were included to serve as the stimuli for the raters to score. These recordings were conducted in Korea, China and India (one from each country). To determine the raters' familiarity with the accents of the speakers, three questions were asked:

> (1) In which countries have you taught English and for how long? (2) In which countries have you lived (but not taught English) and for how long? (3) Have you been exposed to any other particular groups of L2 speakers of English whose accents you have become used to hearing and understanding?
> (Carey et al. 2011, p. 207)

The results of the accent-familiarity questions indicated that all 99 rater participants were either 'familiar' or 'unfamiliar' with the three accents being investigated.

The analyses included logistic regression modeling to determine both the association of pronunciation scores with accent-familiarity, and to determine if the location of where the test is administered affects scores. Logistic regression models were fitted separately for each accent to determine if accent-familiarity affected pronunciation scores. 'High' pronunciation scores were determined as ≧6.0 and 'Lo' scores as ≦4.0. The findings determined that Chinese (Cantonese-English) test-takers were 2.62 times more likely to receive a high pronunciation score from raters familiar with Cantonese-English than from unfamiliar raters using a likelihood ratio test (95% CI: 1.14, 6.02; p-value: 0.022); Korean test-takers were 3.48 times more likely to receive a high score from raters familiar with Korean-English than unfamiliar (95% CI: 1.37, 8.83; p-value: 0.008); and Indian test-takers were found to be 4.62 times more likely to receive a high score from a rater familiar with Indian-English than from a rater unfamiliar with Indian-English (95% CI: 1.82, 11.75; p-value: ,0.001) (p. 210). Likelihood ratio tests also determined that the odds of a Chinese (Cantonese-English) test-taker is 9.71 times more likely to receive a high pronunciation score if sitting the test in Hong Kong than elsewhere (95% CI: 2.11, 44.66; p-value: ,0.001); Indian test-takers are 23.89 times more likely to receive a high score if sitting the test in an Indian test center than elsewhere (95% CI: 3.05, 187.26; p-value: ,0.001); and the odds for Korean test-takers receiving a high score sitting the test in Korea or elsewhere could not be determined (p-value: 0.001) (p. 211). What these findings reveal is that raters' accent familiarities do appear to affect pronunciation scoring, where the OPI is administered can affect pronunciation scores, and that the measure of effect accent-familiarity has on pronunciation scores is accent dependent; the potential effect raters' differing accent familiarities has on pronunciation scores appear to be accent specific.

As mentioned previously, their review of the relevant literature included brief discussions of the Perceptual Magnet Effect and the Exemplar Theory. These discussions were included as examples of explanations of how speech perception might occur, and offered justification for the research they conducted. However, since their study only examined raters' scores of speakers with familiar and unfamiliar accents and did not measure communicative success, neither theory could be confirmed. Such information might be gained if a single study were to examine both the pronunciation scores raters administer and also measure intelligibility between speaker and rater. Such a study would provide a greater insight into not only how raters' accent-familiarity differences affect inter-rater reliability, but also potential reasoning why differences occur. By examining both pronunciation scores and intelligibility, as the present study has, it might be possible to determine that accent-familiarity increases intelligibility, and as a result justifies higher pronunciation scores, especially when pronunciation scores are dependent on how easy or difficult it is for the rater to cope with the test candidate's speech.

2.6.5 Winke, Gass and Myford 2011

Another study of great importance concerning the potential effects raters' accent-familiarity have on high-stakes tests of spoken English is that of Winke, Gass and Myford (2011; see also Winke et al. 2012; Winke & Gass, 2013). A larger study in terms of both rater participants and speaking samples for scoring than other studies examining raters' accent familiarities' effect on test scores, the main aim of the study was to determine if raters' L2 knowledge of test-takers' L1s lead to bias. One hundred seven raters, "mostly of learners of Chinese, Korean, and Spanish" (p. i) listened to and scored selections from 432 speech samples taken from 72 native Korean, Spanish and Chinese speaker TOEFL iBT test-takers. The rating data was analysed using MFRM, and the qualitative data retrieved from stimulated recall sessions were analysed using the QSR NVivo 8 program to determine how aware the raters were of their biases. MFRM results revealed that raters with Spanish or Chinese as an L1 or L2 were significantly more lenient toward test-takers whose L1 they shared or had studied, and the qualitative analyses revealed that the raters were aware of their biases. Their findings support those of Carey et al. (2012) and Xi and Moullaun (2009) that raters' linguistic experiences pose a threat to the reliability and validity of scores on high-stakes tests of spoken English. Interestingly, like the findings from Carey et al. (2012) the effects of raters' familiarity with Korean-English failed to yield significant differences of scores.

The greatest influence of Winke et al.'s study had on the present study is their choice of MFRM analyses. The MFRM analyses provided the most robust and definitive evidence supporting the notion that accent-familiarity as a rater characteristic poses a real threat to high-stakes tests' validity and reliability. They used the FACETS computer program to examine seven facets: test-takers, the raters, the speaking tasks, test-takers' L1, the raters' "level of knowledge of their L2 (heritage speaker, more than 2 years' experience, or less than 2 years' experience" (p.21), and the raters' ESL/EFL teaching experience that was determined by either "more than 1 year vs. no experience or less than 1 year" (p.21). One of the key findings from their analyses was that they determined a group-level differential severity/leniency effect from the ratings of the three L1 subgroups of test-takers. A chi-square test of the test-taker L1 facet determined that the scores of at least two of the subgroups were statistically significantly different ($\chi^2(2)=28.6$, $p=.00$). This was surprising because the samples used in the study were provided by ETS, and according to the ETS ratings were reported to all be of the same average proficiency. The results suggest that their raters exercised dissimilar leniency/severity when rating the different groups of speakers, and rated the Spanish L1 test-takers the highest and the Chinese test-takers the lowest. T-tests conducted between the three pairings of L1 test-taker groups revealed significant differences between the ratings for the Chinese and Spanish test-takers ($t(87)=3.52$, $p=.00$), as well as between the Korean and Spanish test-takers ($t(57)=4.44$, $p=.00$). They reported no significance determined between the Chinese and Korean L1 test-takers ($t(50)=2.36$, $p=.02$), though the some may argue that the p-value could reflect significant differences of scores had their alpha level been 0.05. In any case, test-taker L1 affected the ratings from their raters. Another important finding were the results from the analyses of the raters' L2 facet. The interactions between raters sharing an L2 with the test-takers' L1 demonstrated statistically significant ($p=.01$) signs of bias concerning scores for the Chinese and Spanish L1 test-takers, though no significance was observed from raters with Korean as an L2 and Korean L1 test-takers ($p=.108$).

Winke et al.'s study represents the most concise and in-depth examination of the question as to how raters' linguistic experiences affect ratings on high-stakes tests of spoken English, but some questions remain. The effects of raters' linguistic experiences do not appear to be equal for all accents or test-taker L1 groups. For example, their findings support Carey et al.'s (2011) lack of significant evidence that raters' familiarity with Korean affects scores for Korean speakers, though familiarity with Spanish did. It is possible, since all of the rater participants in Winke et al.'s study were living in the United States at the time of participating in their study that they may have developed some measure of familiarity with the accent, and they mention it directly stating, "American students studying in Michigan most likely have more opportunities to hear Spanish-accented English than Korean- or Chinese-accented speech; thus, overall, they may be more familiar with it – which follows our theory that they then may rate it more leniently" (p.53). For that reason, it would be beneficial to recruit raters in a similar study from various locations, and determine if similar significance is still determined concerning Spanish L1 speakers' scores. Another point, as stated previously, examining scores of holistic English speaking ability seems too broad in its scope to truly determine if familiarity with test-takers' L1s has a significant effect. True, Winke et al. established that raters' sharing the same L2 as test-takers' L1 can have a significant effect, but by examining holistic scores alone it is not possible to determine what precisely about the test-takers' performance is being scored differently. Fortunately, Winke et al. included qualitative data that revealed raters' attention to accent, and their awareness of how both familiar and unfamiliar affected their score decisions. If pronunciation scores are examined in isolation, the need for qualitative data collection may not be necessary, as bias can be determined whether the rater is aware of it or not. Of course, it is interesting to know if raters are aware of their bias, but it is more important to ascertain whether there is bias than the awareness of bias. Finally, while their study duplicates actual rating experiences from the TOEFL iBT, which is useful, the results do not yield any information concerning how exactly familiarity with test-takers' L1 or accent affects raters' speech perception; it only determines how they score test-takers. A study is needed that will both examine ratings, and attempt to determine whether or not familiarity affects the speech perception of raters when processing test-takers' speech in familiar and unfamiliar accents. Measuring intelligibility could provide that important missing element.

2.7 Clarifying the terms: intelligibility, comprehensibility and interpretability

As was explained in Chapter 1, Smith and Nelson (1985) suggested that researchers should clearly define how the terms *intelligibility*, *comprehensibility* and *interpretability* are defined in their work due to fact that these words have been interpreted both inconsistently and at times interchangeably. Thirty years later, these terms continue to be sources of confusion and disagreement among researchers (Deterding & Kirkpatrick, 2006; Coetzee-Van Rooy, 2009; Rajadurai, 2007). Before describing and discussing the research and instruments designed to complete this study, it is important to provide additional clarity concerning some key terms.

In this section these three terms are each discussed and defined according to their use in this study. The discussions also include examples of how the terms have been defined and applied in different ways. This section is included here not only because intelligibility is an important focus of this study that must be explicitly defined, but because each of these terms commonly appear as constructs of high-stakes tests of spoken English. This study is primarily concerned with pronunciation and intelligibility constructs; however, it is still necessary to clearly define and distinguish how comprehensibility and interpretability differ from intelligibility. Therefore, all three terms are discussed.

2.7.1 Intelligibility

Intelligibility is a term that not all researchers can easily agree upon concerning its meaning and scope. Talia Isaacs commented (2008), “intelligibility is an evasive concept that we know little about” (p.557), and Jenkins (2000) stated, “there is yet no broad agreement on a definition of the term ‘intelligibility’: it can mean different things to different people” (p.70). What follows are explanations of some of the better-known interpretations that have emerged in the linguistics and applied linguistics literature.

Catford’s (1950) interpretation of intelligibility is perhaps the oldest, and continues to have a ripple effect in many current adaptations of the term. He considered intelligible speech to be dependent on the listener understanding the words a speaker said. ‘Understanding’ in this interpretation was heavily dependent on the listener’s ability to respond appropriately to the speaker’s utterance. This was a two-dimensional version of intelligibility in that it included both word or utterance recognition and also notions of syntactical, lexical and morphological accuracy. An important aspect of Catford’s version is that it placed the listener into the equation, and did not suggest the speaker solely determined intelligibility. This notion that intelligibility is a collaboration continues to echo in many current interpretations (e.g. Smith & Nelson, 1985; Fayer & Krasinski, 1987; Bamgbose, 1998), and is perhaps best stated by Morley (1991) who called it a “slippery concept”, and stated, “intelligibility may be as much in the mind of the listener as in the mouth of the speaker” (p.499).

Another reason it is suggested here that Catford’s interpretation of intelligibility continues to ripple in current versions is that it also included aspects of comprehensibility and interpretability. This was largely due to the matter that intelligibility was dependent on the listener responding appropriately to the utterance, which suggests that meaning and interpretation were components of intelligibility. It is arguable though that these notions of comprehension and interpretability were weaved into his pairing of intelligibility with ‘*effectiveness’* (p.7). Effectiveness, according to Catford, occurred when the appropriate response by the listener was achieved. Intelligibility, Catford explained, could be achieved without the speaker's utterance being effective, or more clearly – understood by the listener according to the speaker's intention. The following illustration explains how intelligibility does not guarantee effectiveness:

> Let us imagine, for example, a foreign guest at an English tea party. On the table there are two kinds of sweetmeats – *cakes* and *tarts*. The guest is partial to the combination of jam and pastry, and wants to obtain a tart. But his limited vocabulary does not run to more than one word for baked sweetmeats, and so he asks for a cake. His request is perfectly intelligible to his English hostess, who responds appropriately to the linguistic form by passing the plate of cakes. But the guest is confused and disappointed because his hostess's response is not appropriate to his purpose in speaking. His utterance, in other words, is *ineffective*, though intelligible.
>
> (Catford, 1950, p. 8)

Catford's notion of effectiveness as a partner to intelligibility introduced an interesting notion, namely that intelligibility did not guarantee illocutionary force, but that intelligibility was an important first step toward achieving Smith and Nelson's notions of comprehensibility (locutionary force) and interpretability (illocutionary force). Catford suggested that the only way an utterance that was not intelligible could be effective is if some other form of nonverbal message or paralanguage made the listener aware of the speaker's intention. So Catford's version of intelligibility functions as a kind of gatekeeper to comprehension and interpretability.

An early version of intelligibility that sought to distinguish it from other terms like comprehensibility and interpretability was Smith and Rafiqzad's (1979). They differentiated intelligibility from comprehension by limiting the scope of intelligibility to only word recognition from speaker to listener. Their study utilized a cloze test where listeners completed gap-fill tasks, and was among the first to measure intelligibility in terms of quantifiable success-rates. The greater the number of accurate gap-fill items a listener could transcribe, the greater the intelligibility between speaker and listener occurred. Kenworthy (1987), too, stated, "the more words a listener is able to identify accurately when said by a particular speaker, the more intelligible that speaker is" (p.13). This interpretation is important to this research, as it was the first to limit the term to only the ability of a listener to determine the phonological content of a speaker, and places intelligibility apart from meaning (locution) or intentions of the speaker (illocutionary force). In this study, however, Kenworthy and Smith and Rafiqzad's (1979) notion that intelligibility is quantifiable through accurate transcriptions of a speaker's utterance does not suggest any sense of universal intelligibility of a speaker; intelligibility is only determined on individual bases occurring between one speaker and one listener.

Perhaps the best-known work on the topics of intelligibility, comprehensibility and interpretability is Smith and Nelson's (1985) paper. It was the first research to consider the problems inherent in the three terms and how they had been (and continue to be) used interchangeably. The work they did was a summary of the various interpretations of the three terms. Concerning intelligibility, their definition echoed Smith and Rafiqzad's (1979), and provided the most directly worded and without elaboration explanation. They defined intelligibility as simply, "word/utterance recognition" (p.334). Also like Smith and Nelson, they isolated intelligibility from notions of locutionary and illocutionary force with comprehensibility and interpretability respectively. They also astutely noted that intelligibility is not centered on either the speaker or the listener, but is interactional between the two participants in any communication exchange. When this interpretation is applied in a language testing construct, the responsibility and potential for success in a speaking task must not solely be the test-takers'. It demands that raters share a potentially equal role in the outcomes or scores. And like test-takers, raters can be varied in their abilities, too.

The reason intelligibility and the other terms continue to be interpreted in numerous ways and at times interchangeably, is that researchers adapt them to suit their needs. Rajadurai (2007) suggests that, "the lack of consistency in intelligibility studies and their findings may be attributed to differences in definitions, methodologies, and samples used as well as variables investigated" (p. 89). If the research demands intelligibility includes the listener understanding the meaning of the utterance or intention of the speaker, there are precedents of such use to choose from, or the researcher may simply invent new definitions.

When intelligibility includes aspects of locutionary and/or illocutionary force the distinction between intelligibility, comprehensibility and interpretability, as defined by Smith and Nelson (1985), is reduced or lost. Such an interpretation is that of Munro, Derwing and Morton's (2006) (also in Derwing & Munro, 1997; Munro & Derwing, 1995, 1999) that suggests, "Intelligibility is the extent to which a speaker's utterance is actually understood" (p. 112). Interestingly, Derwing and Munro (1997) suggest that their interpretation was not unlike Smith's (1992) that focused on word and utterance recognition (p.2), but this connection seems dubious. What is dubious is the phrasing "actually understood" that implies locutionary and/or illocutionary force. Bamgbose (1998) expanded this notion of a broader interpretation of intelligibility stating it is "a complex of factors comprising recognizing an expression, knowing its meaning, and knowing what that meaning signifies in the sociocultural context" (p. 11). Kachru (1986) and Rajadurai (2007) argue that intelligibility must include considerations of context such as topic, the situation of the exchange and participants. Like Smith and Nelson (1985) explained, it does not seem prudent to attempt to warehouse so many concepts and measurements of performance and production into one term or aspect. Intelligibility defined and distinct from comprehension, comprehensibility, interpretability and understanding, seems the most efficient and reliable means to address this term and avoid confusion especially for testing purposes.

There is one aspect of Smith and Nelson's (1985) notion of intelligibility that this research disagrees with. Namely, they suggest that intelligibility, or word/utterance recognition, is potentially less important than comprehensibility or interpretability. This analysis of their view of intelligibility is derived from their comment, "intelligibility, comprehensibility and interpretability are not equally weighted . . . the most serious misunderstandings occur at the level of comprehensibility and interpretability" (p.335). This notion suggests intelligibility is somehow easier than negotiating comprehensibility or interpretability. They aptly note that a listener can possibly recognize and identify the lexical content of an utterance, but that knowledge of context and other factors influence locutionary and illocutionary force of utterances. This may be true, but similar to Jenkins (2000), I believe intelligibility functions as a gatekeeper of sorts to successful transmission of locutionary and/or illocutionary force. Intelligibility is not a guarantee for successful comprehension or interpretability, but it is the best first step toward such success.

Jenkins (2000) compliments Smith and Nelson's (1985) (also Field, 2005) notion of intelligibility adding it is, "the production and recognition of the formal properties of words and utterances and, in particular, the ability to produce and receive phonological form" (Jenkins 2000, p.78). What distinguishes Jenkins' interpretation from the other interpretations suggesting intelligibility is determined by accurate word/utterance transcription is that she considers intelligibility from the viewpoint of its occurring between English speakers from different L1 backgrounds. This distinction is worth mentioning because it represents the continuance of the notion of intelligibility being both collaborative between speaker and listener, and that it is primarily concerned with successful word/utterance identification from speaker to listener.

Field's (2005) interpretation, that this research applies also limits intelligibility to only the listener's ability to correctly transcribe the phonetic content, or words, a speaker says.

> Intelligibility is measured by the ability of judges to transcribe the actual words of an utterance, comprehensibility by an overall rating of how easy it is to understand a given speaker . . . This article consequently restricts the term *intelligibility* to features of the speech signal. As used here, it refers to the extent to which the acoustic-phonetic content of the message is recognizable by a listener.
>
> (Field, 2005, p.401-2)

This interpretation of intelligibility does not include aspects of speakers' locutionary or illocutionary forces that could make measurement unreliable for the reasons previously described. Success or failure of intelligibility is determined solely through transcription task success. This application is suitable for this study because it narrows the scope of the term appropriately to a single, measurable aspect of oral communication between speaker and listener, or test-taker and rater.

Future researchers will no doubt make their own arguments for either continuing to apply any or all of the already numerous interpretations of intelligibility, or proceed to invent new ones. No doubt the interpretation applied in this research will be scrutinized and criticized as flawed or even possibly preposterous. Such interpretations are expected considering the already wide range of interpretations applied to intelligibility that researchers have to choose from. For this reason, it is worthwhile to echo Smith and Nelson's appeal that researchers clearly define their use of these terms.

2.7.2 Comprehensibility

Comprehensibility, like intelligibility, is also a term whose meaning researchers cannot easily agree about. Possibly the best effort to provide a definition as clear and unambiguous as possible is Smith and Nelson's (1985) attempt stating it is, "word/utterance meaning (locutionary force)" (p. 334). Gass and Varonis' (1984) formula of comprehensibility, described previously in this chapter, is perhaps the most complicated explanation, or non-explanation, of the term. Oddly, the manner they chose to measure comprehensibility with in their study was through transcription accuracy, so it is puzzling as to why they constructed such a convoluted formula for comprehensibility. Gass and Varonis (1984) could have defined the term as simply as Smith and Nelson's (1985) intelligibility definition as, "word/utterance recognition" (p.334). Other interpretations of the term seem to fall between the simplicity of Smith and Nelson and the complexity of Gass and Varonis.

The interpretation of comprehensibility this study adheres to is that of Derwing and Munro (2009) who state:

> We define COMPREHENSIBILITY as the listener's perception of how easy or difficult it is to understand a given speech sample. This dimension is a judgment of difficulty and not a measure of how much actually gets understood. Our research shows that comprehensibility ratings

correspond to the amount of time, or the effort it takes to process utterances, even if they are perfectly understood in the end.

(Derwing & Munro, 2009; p. 478)

This explanation of comprehensibility was selected both because it fits with the tasks the raters in this research completed, and because it compliments Field's (2005) notion of intelligibility adopted in this research. Field (2005) provided a less descriptive version of comprehensibility stating it is, "an overall rating of how easy it is to understand a given speaker" (p.400). Kennedy and Trofomovich (2008), also interpreted comprehensibility as "*perceptions* of how easily they understand an utterance (p. 461). Derwing and Munro's (2009) version of the term applied in this study places comprehensibility more outside of the notions of locutionary or illocutionary force than Field's or Kennedy and Trofomovich. The interpretation applied in this study emphasizes 'listener effort' as the determining factor. This definition more closely matches the rater task of scoring pronunciation applied in this study and many high-stakes tests of spoken English that rely on raters' listener effort to determine scores.

Derwing and Munro's (2009) definition of comprehensibility is devoid of locutionary force, whereas Field's (2005) and Kennedy and Trofomovich's (2008) are inclusive. Interestingly, Derwing and Munro (2009) decided to interpret 'intelligibility' in their paper as "the degree of a listener's actual comprehension of an utterance" (p.479). All three studies, however, included transcription tasks to determine intelligibility, and each addressed comprehensibility similarly by means of raters scoring speakers using a rating scale. Transcription tasks were also used in all three studies to determine intelligibility. It is interesting how all three included similar tasks, but each defined comprehensibility differently. Field's definition is closest to Smith and Nelson's implying locutionary force as a component of the construct. However, 'understanding' as a component of the construct was not really measured in Field's study. Only the difficulty the raters experienced were reported by means of a rating scale. It can only be presumed that the raters' scores are accurate reflections of their effort needed to 'understand' the speakers' utterances, but it cannot be confirmed. The interpretation of 'comprehensibility' departed further from Smith and

Nelson's definition by Derwing and Munro (2009). In their interpretation they eliminated locutionary force entirely from comprehensibility. Though they made no mentioning of their reasoning for this peculiarity of their interpretation, it is possible it was due to locutionary force being difficult (or impossible) to measure by means of a rating scale.

2.7.3 Interpretability

The term interpretability has not had the same kind of variety in its use or application as intelligibility and comprehensibility have by researchers, though it is still deserving of discussion and clarification. This study adheres to the most straightforward and simple definition of interpretability available: Smith and Nelson's (1985) definition, "meaning behind word/utterance (illocutionary force)" (p. 334). Smith and Nelson's definition of interpretability is very similar to Catford's (1950) notion of 'effectiveness' that was described as, "it is normally the speaker's intention that the hearer should respond to his utterance in a manner which is appropriate to his purpose in speaking" (p.7). Smith and Nelson's interpretability is also similar to Kenworthy's (1987) 'communication' that "involves reading the other's intention" (p. 16). Clearly, Catford and Kenworthy were distinguishing illocutionary force from locutionary force. 'Interpretability' is dependent on appropriate reactions of utterances (see also James', 1998 definition of 'communicativity'; p.217).

In semi-direct (and direct) speaking tests, illocutionary force plays a vital role: test-takers must respond appropriately to test prompts; however, the interpretability of test candidates' speech by raters is not measured. Perhaps, it cannot be measured. Consider again Catford's observation that speech can be intelligible but still not effective, and how this was established by means of the listener mistaking a cake for a tart. In the case of raters of semi-direct tests of spoken English there is no system that confirms whether or not raters accurately determined the illocutionary force of test candidates' utterances; it is not possible for the listener to illicit either the correct or incorrect response, so interpretability of test-candidates' utterances cannot reliably be determined through rating. Like Derwing and Munro (2009) stated, “how do you measure how much someone has understood?” (p.479).

2.7.4 Observations of the varied use of the terms

The meaning (locutionary force) of an utterance or intentions of a speaker (illocutionary force) can never be completely confirmed using only a rating scale. In face-to-face tests like the IELTS OPI it is possible for an interlocutor to confirm the locutionary and/or illocutionary forces of a test candidate’s utterances, but it is not possible in semi-direct tests like the TOEFL iBT. From the TOEFL iBT rater perspective the locutionary and illocutionary force of test candidates’ utterances can only be speculated but never confirmed. Likewise, the intelligibility of test candidates’ speech from the rater perspective is elusive. It is possible (and demonstrated in this research) that a rater may believe they heard one word when, in fact, the speaker uttered a different word. Raters are never asked to transcribe test-takers’ utterances to measure intelligibility in actual test rating procedures.

There are currently no high-stakes, semi-direct tests of spoken English that verify whether or not raters have accurately determined the locutionary force (or illocutionary force) of test candidates’ speech. Semi-direct tests of spoken English, as discussed earlier in this chapter, have been widely considered to have stronger reliability than direct speaking tests because of the greater consistency of test delivery (e.g. Huges, 1989; Lazaraton, 1996; O’loughlin, 2001). While delivery may be more consistent in semi-direct tests because they do not use interlocutors to deliver task prompts, it is not possible to suggest that semi-direct tests are more reliably scored. As has been demonstrated through the discussions of these terms, measuring *intelligibility, comprehensibility* and *interpretability* are difficult in the best of circumstances, and arguably unlikely in other testing circumstances.

Clearly researchers feel the meanings of these terms remain open to interpretation. It seems that researchers’ decisions concerning how they interpret the terms have more to do with how and what they hope to measure (or not measure) than any attempt to agree with one or more previous definitions. Locutionary force and illocutionary force are complicated constructs. It is arguable that the reason Smith and Nelson (1985) were able to define the three terms so simply was that their definitions were not being applied to a research project in that paper. It is arguable that they were drafted not to describe test or fluency constructs, but simply for the sake of proving they could be defined simply. Test and fluency constructs, it seems, are far more complicated, and it is likely these terms will remain flexible and inconsistent in their interpretations. For this study Derwing and Munro's (2009) adaptation of comprehensibility, paired with Field's (2005) version of intelligibility work well together. They describe how raters decipher the lexical content of phonetic messages from the speaker participants (intelligibility), and also determine how raters measure the difficulty they experience deciphering the lexical content of the speakers' utterances (comprehensibility). Interpretability is not included in the scope of what this research project aimed to measure or determine, so is not discussed further in this thesis.

In the next chapter, the pilot study is presented. The study examines how raters' familiarity with Japanese-English affected pronunciation scores and intelligibility. It was the first attempt to test the hypotheses set forth by this study.

Chapter 3 : The pilot study

Overview

This chapter describes the first test, or pilot study, conducted in this research. Its appearance here in this chapter prior to the methodology chapter was decided due to the impact this early study had on the final design of the main study. By presenting it here unnecessary redundancy in the next chapter can hopefully be avoided, and offers a more chronological account of the evolution of the research instruments and analyses. This pilot study was first described in Browne and Fulcher (2016), though the word limitations of the book chapter prevented a full discussion of the study. What follows is the full description and explanation of that test, the subsequent findings and discussion of its merits and limitations.

From May to July 2012 the study was conducted. It was designed to attempt to determine if raters' familiarity with Japanese accented English (Japanese-English) causes statistically significant differences in pronunciation scores on high-stakes semi direct tests of spoken English. The test also measured the intelligibility success-rates between speaker and rater, a first for any study, to determine if the theory that intelligibility differences due to accent-familiarity could be a contributing a factor of score variances observed in previous studies (i.e. Carey et al., 2011; Winke, et al., 2012). If 'listener effort' determines pronunciation score selection in high-stakes tests, measuring intelligibility differences as well as the pronunciation scores could provide a deeper insight into the potential threat of raters with differing measures of familiarity with test-takers' L1 affected English accents.

The findings revealed raters' familiarity with Japanese-English significantly affected pronunciation scores. The greater the familiarity with Japanese-English, the more lenient the raters were. It was also determined that raters' familiarity with Japanese-English significantly affected intelligibility, and likewise, the greater the familiarity the more intelligible the speakers were to the raters. Significant correlations between accent-familiarity level and both leniency and intelligibility were also revealed.

The pilot was designed as a semi-direct test, and utilized an online survey-host, Survey Monkey, to deliver the test. Similar with other semi-direct test scoring procedures, no personal information about the test-takers, such as gender, age, race, or nationality were revealed to the rater-participants. The study was deemed largely a success, and contributed to the development of the main research instrument described in Chapter 4. What follows are the details of this pilot study.

3.1 The test

The test was comprised of three main parts. Japanese-English was selected as the nonnative accent to be examined for the pilot. It was considered better to focus on only one accent in the first study to determine the effectiveness of the design, and make any necessary changes before examining other accents. The test was designed to be completed in no longer than twenty minutes. The reasoning for this was that since rater participation was voluntary and without compensation, if the test required more than twenty minutes to complete most potential rater-participants would not complete the entire test. Overall, the test design proved to be effective, and yielded valuable insight into the effects of raters' accent-familiarity with Japanese-English.

Part one of the test collected biographical, professional and other necessary information about the rater-participants' linguistic experiences. Raters answered questions about their L1, home country, country they were currently residing, previous ESL/EFL teaching experience and familiarity level with Japanese-English. (See Appendix A for a copy of the entire test).

Part two of the test was the main part of the test, and was divided into six sub-sections- one for each speaker-participant. In each sub-section the raters were instructed to first listen to a recording of a speaker reading two sentences, complete intelligibility gap-fill items for that recording by completing an incomplete transcript of the sentences, and then score that speaker's pronunciation with the provided rating scale (see Table 3.1). Like the TOEFL iBT rating scales for delivery, this test's scales required raters to estimate the degree of 'listener effort' they experienced interpreting the phonological content of each speaker.

Table 3.1: The pronunciation rating scale

Score	Description
5	Speech is generally clear and requires little or no listener effort. Only one listening required.
4	Speech is generally clear, with some fluidity of expression, but it exhibits minor difficulties with pronunciation and may require some listener effort at times. Only one listening required.
3	Speech is clear at times, though it exhibits problems with pronunciation and so may require more listener effort. It was necessary to listen more than once before attempting to complete the gap fill.
2	Consistent pronunciation difficulties cause considerable listener effort throughout the sample. It was necessary to listen more than once before attempting to complete the gap fill.
1	Cannot comprehend at all.

The recordings were accessed via embedded videos in the test. Raters had control over the recordings in that they could start, stop or replay the videos at their discretion. The videos did not include moving or still images of the speakers, so no additional paralanguage stimuli were included that could affect scores. Each sub-section also included opportunities for raters to comment about the speaker or experiences using the research instrument.

The gap-fill tasks for each speaker included four or five missing words from each recording consisting of a total of 28 intelligibility items. Intelligibility success was determined as either successful or unsuccessful based on accuracy. No partial credit was given though spelling errors were not penalized.

Perhaps the most unique aspect of this study's instrument was the sentences the speaker-participants read. The sentences were largely conceived based on the original BKB-R sentence lists (Bench, Kowald & Bamford, 1979), which were designed to measure the degree of sensorineural hearing loss in partial-hearing children (see Appendix A for examples from the original BKB-R lists). According to the American Speech-language-hearing association, sensorineural hearing loss "reduces the ability to hear faint sounds. Even when speech is loud enough to hear, it may still be unclear or sound muffled"[10]. In short, it affects a listener's perception of the phonetic content of speech. The measure of loss an individual has can be tested through intelligibility, or word/utterance identification and discrimination tasks like in the BKB-R test (more on this topic is discussed in Chapter 4). Accented speech, too, can adversely affect the ability for a listener to find speech intelligible (Flege, 1984; Munro, 2008). This study theorizes that accent-familiarity enhances the listener's ability to process the phonological content of speech in that accent, and may be measured in a similar way as sensorineural hearing loss can be measured. This instrument measures the perceptual benefit of accent-familiarity in a similar way the BKB-R sentences test measures loss of speech perception. More is discussed in Chapter 4 about the sentence structure ideology using the BKB-R sentences as a guide.

The sentences constructed for this study shared a uniformity of length and lexical level, though they differed from the BKB-R sentences, as their contexts were intentionally complex. As discussed previously, Gass and Varonis (1984) determined that knowledge of context is the greatest contributor to comprehension. It was reasoned that if the test sentences contained 'real world knowledge' contexts, raters could possibly deduce the correct gap-fill target word(s) based on logical lexical choices inferred from the identifiable contexts of the sentences. This application of context knowledge for task completion, it was reasoned, would adversely affect the data. This study aimed to measure how raters cope with accented speech, and their ability to decipher the phonological content of accented speech. If raters were able to employ deductive reasoning from context knowledge to make answer choices, the ability to isolate raters' speech perception based on pronunciation alone would be diminished. With this method, raters had to be able to cope with the pronunciation of the speaker alone to complete the gap-fill tasks successfully.

To eliminate context familiarity as an influence to task completion accuracy, one or more of the original gap-fill words from the BKB-R sentences were replaced with words that resulted in contexts that could be deemed, 'possible but not probable.'

[10] Taken from the American Speech-Language-Hearing Association webpage for Sensorineural Hearing Loss. http://www.asha.org/public/hearing/Sensorineural-Hearing-Loss/

In other words, the sentences were not nonsensical. The replacement word choices were not randomized; rather, the original word(s) was first located on the *JACET 8000* (2003) list of the 8,000 most frequently used English words by Japanese speakers of English, and identified for word class (e.g. noun, verb, adjective, etc.). The nearest word on the list sharing the same word-class as the original word that also fit the restrictions of 'possible but not probable' was selected from the list as the replacement. Preference was given to words that were ranked with higher frequency of use than the original to maintain consistency of lexical difficulty to the original. This method of word replacement served four functions: 1) it was employed to eliminate the predictability of the items for the gap-fill. 2) by employing the JACET 8000 to locate the original and then use the nearest word that fit the criteria for selection, it reduced the researcher's manipulation of the sentences to a more randomized method of choosing replacement words. 3) it operated to keep the selected words manageable for the speaker-participants to recognize and read aloud without the assistance or explanation from the researcher concerning their 'correct pronunciation.' The lexical level of the sentences was all within the 3,000 most frequently used words Japanese speakers of English use. 4) by controlling the lexical level of the sentences guaranteed that the rater-participants, too, should be familiar with the contents of the sentences if they are capable of finding the speakers intelligible. The resulting sentences were syntactically accurate though contextually complex or unpredictable (see Table 3.2).

Table 3.2: The sentences included in the test with the intelligibility items underlined

Speaker 1	They had a <u>tiny</u> <u>day</u>.
	The old <u>soaps</u> are <u>dirty</u>.
Speaker 2	They are <u>paying</u> some <u>bread</u>.
	The <u>play</u> had nine <u>rooms</u>.
Speaker 3	The institution <u>organism</u> was <u>wet</u>.
	The <u>dog</u> made an <u>angry</u> <u>reader</u>.
Speaker 4	The <u>ladder</u> is <u>across</u> the <u>door</u>.
	He <u>cut</u> his <u>skill</u>.
Speaker 5	The <u>union</u> cut some <u>onions</u>.
	She <u>sensed</u> <u>with</u> her <u>knife</u>.
Speaker 6	<u>Mine</u> <u>took</u> the money.
	The <u>matches</u> <u>lie</u> on the <u>infant.</u>

The sentences were proofread by three university English teachers prior to use for recordings. All teachers reported that the contexts of the sentences were unpredictable but possible, and that any English language teacher of any level should be familiar with the lexical content. Additionally, discussions were held concerning the potential for features of Japanese-English that could be deemed *typical* of Japanese-English pronunciation from use of the sentences. Elements of problematic Japanese-English phonology incorporated in the test included /r/ - /l/ distinction (see Riney & Flege, 1998), the lax vowels /I/, /ʊ/, /ʌ/ and /ə/, and the voiced dental fricative /ð/ (see Carruthers, 2006 for a complete discussion of pronunciation difficulties of Japanese speakers of English). Apart from the instructions and explanations delivered in the test, raters were provided with no additional training that sought to normalize the scores they administered.

3.2 The participants

3.2.1 The speaker-participants

The study included six speaker-participants. Five native Japanese university students studying English as non-English majors at two Japanese universities (Waseda University, male n=2; Tsukuba University, male n=1; female n=2) provided the Japanese-English samples. A native speaker, the researcher – an American male from the Southern United States (New Orleans, Louisiana), Speaker A, was included in order to attempt to serve as a 'highly intelligible speaker.' It was reasoned that since the rater-participants received no formal training that the raters might more easily grasp the tasks and function of the design if the first speaker they encountered was *easily intelligible*. In particular, it was reasoned by presenting to the rater-participants a speaker they could easily cope with first that the rater-participants might better understand the remaining tasks they were being asked to perform.

3.2.2 The rater-participants

Eighty-seven volunteer rater-participants comprised of ESL/EFL teachers and researchers of linguistics and applied linguistics were recruited from sixteen countries via email invitations (see table 3.3). It was decided that since the TOEFL iBT also uses nonnative speakers of English as raters (e.g. Xi & Mollaun, 2009), and is a test that "measures the ability of nonnative English speakers to use and understand the English language as it is heard, spoken, read and written in the university classroom"[11] the use of such qualified participants from any racial or cultural background could qualify as rater-participants. It did not seem prudent to exclude qualified, nonnative speakers of English that are English language teachers and/or researchers from participating. Among the rater-participants were university EFL/ESL teachers (n=48), Business English teachers working in companies (n=4), High School or Junior High EFL/ESL teachers (n=11), elementary school EFL/ESL teachers (n=6), private English-School EFL/ESL teachers (n=12) and graduate students enrolled in Applied Linguistics or TESOL MA and PhD programs (n=37). Thirty-six participants reported they were both working as EFL/ESL teachers and enrolled in graduate programs at the time of participating. Seventy-three rater-participants were native English speakers and nine were nonnative speakers; four of the nonnative speakers were native Japanese speakers. Since the study aimed to determine the affect familiarity with Japanese-English has on pronunciation scores and intelligibility success rates, raters self-reported their level of familiarity with Japanese-English according to the following four-level scale:

1. No Familiarity. (n=13)
2. Limited Familiarity – You have heard Japanese speakers of English but without regularity, and/or have not had Japanese students during the last two years. (n=32)

[11] The answer for "What is the TOEFL iBT test?" Taken from the Frequently Asked Questions About the TOEFL iBT Test web page. Retrieved December, 9, 2014 from https://www.ets.org/toefl/ibt/faq/

3. Some Familiarity- You have spent at least the last two years with students from Japan, have visited Japan and/or regularly watch TV or movies in Japanese. (n=4)
4. Very Familiar- You are a native speaker of Japanese, have lived in Japan for one or more years, and/or studied the Japanese language for 1 or more years. (n=38)

Table 3.3: Rater-participants' Home Country List

United Kingdom	35
USA	24
Canada	7
South Africa	4
Japan	4
Australia	3
Brazil, France, Jamaica, Libya, Malta, Spain, St. Lucia, Sudan, Syria, Ukraine	1(per country)
Total	87

3.4 Results and discussion

In this section the analyses procedures and results are presented. Discussions of the findings are included throughout this section in order to provide a clear narrative of the findings. The pronunciation scores' data and results are discussed first, followed by the intelligibility task results and analyses.

The quantitative data collected in the study were analysed using Facets 3.71, a Many Facets Rasch Measurement (MFRM) software and SPSS (version 20) data analysis software. MFRM was conducted in order to examine both the pronunciation scores and intelligibility data. In the MFRM analyses only the five Japanese-English speakers' data were included (the native speaker, Speaker A was excluded), as the analyses were meant to examine only how the raters coped with Japanese-English speakers and how familiarity differences with Japanese-English might result in differences of scores and intelligibility. The inclusion of the data for Speaker A, the

English NS, would have skewed the the results. Speaker A was, however included in the other analyses.

The pronunciation score data and intelligibility task data were analysed separately in Facets due to the differences of tasks the rater-participants were required to perform. Pronunciation scoring is clearly what can be considered a 'rater task', but since raters in high-stakes tests of spoken English are not required to complete intelligibility tasks, it was determined that such a task more reflected a 'test candidate task', and so were separated. The Facets program is capable of calculating the analyses of both tasks, but since the different tasks measured such different dimensions of the rater-participants the analyses were calculated separately. Linacre (personal correspondence) recommended that separate analyses for the two tasks was the most appropriate means to understanding how raters' familiarity with Japanese-English affects pronunciation scores, and also affects intelligibility success rates.

3.4.1 Pronunciation scores

The analyses of the pronunciation data examined three facets: the raters, the speakers and the raters' familiarity level of Japanese-English included as a grouping facet[12]. The results support the findings of earlier studies that raters' familiarity with speakers' accents could have a significant effect on pronunciation scores (e.g. Carey et al., 2011; Winke et al., 2011).

The most informative and important piece of output from Facets analyses is the variable map shown in Figure 3.1. It summarizes the key information of each facet in a single figure highlighting the results of rater behavior, speaker performance and familiarity grouping. The scale utilizes measurements in terms of "logits" that reflect probability estimates in an equal-interval scale. The figure is separated into five vertical columns (reading from left to right):

[12] Grouping facets are data that are only used for grouping purposes in the Facets software, and do not affect test related outcomes.

1. Column one displays the logit scale. The scale ranges from -7 to 2. The scale provides a reference for measurements of all other columns. The measure 0 represents even likelihood, or 50-50 odds of prediction.
2. The second column displays each rater's total scoring allotment of the five speakers. The numbers displayed in the column represent each rater-participant's identification number. The raters at the top of the column were the most lenient, and raters at the bottom most severe in their scoring.
3. The third column shows the results of the raters separated into groups according to their reported familiarity levels displaying how high or low each group scored the five speakers. Like column two it reflects the level of severity the rater groups demonstrated scoring the speakers. Raters reporting very familiar were most lenient, and raters with no familiarity the most severe.
4. Column four represents the overall performance of each speaker according to the grades they received from the raters. Speaker E, at the top of the column scored the highest, and is the most capable speaker according to the data; likewise, Speaker F, at the bottom of the column, scored lowest, and was determined to be the least capable speaker.
5. The fifth column displays the five-point scale that was implemented to score the speakers' pronunciation in the test. The broken horizontal lines between the numbers on the scale represent the measurable point where the likelihood of receiving a higher or lower score exists according to the analysis.

```
+-----+----------------------------------+--------------------------------------+----------+-----+
|Measr|+rater                            |+Familiarity Level                    |+Speaker  |Scale|
|     |   (Most lenient - top)           |    (Most lenient - top)              |(most     |     |
|     |                                  |                                      |capable-top)    |
+-----+----------------------------------+--------------------------------------+----------+-----+
+   2 +                                  +                                      +          + (5) +
|     |                                  |                                      |          |     |
|     |                                  |                                      | Speaker E| --- |
|     | 14  48                           |                                      |          |     |
+   1 + 50  63  65                       +                                      + Speaker C+     +
|     | 25  3                            |                                      |          |     |
|     | 23  61                           | Very Familiar                        |          |     |
|     | 19  2   24  26  37  62  85       |                                      |          |  3  |
*   0 * 17  55                           * Limited Familiarity  Some Familiarity*          *     *
|     | 13  18  34  40  6   66  74  9    | No Familiarity                       | Speaker B|     |
|     | 38  54                           |                                      |          |     |
|     | 1   15  21  22  29  31  44  52  8|                                      | Speaker D|     |
+  -1 + 35  36  56  68  80               +                                      +          +     +
|     | 58                               |                                      | Speaker F| --- |
|     | 30  32  4   5   51  71  72       |                                      |          |     |
|     | 12  39  43  60  78  81           |                                      |          |     |
|     | 20  49  76                       |                                      |          |     |
+  -2 + 11  16  27  28  75  77  82       +                                      +          +     +
|     | 42  7   79                       |                                      |          |     |
|     | 10  41  45  46  86               |                                      |          |     |
|     | 67                               |                                      |          |     |
|     | 47  53  57  64                   |                                      |          |  2  |
+  -3 +                                  +                                      +          +     +
|     | 33  84  87                       |                                      |          |     |
|     | 70                               |                                      |          |     |
|     | 59                               |                                      |          |     |
+  -4 +                                  +                                      +          +     +
|     |                                  |                                      |          | --- |
|     | 83                               |                                      |          |     |
+  -5 + 69                               +                                      +          +     +
|     |                                  |                                      |          |     |
|     |                                  |                                      |          |     |
+  -6 +                                  +                                      +          +     +
|     | 73                               |                                      |          |     |
|     |                                  |                                      |          |     |
+  -7 +                                  +                                      +          + (1) +
+-----+----------------------------------+--------------------------------------+----------+-----+
|Measr|+rater                            |+Familiarity Level                    |+Speaker  |Scale|
+-----+----------------------------------+--------------------------------------+----------+-----+
```

Figure 3.1: Facets Variable Map of Pronunciation Scores including Four Levels of Familiarity

To fully appreciate the simplicity of the variable map, one needs to understand the function of the Logit scale in column 1, and how it is used to measure the likelihood of how raters or groups of raters will score the different speakers. For example, rater 50 is located at 1 on the Logit scale and rater 69 is at -5 on the Logit scale. If raters 50 and 69 were to score Speaker F, the scale predicts that rater 50 would score the speaker's pronunciation at 3 using the test's rating scale, but rater 69 would give a score of 1. This is calculated by measuring the distance from 0 on the Logit scale to the placement of the rater; if the number is positive that same distance will be added to the Speaker's position in column 4 to determine the prediction of column 5, and if the rater's Logit value is a negative value, the distance from 0 to the rater's placement in column 2 is subtracted from the Speaker's position in column 4 to determine the scale's prediction for column 5. It is necessary to understand the function of the Logit scale to effectively appreciate and understand the benefit of the variable map. The scale can be used to measure the likelihood of how raters or groups of raters will score different test-candidates.

One of the research questions concerning the raters was whether or not the raters would act as independent experts. This question was included because the lack of rater training could be argued as reasoning to dismiss the results of the study. The 'rater agreement statistics' Facets provides revealed a slightly higher percentage of actual agreements (2399; 37.9%) than expected agreements (2338; 37%). These results suggest that though the raters had no official training, they functioned largely like independent experts, and performed according to what the Rasch model expects (see Green, 2013).

Facets also calculated a rater separation index of 1.49 (G), which can be applied to the formula (4G+1)/3 to determine a strata index of 2.32 that determines the number of strata of severity. This suggests that there were between two and three levels of severity demonstrated by the raters concerning pronunciation scoring. These calculations also included a reliability measure of the separation index to be .69. A reliability score of 1.0 would suggest that all raters demonstrated non-interchangeable severity, and a score of zero that all of the raters had interchangeable severity. The .69 further suggests that variance in rater severity occurred, which answers the research question concerning whether or not the raters would demonstrate equal measures of rater severity.

The Facets results include 'In-Fit' measures that determine how raters (or items) performed within the model. These results can be used by test management to identify individual raters that do not *fit* the Rasch model. Since this research predicted variance in individual raters' severity due to potential affects of accent-familiarity, no rater-participants were removed from the sample due to their individual infit results.

The fit statistics concerning the grouping facet were considered to determine if each group of raters fit the Rasch model. The results revealed that all four groups' In Fit and Out Fit statistics reflected acceptable values (0.5~1.5) (see Green, 2013 p. 219). Table 3.4 shows the results of the grouping facet measures. Notice that as familiarity level increases, the infit mean square values drop until the 'very familiar' group suddenly rises and threatens to not fit. These varying fit statistics, while not significant do suggest familiarity was affecting the 'fit'. These results suggest differences in rater-group severity also occurred.

Table 3.4: Pronunciation score Facets rater familiarity level group measures

Familiarity Level	Total Score	Obsvd Avrg	Measure in Logits	Model S.E.	Infit MnSq	ZStd
No	144	2.22	-0.38	0.22	0.91	-0.4
Some	48	2.4	-0.09	0.38	0.71	-0.9
Limited	388	2.47	0.03	0.14	0.92	-0.6
Very	526	2.77	0.43	0.12	1.09	0.9
Mean	276.5	2.46	0	0.21	0.91	-0.3
S.D.	190	0.2	0.29	0.1	0.14	0.7

In order to determine whether or not significant variance of pronunciation scores occurred between groups of raters divided according to accent-familiarity levels, the results of Pearson's chi-square tests, included in the Facets output were reviewed. The results determined significant differences in leniency scoring pronunciation occurred between the four groups of raters ($\chi^2(3)=12.3$, $p=.01$). This result confirmed that raters' familiarity differences with Japanese-English did result in significant differences of pronunciation scores.

To complement the MFRM findings, the significance of group differences between the four familiarity groups of raters were analysed by conducting One-way ANOVA using SPSS. Prior to conducting the ANOVA, the data were analysed also using SPSS to determine if the distribution was normal. Shapiro-Wilk results revealed the data were not normally distributed, and histograms showed the data for all but speaker D were negatively skewed; Speaker D was positively skewed. Though the data were not normally distributed, results from parametric tests like ANOVA and t-tests maintain their reliability if sample sizes are large enough as in this thesis. Ghasemi and Zahediasl (2012) explain that when sample sizes are larger than 100 "we can ignore the distribution of the data" (p.486). Norman (2010) clarifies this point stating:

> we learn that "parametric tests are based on the assumption of normality''. Regrettably, we forget the last part of the sentence. For the standard t tests ANOVAs, and so on, it is the assumption of normality of the distribution of

means, not of the data. The Central Limit Theorem shows that, for sample sizes greater than 5 or 10 per group, the means are approximately normally distributed regardless of the original distribution.

(Norman, 2010; p.628)

The ANOVA results are shown in Table 3.5, and reveal that Speakers B, D and E received statistically significant differences of scores ($p \leq .05$). The effect sizes reflected in the Eta Square values demonstrate that the effect was large for Speakers B and D, and medium for Speaker E. Raters' differing levels of familiarity with Japanese-English did result in significant differences of pronunciation scores, though not for all of the speaker participants.

Table 3.5: ANOVA results of four familiarity groups' pronunciation scores

Spkr.		Sum of Squares	*df*	Mean Square	*F*	*p*	Eta Squared
B	Between Groups	11.54	3	3.85	6.77	0.000	0.232
	Within Groups	47.18	83	0.57			
	Total	58.71	86				
C	Between Groups	1.31	3	0.44	0.51	0.674	0.027
	Within Groups	70.28	83	0.85			
	Total	71.59	86				
D	Between Groups	7.82	3	2.61	6.79	0.000	0.230
	Within Groups	31.48	82	0.38			
	Total	39.3	85				
E	Between Groups	8.66	3	2.89	4.90	0.004	0.167
	Within Groups	47.76	81	0.59			
	Total	56.42	84				
F	Between Groups	0.09	3	0.03	0.07	0.975	0.004
	Within Groups	33.52	83	0.40			
	Total	33.61	86				

The results of the Chi-square and ANOVA confirmed that significantly different pronunciation scores occurred between the four groups, but one hypothesis of this study is that raters 'very familiar' will demonstrate the most leniency. To test this hypothesis, the other three familiarity level rater sub-groups were collapsed into one new subgroup "less than 'very familiar' with Japanese-English", and independent t-tests were conducted to determine if significant differences of pronunciation scores occurred. Table 3.6 shows the results of the tests. The results reveal that the 'very familiar' raters scored three of the speakers significantly higher than all other raters. The Cohen's *d* results for those three speakers can be used to calculate the common language effect sizes (CL), or *probability of superiority,* that determine that there is between a 70-80% chance that the same speakers would receive higher scores if a randomly selected rater participant from the 'very familiar' raters were chosen to rate them than if a randomly selected rater participant with 'less familiarity' conducted the ratings. CL are calculated from Choehn's *d* using the following formula where Φ is the cumulative distribution function and the population's Cohen's *d* is δ (Ruscio, 2008).

$$\mathrm{CL} = \Phi\left(\frac{\delta}{\sqrt{2}}\right)$$

Though the CL results are results are affected by the distribution of the data, the combination of results suggest that the hypothesis that raters 'very familiar' with Japanese-English will score Japanese-English speakers higher than raters with less familiarity appears to be plausible.

Table 3.6: Independent t-test results of 'very familiar' and all other raters' pronunciation scores

	"Very familiar" with Japanese-English			Not "very familiar" with Japanese-English						
Spkr.	*N*	*Mean*	*S.D.*	*N*	*Mean*	*S.D.*	*t*	*df*	*d*	
B	38	2.76	0.85	49	2.10	0.68	3.90	69.8	0.857	**
C	38	2.97	0.94	49	2.90	0.90	0.38	85.0	0.076	
D	38	2.61	0.64	48	2.02	0.60	4.33	77.2	0.951	**
E	38	3.39	0.68	47	2.83	0.84	3.35	83.0	0.733	**
F	38	2.11	0.69	49	2.14	0.58	-0.28	85.0	-0.047	

*Note: ** = p ≤ .001*

The results of the pronunciation data analyses revealed that raters' differences of accent familiarity with Japanese-English did result in significant differences in pronunciation scores. What the results did not determine was why accent familiarity appears to have had such an effect on pronunciation scoring. The next section may provide an answer why score variance occurred.

3.4.2 Intelligibility

In this section the results of the analyses of the intelligibility data are presented and discussed. Similar to the pronunciation score data, the intelligibility data was also analysed examining two facets and one grouping facet: the raters, the gap-fill items and raters' reported level of familiarity with Japanese-English respectively. Results from MFRM and other analyses revealed significant differences of intelligibility occurred between raters with different levels of accent-familiarity, suggesting that increased accent-familiarity leads to increased intelligibility success rates, at least concerning Japanese-English. First discussed are the MFRM results followed by additional findings determined from other analyses that reveal the impact accent-familiarity had on intelligibility.

The Facets Variable Map shown in Figure 3.2 offers a visual representation of the Facets results. The content of each column is as follows:

1. Column one displays the logit scale ranging from -4 to 6. With dichotomous data, as in this intelligibility analysis, the logit measures provide success-probability-calculations corresponding to item difficulty calibration, and rater, or rater-group, ability measurements (see Appendix A for a Logit-to-probability conversion table).
2. Column two shows how the individual raters performed in the intelligibility gap-fill exercises.
3. Column three displays how the different rater groups performed on the gap-fill exercises when grouped according to their self-reported measure of familiarity with Japanese-English.
4. Column four provides a visual representation of the difficulty of each intelligibility gap-fill item included in the test. Each item is labeled first by the speaker who delivered the word and the target word. Items at the top of

the column were the most difficult, and items at the bottom of the column were easiest. It is interesting that all five speakers delivered both easier and more difficult items for transcription. As with the pronunciation data, Speaker A was not included in the analysis.

```
+------------------------------------------------------------------------------------------------------------------------+
|Measr|+rater                          |+Familiarity Level   |+Item                                                      |
|     |Most Capable                    |Most Capable         |Most difficult                                             |
|-----+--------------------------------+---------------------+-----------------------------------------------------------|
|   4 +                                +                     + Speaker C angry     Speaker E knife                        |
|     |                                |                     | Speaker C dog                                             |
|     | 24                             |                     |                                                           |
|     |                                |                     |                                                           |
|     |                                |                     |                                                           |
|   3 +                                +                     + Speaker D door      Speaker E onions                       |
|     | 19                             |                     | Speaker E with                                            |
|     |                                |                     |                                                           |
|     |                                |                     |                                                           |
|     | 3  6  18 48 55                 |                     | Speaker E union                                           |
|   2 + 54                             +                     +                                                           |
|     | 63                             |                     |                                                           |
|     | 4  14 62 85                    |                     |                                                           |
|     | 35 50 74                       |                     | Speaker B rooms                                           |
|     | 2  15 16 30 31 37 52 72 82     |                     | Speaker C reader    Speaker C wet                         |
|   1 + 45 47 59 65 68                 +                     + Speaker F took                                            |
|     | 1  5  25 29 51                 |                     |                                                           |
|     | 12 40 78                       |                     |                                                           |
|     | 11 21 23 64 67 71 77 79 81     | Very Familiar       |                                                           |
|     | 38 46                          |                     | Speaker F matches                                         |
*   0 * 10 13 17 26 39 41 42 57 66 84 * Some Familiarity    * Speaker B bread                                           *
|     | 32 34 83                       | Limited Familiarity | Speaker C organism  Speaker D across                      |
|     | 7  36 49 56 58 60 70 80 86     | No Familiarity      | Speaker F lie                                             |
|     | 8  22 28 53 73 87              |                     | Speaker B play      Speaker D cut       Speaker E sensed  |
|     | 61 75                          |                     |                                                           |
|  -1 + 43                             +                     +                                                           |
|     | 27 33 44                       |                     |                                                           |
|     | 9  69 76                       |                     |                                                           |
|     |                                |                     |                                                           |
|     | 20                             |                     | Speaker D ladder                                          |
|  -2 +                                +                     +                                                           |
|     |                                |                     |                                                           |
|     |                                |                     |                                                           |
|     |                                |                     |                                                           |
|     |                                |                     |                                                           |
|  -3 +                                +                     + Speaker D skill                                           |
|     |                                |                     |                                                           |
|     |                                |                     | Speaker F infant                                          |
|     |                                |                     |                                                           |
|     |                                |                     | Speaker F mine                                            |
|  -4 +                                +                     +                                                           |
|     |                                |                     |                                                           |
|     |                                |                     |                                                           |
|     |                                |                     |                                                           |
|     |                                |                     |                                                           |
|  -5 +                                +                     + Speaker B paying                                          |
|     |                                |                     |                                                           |
|     |                                |                     |                                                           |
|     |                                |                     |                                                           |
|     |                                |                     |                                                           |
|  -6 +                                +                     +                                                           |
|-----+--------------------------------+---------------------+-----------------------------------------------------------|
|Measr|+rater                          |+Familiarity Level   |+Item                                                      |
|     |Least Capable                   |Least Capable        |Least Difficult                                            |
+------------------------------------------------------------------------------------------------------------------------+
```

Figure 3.2: Facets Variable Map of Intelligibility Gap-fill Outcomes Including Four Levels of Familiarity

The MFRM results of the intelligibility data support the hypothesis that the more familiarity with Japanese-English a rater has, the greater success that rater will have finding Japanese-English speakers intelligible. This was determined by Chi-square results that revealed a significant difference in overall intelligibility occurred between at least two of the four accent-familiarity rater sub-groups ($\chi^2(3)=26.6$, $p=.00$). Table 3.7 shows additional detail from the Facets analyses of the accent-familiarity rater sub-groups' performance. The more accent-familiarity each rater group had, the greater the overall intelligibility-success occurred. Raters 'very familiar' with Japanese-English were twenty percent more successful than the raters with no familiarity, which is evidenced by the differences in observed average scores.

Table 3.7: Facets accent-familiarity level measurement report for intelligibility items

Familiarity level	Total Score	Total Count	Obsvd Avrg	Measure in Logits	Model S.E.	Infit MnSq	ZStf
No	156	312	0.50	-0.41	0.16	0.99	0.0
Limited	435	768	0.57	-0.13	0.10	0.91	-1.7
Some	59	96	0.61	0.07	0.30	0.84	-1.0
Very	634	912	0.70	0.46	0.10	1.08	1.4
Mean	321.0	522	0.59	0.00	0.17	0.96	-0.4
S.D.	227.4	331	0.07	0.32	0.08	0.09	1.2

To compliment the chi-square, ANOVA tests examining how the four accent related rater sub-groups differed on each intelligibility item were conducted using SPSS. Again, the data were not normally distributed. Table 3.8 shows the eight items that resulted in significant intelligibility variance between rater-subgroups from the ANOVA tests. Here we can see the first potential evidence suggesting why pronunciation score variance was observed for Speakers B, D and E between rater-subgroups in Table 3.5. The results in Table 3.8 also indicate significant differences of intelligibility occurred with the same speakers, as well as with Speaker F. The three items from Speaker D in the table represent all items from the first sentence the speaker uttered. The effect sizes of the items from Speaker D were large, and further suggest that familiarity level significantly affected intelligibility. Such similarities between the ANOVA results of the pronunciation scores and intelligibility items could be used to make the argument that intelligibility differences between rater-subgroups may account for the pronunciation score variance observed. Nevertheless, the findings of significant differences of intelligibility with one third of the total test items suggests accent familiarity had a strong impact on intelligibility.

Table 3.8: Results of ANOVA tests conducted of the Japanese-English intelligibility items

Item	Spkr.		Sum of Squares	*df*	Mean Square	*F*	*p*	Eta Squared
paying	B	Between Groups	6.41	3	2.14	4.53	0.005	0.133
		Within Groups	41.92	89	0.47			
		Total	48.32	92				
rooms	B	Between Groups	1.59	3	0.53	3.64	0.016	0.109
		Within Groups	12.93	89	0.15			
		Total	14.52	92				
ladder	D	Between Groups	7.64	3	2.55	19.65	0.000	0.398
		Within Groups	11.53	89	0.13			
		Total	19.16	92				
across	D	Between Groups	5.32	3	1.77	8.91	0.000	0.231
		Within Groups	17.71	89	0.20			
		Total	23.03	92				
door	D	Between Groups	0.55	3	0.18	4.97	0.003	0.144
		Within Groups	3.28	89	0.04			
		Total	3.83	92				
sensed	E	Between Groups	2.39	3	0.80	3.44	0.020	0.104
		Within Groups	20.64	89	0.23			
		Total	23.03	92				
matches	F	Between Groups	4.05	3	1.35	6.75	0.000	0.185
		Within Groups	17.78	89	0.20			
		Total	21.83	92				
lie	F	Between Groups	2.35	3	0.78	3.34	0.023	0.101
		Within Groups	20.88	89	0.24			
		Total	23.23	92				

In order to determine whether or not the 'very familiar' rater sub-group had significantly better intelligibility-success rates than all other raters, independent t-tests were conducted with each intelligibility item. Table 3.9 shows the results, and reveals that the 'very familiar' raters were significantly better transcribing eight of the

24 items. These results also indicate that accent familiarity affects intelligibility, and answers the research question concerning whether or not the 'very familiar' raters demonstrate a superior ability finding Japanese-English speakers' utterances intelligible.

Table 3.9: Significant results from independent t-tests measuring the intelligibility differences between the raters 'very familiar' with Japanese-English and all other raters

		"Very familiar" with Japanese-English			Not "very familiar" with Japanese-English						
Item	Spkr.	*N*	*Mean*	*S.D.*	*N*	*Mean*	*S.D.*	*t*	*df*	*d*	
paying	B	38	0.74	0.45	49	0.27	0.45	3.36	69.7	1.04	**
rooms	B	38	0.95	0.23	49	0.71	0.46	2.85	84.0	0.66	*
ladder	D	38	0.58	0.50	49	0.04	0.20	6.99	49.8	1.42	**
across	D	38	0.82	0.39	49	0.35	0.48	5.02	90.7	1.07	**
door	D	38	1.00	0.00	49	0.92	0.28	2.06	51.0	0.40	*
union	E	38	0.97	0.16	49	0.83	0.37	2.55	70.9	0.49	*
onions	E	38	1.00	0.00	49	0.92	0.28	2.33	51.0	0.40	*
sensed	E	38	0.63	0.49	49	0.35	0.48	2.34	91.0	0.58	*
matches	F	38	0.82	0.39	49	0.47	0.50	4.01	90.9	0.78	**

*Note: *= $p \leq .05$, ** = $p \leq .001$*

Now that the findings from both the pronunciation and intelligibility data sets have been presented and discussed can the question of whether or not significant correlations exist between accent familiarity levels with Japanese-English and how raters score Japanese-English speakers' pronunciation and find their speech intelligible. Using SPSS Pearson's correlations were calculated using the raters' reported familiarity level and their mean pronunciation scores for the five Japanese-English speakers, as well as their mean intelligibility scores shown in Tables 3.10-11. These data were graphed into two scatterplots shown in Figures 3.3-4. SPSS calculated the regression lines, and reveal patterns suggesting that as familiarity level increases, so too, do intelligibility success rates and higher pronunciations scores. Though Figure 3.3 shows a modest 15% of shared variance (reflected in the R_2value 0.15) between pronunciation scores and level of familiarity, such an influence can make a significant impact on individual scores. The shared variance shown in Figure

3.4 of 31% (R_2=.313) indicates a potentially large impact of accent-familiarity level on intelligibility. The hypothesis that correlations exist between accent-familiarity level and pronunciation scores and intelligibility is not rejected, at least concerning familiarity with Japanese-English.

Table 3.10: Pearson's correlation results measuring familiarity level with Japanese-English and pronunciation score

		Familiarity with Japanese-English (1:No;2:Limited;3:Some; 4:Very)	Raters' mean pronunciation scores for Japanese-English speakers
Familiarity with Japanese-English (1:No;2:Limited;3:Some;4:	Pearson Correlation	1	.388**
	Sig. (2-tailed)		0.000
	N	87	87
	R squared		0.151

** Correlation is significant at the 0.01 level (2-tailed).

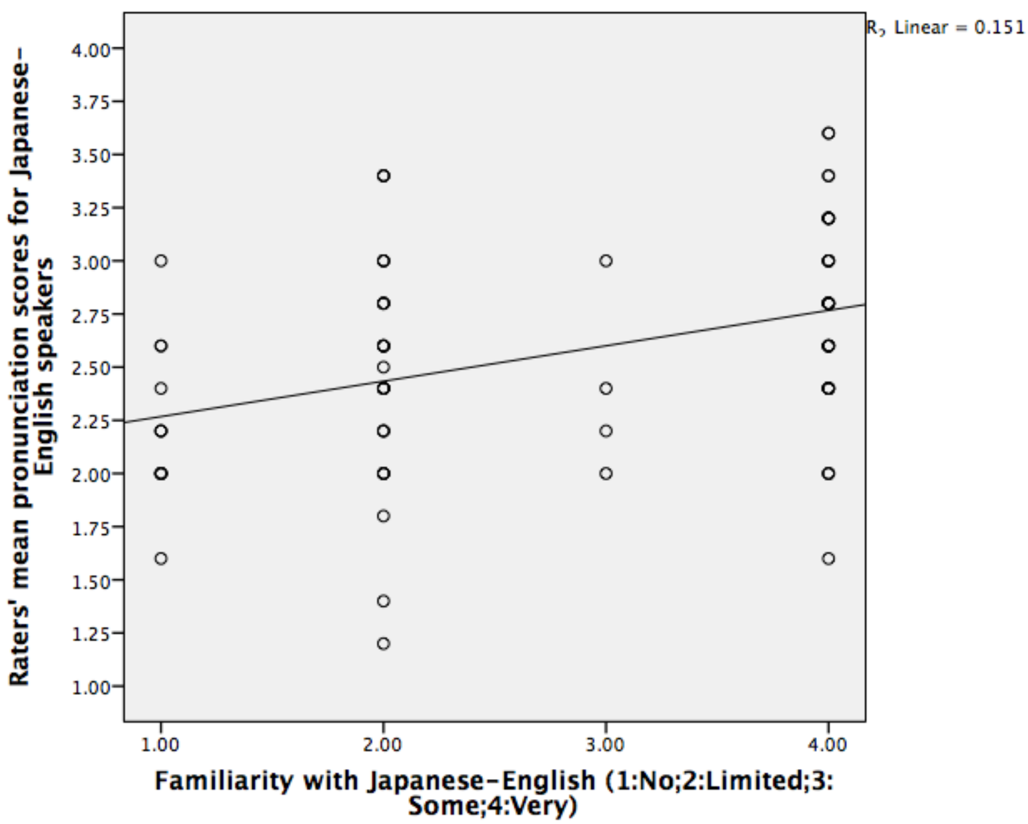


Figure 3.3: Scatter plot with regression line showing the correlation between raters' familiarity level with Japanese-English and how they score those speakers' pronunciation

Table 3.11: Pearson's correlation results measuring familiarity level with Japanese-English and intelligibility success

		Familiarity with Japanese-English (1:No;2:Limited;3:Some; 4:Very)	Raters' mean intelligibility scores coping with Japanese-English
Familiarity with Japanese-English (1:No;2:Limited;3:Some;4:	Pearson Correlation	1	.559**
	Sig. (2-tailed)		0
	N	87	87
	R squared		0.313
** Correlation is significant at the 0.01 level (2-tailed).			

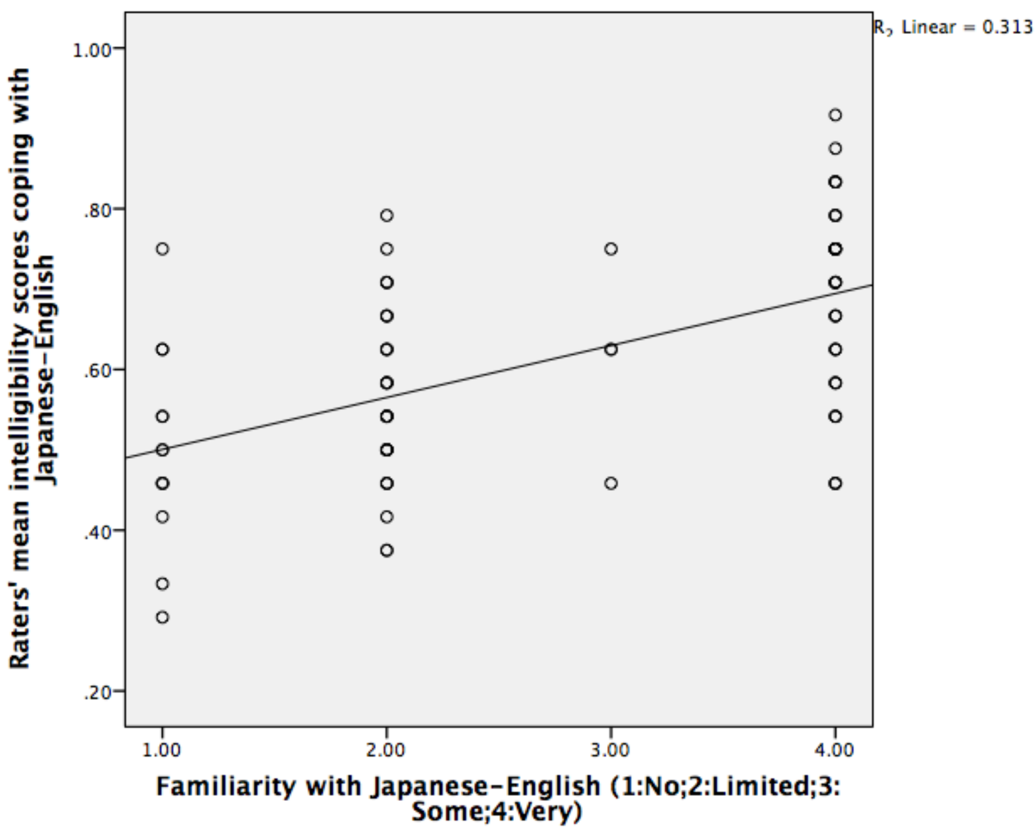


Figure 3.4: Scatter plot with regression line showing the correlation between raters' familiarity level with Japanese-English and intelligibility success with the accent

3.5 Conclusions from the pilot study

The research instrument functioned well, and the outcomes determined that raters' differing familiarities with Japanese-English can result in statistically significant differences in pronunciation scores and intelligibility success rates. Both pronunciation scores and intelligibility success rates were observed to be significantly affected by accent-familiarity, and provide evidence that raters' accent-familiarity differences are a real threat to the validity and reliability of pronunciation scores of high-stakes tests of English. Since this study focused on only one non-native accent it could not determine if familiarity with Japanese-English leads to bias when scoring Japanese-English speakers compared to speakers of other accents. It did, however, through examining how intelligibility success rates differ between familiarity levels, offer the first potential reasoning based on empirical evidence as to why score variance may occur due to raters' accent-familiarity differences. It was therefore decided that the larger, main study would include three different accents. By doing so, it would be possible to simultaneously evaluate the effects raters' familiarity levels have on pronunciation scoring, intelligibility success rates and also determine if familiarity leads to bias.

The Facets results revealed interesting points concerning accent-familiarity's effect on scores. The agreement statistics suggested that the raters did act like independent experts, and not like *rating machines*. These agreement statistics would likely be considered ideal from actual high-stakes test management, yet there were significant discrepancies observed when the data was examined dividing the raters according to accent familiarity levels. In a larger dataset with speakers of numerous accents, it is reasonable that the noise in the data could effectively obscure the threat to scores; however, the evidence is plain to see here. The more familiar the rater groups were with Japanese-English, the more lenient they were toward scoring Japanese-English speakers, and likewise had intelligibility-success coping with the accent.

It is worth mentioning that the pronunciation data analyses revealed that three of the five Japanese-English speakers received statistically significant differences of scores by raters with differing levels of familiarity with Japanese-English but not all of them. In order to attempt to determine if a reason or reasons for this difference existed, additional investigation and analyses were conducted. An independent t-test of the collapsed familiarity groupings ('familiar' and 'unfamiliar') determined that both rater groups scored Speaker C almost identically ($t(85) = -.02, p = .981$) (see table 6). This could be due to the fact that both rater groups found her equally intelligible. As mentioned previously, one intelligibility gap fill item from Speaker C was answered correctly by all participants, one item of hers was entered incorrectly by only two raters and the only intelligibility item that did not favor increased familiarity was also spoken by Speaker C. While Speaker C represents an example of a Japanese-English speaker whose pronunciation is not scored with significant difference by raters familiar or unfamiliar with Japanese-English, the intelligibility data results suggest that there is a connection between intelligibility and pronunciation scoring. Without including intelligibility in the study it could only be supposed as to why Speaker C did not receive significantly different scores; however, a link between intelligibility and pronunciation scoring was observed, and provides a clearer indication as to how familiarity manifests in a measurable way leading to significant differences in pronunciation scoring. Speaker F's pronunciation scores also failed to show significant differences due to familiarity, but intelligibility success rates demonstrated that Speaker F was more intelligible the more familiarity raters had with Japanese-English.

The qualitative data collected from the rater comments provide a possible reasoning concerning why no significant differences in scores occurred for Speaker F. One rater very familiar with Japanese-English commented that Speaker F, "did not sound like a Japanese speaker of English." Since all speaker-participants were former students of the researcher it was known prior to recruitment that Speaker F had spent one year studying English intensively in the Philippines. I, too felt that Speaker F's pronunciation was somewhat *different* than most Japanese-English speakers, but did not predict this would result in any differences with the other included speaker-participants. Of course, only one rater making a comment suggesting the speaker did not represent a stereotypical Japanese-English speaker is not enough evidence to dismiss the lack of statistically significant differences in pronunciation scores, but it may be worth future investigation of how prolonged intensive study of English in Expanding Circle Countries may result in hybridized or unique accents that reduce, alter or eliminate familiarity effects on pronunciation scores in future research.

Ten of the twenty-two lexical items in the intelligibility gap-fill tasks with the Japanese-English speakers included voiced alveolar liquid (/l/; n=4) and voiced palatal liquid (/r/; n=6), which are perhaps the most commonly known points of difficulty for both distinction and production by Japanese-English speakers (Carruthers, 2006). It is suspected that raters familiar with Japanese-English pronunciation did happen to have what Johnson (2005) described as 'exemplars' at a speaker-group-related level, or phonological memories of sound structures, that were activated while engaged in the task of speech perception, and that they contributed to the significant differences in intelligibility success rates. Additionally, these exemplars could also be the reason why the pronunciation scores showed significant differences due to the 'perceptual magnet effect' that functions to decrease the experienced degree of difficulty decoding phonological input when presented with familiar patterns of pronunciation.

For the purposes of the larger study, it was decided that the sentences should be constructed to include typical features of the included accents that represent typical pronunciation peculiarities or difficulties. By doing so it could be possible to examine if the 'exemplar' and 'perceptual magnet' theories apply resulting in increased intelligibility and reduced levels of difficulty interpreting speakers, thus resulting in higher pronunciation scores. For that reason, the larger study's sentence construction design differs from this pilot's shifting away from the more randomized method of using replacement words based only on proximity on the JACET 8000 of the original word to selecting replacement words that featured typical pronunciation production and perception difficulties predetermined for each targeted accent. This pilot instrument could have possibly been better had it featured more examples of typical elements of Japanese-English pronunciation deemed problematic. Nevertheless, the results demonstrate that raters familiar with Japanese-English tended to find speakers of the accent more intelligible and scored them higher than raters less familiar with the accent.

Though the sentences were all deemed “possible but not probable” prior to use in the recordings, the qualitative data revealed sixteen instances of rater-participants failing to understand the instructions that the sentences were *prescribed* by the researcher, and were only read by the speakers. The speech was not representative of spontaneous speech, and the sentences’ contexts were intentionally complex. Nevertheless, some raters complained they could not ‘understand’ the sentences. As a result, it was decided the main study test would include repeated instructions throughout that the sentences the speaker-participants read were not examples of spontaneous speech, and the samples should be scored only for the quality of the pronunciation alone according to the provided scale.

The pilot study functioned well according to design for time of completion. The calculated average time for completion was 20 minutes, 50 seconds after one outlier was removed from the completion time data that was more than ten hours. In total five participants had the test open for more than one hour, but it is supposed that these participants were either interrupted while completing the test or left the test open for a period while not engaged in the test between starting and completion. As a result, the construction of the larger study would need to find ways to further reduce the task-completion time in order to increase the number of speaking participants yet maintain an average completion time under twenty-five minutes. A t-test was conducted to determine if nonnative speaker raters (n=9 nonnative speaker raters; n=73 native speaker raters) required significantly different amounts of time to complete the test compared with native speaker raters. No significance was reported, so it was deemed that the test design did not favor native speaker raters for difficulty of task or time required to complete the test.

Concerning the phrasing of the pronunciation scoring rubrics used in the pilot, it was determined through rater comments that the term “comprehension”, though defined in the instrument, caused some confusion by being included in the pronunciation rating scale for the lowest score choice, “1. Cannot comprehend at all” would be changed to, “1. The speaker’s pronunciation is unintelligible” in the main study. Since pronunciation scoring is based on raters employing their measure of difficulty deciphering the phonetic content of the sentences read by the speakers, the term “comprehension” is agreeably problematic for use in the rating scales, as it evokes in many people the inclusion of locutionary force, illocutionary force and interpretability, which were not construct relevant to this instrument.

Japanese-English is arguably a well-known World English, and this pilot study determined raters’ accent-familiarity level differences with Japanese-English poses a potential threat to the validity and reliability of pronunciation scores when scoring Japanese test-takers. It was therefore decided that the main study should include two other well-known accents and one accent that is not very globally well known. Rater training could never realistically include specific instruction for every possible accent, so by including an *uncommon accent* we might gain important insight into how raters’ accent familiarities not only cause score variance, but that the variance is accent-specific, and some accents stand to benefit more than others as a result of their being *better-known accents*.

The findings from this study provide an adequate amount of evidence that raters familiar with Japanese-English tended to score Japanese-English speakers more leniently for pronunciation than raters with less or no familiarity with Japanese-English. The results also revealed that increased accent-familiarity with Japanese-English positively affected intelligibility with the rater participants ‘very familiar’ with Japanese-English experiencing the greatest success. For these reasons, Japanese-English is not included in the larger, main study, though the researcher does not suggest that additional research of raters’ familiarity level differences with Japanese-English as a threat to pronunciation scores is not necessary or warranted.

Chapter 4 : Methodology

In this chapter, the instrument designed to address the research questions is discussed and described. This study argues that judgments of pronunciation in high-stakes tests of spoken English must consider the role of the rater and their familiarity with test candidates' accents as a facet of the pronunciation construct. The success or failure of communication is not solely determined by the performance of the speaker, but must include the abilities of the listener, or in the case of testing – the rater, as an active participant. Even in semi-direct tests, the rater is not entirely passive. The test this chapter describes was designed to accomplish two main goals. First, the test would measure how raters with differing levels of familiarity with three nonnative English accents score speakers of those accents for pronunciation. Specifically, it would attempt to determine if raters' accent familiarities cause construct-irrelevant, rater-dependent variance of scores. Second, for the first time in a study that is concerned with raters' differing accent familiarities affecting test scores, the test described here measured differences in intelligibility between the speakers and the raters. By investigating both the pronunciation scores and intelligibility simultaneously, it was reasoned a deeper understanding of how accent-familiarity affects speech perception as it relates to pronunciation scoring could be realized.

This chapter is divided into seven main sections. Section 4.1 discusses how the hypothesis that raters' differing familiarities, and levels of familiarities, with various English accents might have an unintended, construct irrelevant impact on pronunciation scores developed. Section 4.2 explains how the nonnative accents investigated in the study were selected. Sections 4.3 and 4.4 introduce the speaker participants and rater participants respectively. Section 4.5 describes the test and its three main parts, and section 4.6 explains the different analyses performed.

4.1 The development of the hypothesis and what prompted this study

The hypothesis that raters' differing accent-familiarity levels affects pronunciation scores developed through more than ten years' professional experience of assessing students' spoken English abilities as an ESL lecturer in Japan. This section, delivered in the first-person provides the personal, professional background to this study. It accounts how this hypothesis was formulated, and is included here because these experiences contributed to how the test to determine the hypothesis developed.

From 2007 to 2011 I worked as a senior lecturer of English at Ritsumeikan Asia Pacific University (APU), an international, dual language university in Oita prefecture, Japan. Approximately half of the 6,000 students at APU were from more than 90 foreign countries and regions. This experience provided me teaching and assessment opportunities with students speaking in accents that were both familiar and unfamiliar to me. Among my duties were assessing the students' overall speaking abilities including pronunciation. In this capacity, difficulties attempting to deduce or develop an unbiased means of measuring pronunciation performance occurred early on. The problem stemmed from self-awareness that the ten years of daily interactions with Japanese-English had caused any level of Japanese-English learner to be easily intelligible. In short, I had 'learned' Japanese-English pronunciation, its idiosyncrasies, common errors, and had developed the means to easily decode it.

Speech processing of Japanese-English had become as easy as processing speech from native speakers from my home country. Chapter 2 introduced the Exemplar Theory (Johnson, 1997; 2003) and Perceptual Magnet Effect (Iverson & Kuhl, 1994; 2000), which both describe precisely my personal experiences and knowledge of Japanese-English at that (and the present) time. The realization that Japanese-English was personally easily intelligible did not obscure the fact that many of my Japanese students' pronunciation was by no means *perfect* or *ideal*. I was still able to distinguish students that were 'better' or 'worse' than others, though all were easily intelligible.

Difficulty began when being required to score students' pronunciation whose accents were unfamiliar to me. Particular problems occurred, for example, coping with speakers of Mongolian-English and Vietnamese-English (to name two). Another realization that compounded the problem was that many of these same students' writing and listening abilities, knowledge of grammar, syntax and vocabulary were comparatively superior to many of the Japanese students whose pronunciation was intelligible to me. My initial reaction was that these Mongolian and Vietnamese students, while more proficient in all other aspects of English proficiency, were deficient in their pronunciation; however, after critical reflection of my role as an equal participant to any communicative act as listener, the problem this research addresses became evident. I realized if I had spent ten or more years in Vietnam or Mongolia it would be likely that the same accent-familiarity benefit I was experiencing with Japanese-English would occur with their accents.

At that time it was APU policy that students' pronunciation be scored by means of a rating scale that relied on the teacher's personal measure of difficulty processing the speech of each student. This approach seemed inappropriate due to my understanding that Japanese-English was personally easier to process than any other nonnative accent. This awareness prompted an investigation to determine a better method. The first sources I investigated were the rating scales for 'delivery' from the TOEFL iBT and the IELTS pronunciation score bands and descriptors. I reasoned that since these tests are administered globally, and employ raters with varied linguistic experiences, the test developers must have considered and accounted for the problem I was encountering. As has been explained, the TOEFL iBT and IELTS also employ rating scales that rely on rater-effort to determine pronunciation scores. I realized then that the problem that I was experiencing was likely also occurring with pronunciation scores on the IELTS and TOEFL iBT. Namely, raters are scoring candidates for pronunciation that are speaking in accents both familiar and unfamiliar to them, and their personal accent-familiarities are affecting the scores. Rather than find a solution to the problem I was having, I uncovered a much larger problem – the validity and reliability of scores for pronunciation on high-stakes tests of English like the TOEFL iBT and IELTS were being threatened by raters' differing accent familiarities.

At the time of this revelation and realization (2007) there were no previous Language Testing studies concerned with this potential threat to high-stakes tests of spoken English. In the years since this theory was first considered, as was described in Chapter 2, other researchers have begun to examine raters' accent-familiarity differences as a threat to speaking tests (e.g. Carey et al., 2011; Winke et al., 2013; Xi &

Molluan, 2009). The instruments, findings and limitations from these earlier studies were of great importance to the development of the focus and design of this study. However, no previous study had considered how accent-familiarity might affect intelligibility. The other studies had only confirmed that accent-familiarity is a potential threat to tests. My personal experiences determined that intelligibility was involved. This led to the conclusion that if both pronunciation scores and intelligibility were recorded and measured, they might explain why score differences occur. If the reason accent-familiarity affects scores can be determined, test developers might be able to address and eliminate it. It is from these experiences and conclusions that the design of this study was inspired.

4.2 Non-native accent selection

Three nonnative English accents were selected for examination in this main study: Spanish-English, Arabic-English and Dhivehi-English. In this section of the chapter the method and motivations used to select these accents are revealed. Following this introduction to the section, insights gained from the pilot study concerning accents and familiarity are described. This is followed by descriptions and discussions of each of the three accents. Details concerning the phonetic inventories of each accent are also included, as these features were vital to the construction of the scripts written for the speaker participants of each accent (described later in this chapter). Finally, a brief discussion of certain limitations encountered when making the accent selections is included.

This research supports the notion of "foreign accent" as described by Wells (1982, p.1) where nonnative English speakers' pronunciation patterns, "reflect many of the phonological and phonetic characteristics of their mother tongue". Concerning the use of the terms 'Spanish-English', 'Arabic-English' and 'Dhivehi-English', these are not intended to suggest any notions of total homogeneity or absolute uniformity. Rather, these terms are used simply to refer to the associated L1 of each speaker participant. With the exception of Dhivehi, both Spanish and Arabic are languages spoken as first languages in numerous countries, and are capable of producing their own distinctive accent characteristics. For this reason, as much as possible the speaker participants recruited for this study included individuals from a variety of different regions where their respective L1s are spoken. This was done in order to provide a wider range of potential pronunciation pattern differences, which might better represent each selected accent.

Japanese-English was not included in the final study. Described and discussed in Chapter 3, it was decided that the evidence determined in the pilot study was sufficient to establish that a significant difference in pronunciation scores occurred between raters with differing levels of accent-familiarity ($p<.01$). Additionally, it was determined that ten of the 24 intelligibility items (seven items $p\leq.001$; three items $p\leq.05$) were significantly more intelligible to raters ‘very familiar’ with the accent than those ‘unfamiliar’. Since these findings were clear, Japanese-English was not included.

The general aim of the non-native accent selection process was to select four accents (one for the pilot study, and three for the main test) that were: 1. geographically distant to each other; 2. from distinctly different language families; and 3. included a range of *global familiarity*.

Global familiarity is an original concept of this study. The concept attempts to predict or estimate the potentiality of familiarity with a nonnative accent based on three factors: 1) number of speakers of the accent’s L1; 2) geographic coverage of the accent’s L1; and 3) media/news coverage of the accent and/or L1. It was considered important to choose accents that provided a range of Global Familiarity. Spanish-English was considered to have a high measure of Global Familiarity; Arabic-English to have what could be described as a ‘medium’ level of Global Familiarity; and Dhivehi-English to have a low level of Global Familiarity. Major tests like the TOEFL iBT are designed to measure English speakers’ abilities regardless of the Global Familiarity of test-takers’ accents. However, if it can be determined that raters’ accent-familiarity levels cause positive score variance or bias it is likely the lesser known accents would be disadvantaged, and the test-takers with better-known accents advantaged. In short, test candidates of lesser-known accents will be less likely to have a rater or raters familiar with their accents than test candidates with accents more globally well-known. If this is the case, all L1 backgrounds do not share the same likelihood of the potential positive benefits that could result from a rater accent-familiarity score effect.

Including accents whose native languages come from different language families was also considered when selecting the nonnative English accents to research. Examining score differences from speakers whose accents derive from differing language families, it was reasoned, would present a greater degree of variance of English L2 pronunciation patterns. For example, Spanish and Italian are both Latin languages, and there may be some crossover familiarity related speech-processing benefits related to Spanish-English familiarity that could also affect Italian-English processing by a rater familiar with Spanish-English but not Italian-English, as in Bent and Bradlow's (2003) "mismatched interlanguage speech intelligibility benefit" (p. 1607). For this reason, accents from different language families were selected.

Finally, geographical locations of the speaker populations were considered when making accent selections. It was reasoned if accents that are geographically distant to one another were selected, it would make the study more 'global' than 'regional' in its focus, rather than only focusing on Asian accents or European accents. A more globally spread selection of accents was decided most ideal for this study.

A lesson learned from the pilot study

The pilot study revealed that phonetic characteristics of Japanese-English pronunciation that could be deemed 'typical' of speakers of that accent caused little effect to intelligibility for raters 'very familiar' with Japanese-English, while 'unfamiliar' raters struggled with them. For example, significant differences in intelligibility success were recorded between raters very familiar with Japanese English and those unfamiliar with the accent concerning items that included voiced alveolar liquids (/l/) and voiced palatal liquids (/r/) phonemes (see also Riney & Flege, 1998). Both distinction and production of these liquids by Japanese-English speakers are known to be problematic (Carruthers, 2006), and the pilot study provided evidence to suggest that accent-familiarity contributed to the intelligibility of items including these phonemes. Raters very familiar with Japanese-English had comparatively better intelligibility success rates when coping with these kinds of typical pronunciation patterns than raters with less familiarity. This finding led to the conclusion that if problematic aspects of both production and perception of English pronunciation for each included accent could be determined, such information would be valuable to the construction of the research instrument.

In order to gain a better insight into such problematic aspects of the included accents was to determine the parameters of their phonetic inventories. It was reasoned this approach could identify phonemes commonly used in English pronunciation that are not present in the selected accents' L1s. The examples of difficulty for Japanese-English speakers concerning /r/ and /l/ are instances where missing phonemes from English phonology are not present in Japanese phonology. As discussed in Chapter one, Japanese-English speakers replace these phonemes with an "apicoalveolar tap /ɾ/" (Aoyama et al., 2004p. 234) included in Japanese phonology. Phonemes used in English but not included in Spanish, Arabic or Dhivehi represent potential problems for English pronunciation. Determining the phonological differences between English phonology and the phonology of each selected accents' L1 was instrumental to developing a strategic method of constructing the speaker participants' scripts. Prior to final script construction confirmation of each selected 'missing' or known-to-be-problematic phoneme was further discussed and confirmed with colleagues that were very familiar with each accent. These conversations also helped determine some common phonemic substitutions typical to each accent.

In order to understand what phonemes commonly used in English pronunciation are not present in each of the accents investigated in this study, it was important to first examine and review the phonetic inventories of English. Table 4.1 shows the phonetic inventory of consonants used by native speakers of most English dialects, and Figure 4.1 presents the phonetic inventory of vowel production of speakers of most English dialects (compiled from the speech accent archive)[13]. The

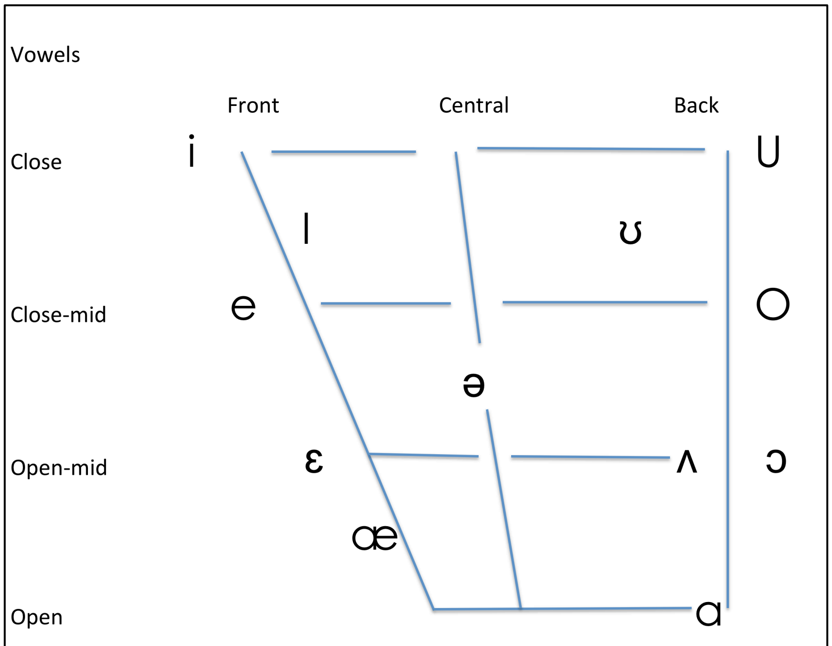
[13] The speech accent archive is an online resource of speech accent samples and descriptions provided by the Department of English program in Linguistics at George Mason University in Fairfax, Virginia, U.S.A. Available at: http://accent.gmu.edu

information in these charts serve as a base of what raters, or in this case - rater participants, might 'expect' to encounter when processing the speaker participants' recordings. As mentioned earlier, the phonetic inventories of the three included accents were compared to these phonetic inventories of English speech. Any and all consonants or vowels present in English phonology found to be missing from the target accents' L1 inventories were deemed of interest to this study.

Table 4.1: Phonetic inventories of English consonants

Consonants (Pulmonic)	Bilabial	Labiodental	Dental	Alveolar	Post alveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d				k g			
Nasal	m			n				ŋ			
Trill											
Tap or Flap											
Fricative		f v	θ ð	s z	ʃ ʒ						h
Affricate					tʃ dʒ						
Lateral fricative											
Approximant				ɹ			j				
Lateral approximant				l							

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible. Some consonants used are not shown.



Dipthongs not included in the chart: /eɪ/, /aɪ/, /ɔɪ/, /ɪə/, /eə/, /əʊ/, /ʊə/, /aʊ/

Figure 4.1: The phonetic inventory of vowels and diphthongs used in most dialects of native English. Retrieved from the speech accent archive

4.2.1 Spanish-English

Spanish-English was the first accent chosen to be included in the main study. Spanish, an Ibero-Romance language that evolved from several dialects of Latin spoken in Iberia[14], and is listed as either the official language or de facto official language in twenty countries with a population of more than four-hundred, twenty-five million native speakers. The United States has an estimated additional fifty million L1 and L2 Spanish speakers, and is currently the second largest Spanish speaking population in the world[15].

---

[14] David A. Pharies (2007). *A Brief History of the Spanish Language*. University of Chicago Press. p. 13.

[15] Retrieved from Wikipedia's "List of countries where Spanish is an official language." http://en.wikipedia.org/wiki/List_of_countries_where_Spanish_is_an_official_language#cite_note-M.C3.A1s_.27speak_spanish.27_que_en_Espa.C3.B1a-29

A contributing factor to the decision to include Spanish English was the findings of Winke et al. (2012). In their study they measured how raters' prior L2 study affected speaking scores on the TOEFL iBT, and found that the raters in their study scored Spanish-English samples relatively higher than Korean-English or Mandarin-English accents regardless of prior Spanish L2 study. Since all of their rater participants were recruited in the United States, it seems possible that these raters had developed at least some measure of familiarity with the accent due to the three factors this research project considers: Global Familiarity, geographic location of populations and media/news exposure. It is possible that though all of the raters in the Winke et al. study had not formally studied Spanish, they may have acquired enough exposure to the accent to establish familiarity with Spanish-English. It seems possible that even through indirect exposure to an accent (indirect meaning, instances where the listener is not actively engaged in conversation or attempting to acquire an L2) can positively affect speech processing of that accent. For these reasons, Spanish-English was selected to represent the most well known non-native accent included in this research. By including a rater population comprised of language professionals from around the world this study may confirm or dispel the Winke et al. finding that Spanish-English samples tend to receive higher scores than other accents when their study recruited rater participants only within the United States.

Phonetic features and characteristics of Spanish-English

Understanding the characteristics and features of Spanish-English, as with the other accents included in this study were of great importance to the design and construction of the research instrument. It was important to first understand and consider what other researchers have determined about the phonetic inventories of the speakers' L1s. Additionally, it was important to determine what common difficulties Spanish L1 speakers encounter concerning English pronunciation, and what possible phonetic substitutions Spanish-English speakers use. While what follows is not a completely comprehensive review of all Spanish and Spanish-English phonetics and phonology, it should be sufficient to both describe and justify the choices made for the design of this study. As will be discussed later in this chapter, eight sentences read by speakers of each accent were included in the final test. It was not possible, therefore, to include every potentially problematic aspect of Spanish-English pronunciation, so it was not necessary to determine every potential phonetic pitfall speakers of the accent face.

Table 4.2 and Figure 4.2 show the phonetic inventories of Spanish consonants and vowels respectively. Shown in red are the phonemes not included in Spanish pronunciations, yet are included in English phonological inventories. These were of the most concern to this study. Among the points of interest were the absences of all diphthongs in Spanish vowel usage, which feature prominently in most dialects of native English speech, as well as the missing voiced plosives and fricatives. The phonemes used in Spanish pronunciation but not in English, shown in blue, are included this way to prevent the need to refer back to Table 4.1 and Figure 4.1.

Spanish-English has its own distinctive characteristics concerning pronunciation. Ikeno et al. (2003) determined Spanish-English speakers to have particular difficulty with schwa vowels tending to use longer, less-reduced schwa vowels than most L1 speakers of English. Most of the unvoiced plosives, fricatives and affricates used in English are not present in Spanish. Williams (1977) examined stop consonant voicing of the voiced bilabial plosive /p/ and unvoiced /b/, and determined that Spanish-English speakers cannot perceptually separate the voicing contrasts of the two phonemes. Substitution of the voiced plosive with the unvoiced plosive can be expected from many Spanish-English speakers. Goldstein and Washington (2001, p.156) observed Spanish-English children making the following consonant substitutions for phonemes present in English but not in Spanish: /v/ ➔[b] and /ʃ/ ➔ [tʃ] /ʃʌvəl/ (shovel)➔ [tʃʌbəl]; [ɹ]➔/ɾ/ - /tɹən/ ➔ [tɾen].

Table 4.2: The phonetic inventory of Spanish consonants shown in black and blue; red consonants are English consonants not included in Spanish pronunciation

CONSONANTS (Pumonic)	Bilabial	Labiodental	Dental	Alveolar	Post alveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d				k g			
Nasal	m			n			ɲ	ŋ			
Trill											
Tap or Flap											
Fricative		f v	θ ð	s z	ʃ ʒ			x ɣ			h
Affricate					tʃ dʒ						
Lateral fricative											
Approximant				ɹ			j				
Lateral approximant				l			ʎ				

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

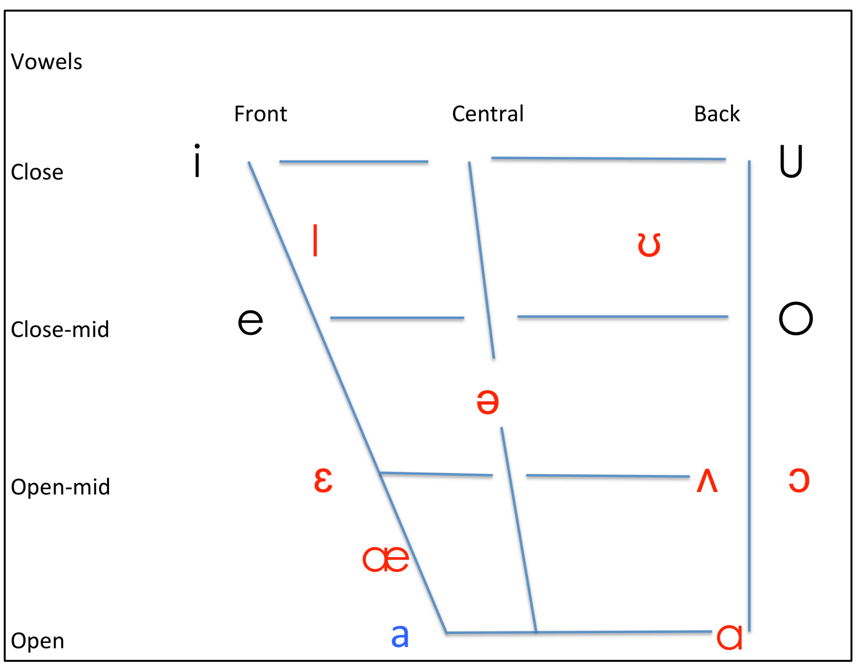


Figure 4.2: The phonetic inventory of vowels used by most dialects of Spanish; red symbols represent vowels used in English but not in Spanish, and blue symbols represent vowels used in Spanish but not included in English pronunciation

4.2.2 Arabic English

Arabic-English was selected because it served as a perfect compliment to Spanish-English and Dhivehi-English concerning how it met all of the desired elements for inclusion in the study. Arabic-English also differs from the other included accents according to the three criteria set-forth to determine inclusion in the study. Additionally, the researcher had colleagues working at the University of Qatar during the instrument-design and data collection phases, so it was reasoned that recruitment of both Arabic-English speaker participants and rater participants familiar with Arabic-English would be easily accomplished.

Arabic is a member of the Semitic language family, and is geographically distinct from Spanish and Dhivehi speaking regions. The Arabic speaking parts of the world comprise some twenty-two nations reaching from the Middle East to Morocco and the Horn of Africa. Figures concerning the global population of native Arabic speakers vary from an estimated 195 million[16] to four hundred twenty-two million speakers[17]. Arabic-English speakers currently receive frequent exposure in news and media outlets from both within and outside of the Arabic speaking world giving Arabic-English a measure of *global familiarity* that is higher than Dhivehi-English, but estimated to be less than that of Spanish-English. This presumption would later prove to be an accurate prediction of the rater population's overall familiarities with the three included accents (details in section 4.4 of this chapter).

---

[16] The Arab Culture and Civilization website that is sponsored by the Middle East Policy council and managed by The National Institute for Technology and Liberal Education reports 195 million native speakers and an additional 35 million L2 speakers. http://acc.teachmideast.org/introduction.php?module_id=1
[17] Retrieved from UNESCO World Arabic Language Day. http://www.unesco.org/new/en/unesco/events/prizes-and-celebrations/celebrations/international-days/world-arabic-language-day/

Phonetic features and characteristics of Arabic-English

For the purposes of explaining what informed the Arabic-English sections of the research instrument, a brief explanation of different characteristics of Arabic-English are described here. As with the explanations of Spanish-English and Dhivehi-English, what follows is not an exhaustive exploration into every nuance of possibility with the accent, but simply an explanation of what the key findings were that inspired the test's design. Included are discussions of the phonetic inventories of Arabic vowels and consonants, as well as different peculiarities of Arabic-English other researchers have determined.

Arabic and English phonologies share many phonemes, but not all. Table 4.3 and Figure 4.3 show the phonetic inventories of Arabic consonants and vowels respectively. Beginning with the consonants, the following phonemes were of particular interest to the design of the test. As Table 4.3 shows Arabic phonology does not include the voiced bilabial plosive /p/. It has long been discussed that this phoneme causes problems for native Arabic speakers learning English (Aziz, 1974). Flege and Port (1981) explain that though Arabic learners of English recognize the contrast of /p/ and /b/, they must learn to produce the phoneme through instruction. Additionally, Flege and Port when examining glottal pulsing of Saudi learners of English determined that a significant number of the Saudis ($p<0.01$) produced the phoneme /p/ both in the word-initial position and word-final position with glottal pulsing when compared to American speakers, and with other Saudis that had lived in the U.S. for several years ($p<0.01$) (p.5).

Table 4.3: The phonetic inventory of Arabic consonants found in most dialects shown in black and blue; red consonants are English consonants not included in Arabic pronunciation

CONSONANTS (Pumonic)	Bilabial	Labiodental	Dental	Alveolar	Post alveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d				k g	q		ʔ
Nasal	m			n				ŋ			
Trill				r							
Tap or Flap				ɾ							
Fricative		f v	θ ð	s z	ʃ ʒ				χ ʁ	ħ ʕ	h
Affricate					tʃ dʒ						
Lateral fricative											
Approximant				ɹ			j				
Lateral approximant				l							

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

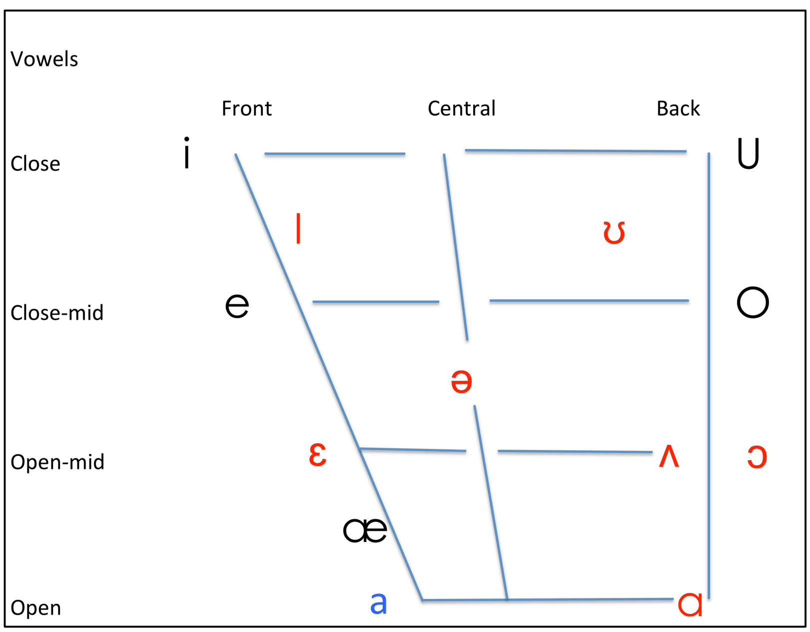


Figure 4.3: The phonetic inventory of vowels used in most dialects of Arabic; red symbols represent vowels used in English but not in Arabic, and blue symbols represent vowels used in Arabic but not included in English pronunciation

4.2.3 Dhivehi-English

Dhivehi-English, also known as Maldivian-English, was selected as the third nonnative English accent, and represents an accent that has very little global familiarity. Dhivehi, a member of the Indo-Aryan language family closely related to Sinhala (Sri Lanka)[18] is the native language of the Republic of the Maldives. An atoll nation in the Indian Ocean, the Republic of the Maldives has a population of less than 350,000 people dispersed on 192 of the nation's 1,192 islands. There are no available statistics regarding the number of Maldivians living abroad or what the size of the Dhivehi-English speaking population is[19]. Dhivehi-English was selected for inclusion in this study in order to include an accent with a considerably smaller speaker population when compared to Spanish-English. Winke et al. (2013) found in their study that, "raters tended to assign relatively higher ratings to Spanish-accented speech samples" (p.244) compared to Chinese (Mandarin) and Korean L1 speakers. The raters in that study were all American university students, and though they all may not have formally studied Spanish as an L2, which was how accent-familiarity was determined in that study, it is possible that many of the raters had experienced enough passive exposure to Spanish-English by living in the United States to cause the positive bias they observed. Dhivehi-English, it was reasoned, due to the rarity of the accent resulting from the relatively small and geographically isolated speaker population might receive comparatively lower overall pronunciation scores than Spanish or Arabic L1 speakers. In short, Winke et al. have shown evidence to suggest that Spanish-English is favored; it might be due to the fact that it is a 'well-known accent', and Dhivehi-English could function well as a 'not well-known accent'. If it can be determined that there are accents that are unfairly advantaged or disadvantaged on tests like the TOEFL iBT because of accent-familiarity it would have devastating effects on the reliability and validity of such tests.

---

[18] Cain, B and J. Gair. (2000). Dhivehi (Maldivian). Lincom Europa. Muenenchen. p. 1.
[19] Retrieved from Wikipedia's page titled, "Maldives."
http://en.wikipedia.org/wiki/Maldives

Dhivehi-English was also selected, in part, because the researcher has a colleague that could both assist in recruiting speaker participants with the accent and rater participants familiar with the accent. It was decided that in order to include an accent that is largely unknown in terms of *global familiarity*, extremely isolated in its geographic location and that receives little to no international media/news coverage that it would be necessary to have at least one close contact from or working closely with people from that accent group. The inclusion of Dhivehi-English did include certain challenges, namely there are no published research concerning the English spoken in the Maldives, and there were difficulties concerning recruiting rater participants willing to complete the test (discussed in section 4.4 of this chapter).

Phonetic features and characteristics of Dhivehi-English

Divehi-English, as previously mentioned, is a largely unknown English accent or 'World English'. As mentioned, there are no books or articles published that describe, or even include Dhivehi-English, and very little has been published on the native language, Dhivehi. Leonid Kulikov (in Reynolds, 2004) stated, "the Maldivian language has probably received less scholarly attention than any other official (i.e. used as the official language of a country) language in the world" (p.249). However, the inventories of Dhivehi consonant and vowel phonology have fortunately been described (see Cain, 2000; Cain & Gair, 2000; Fritz, 2002). Similar to how Spanish-English phonology was investigated in this study, the primary focus was to determine what phonemes that are present in English phonological inventories are not included in Dhivehi phonology (see table 4.4 and figure 4.4). These catalogues of Dhivehi vowel and consonant phonology were extracted from Cain and Gair (2000) and Fritz's (2002) descriptions. Additionally, personal correspondence with two university ESL/EFL lecturers, both natives of the Maldives, each with several years experience teaching English at the university and high school levels in the Maldives and abroad, provided invaluable contributions toward developing a rudimentary description of the potential characteristics of Dhivehi-English phonology.

Table 4.4: The phonetic inventory of Dhivehi consonants shown in black and blue; red consonants are English consonants not included in Dhivehi pronunciation

CONSONANTS (Pumonic)	Bilabial	Labiodental	Dental	Alveolar	Post alveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		t̥ d̥ ^{n}d̥		k g ^{n}g			
Nasal	m			n			(ñ)	ŋ			
Trill											
Tap or Flap											
Fricative		f v	θ ð	s z	ʃ ʒ	š̥					h
Affricate					tʃ dʒ						
Lateral fricative						l̥					
Approximant				ɹ			j				
Lateral approximant				l							

Where symbols appear in pairs, the one to the right represents a voiced consonant. Shaded areas denote articulations judged impossible.

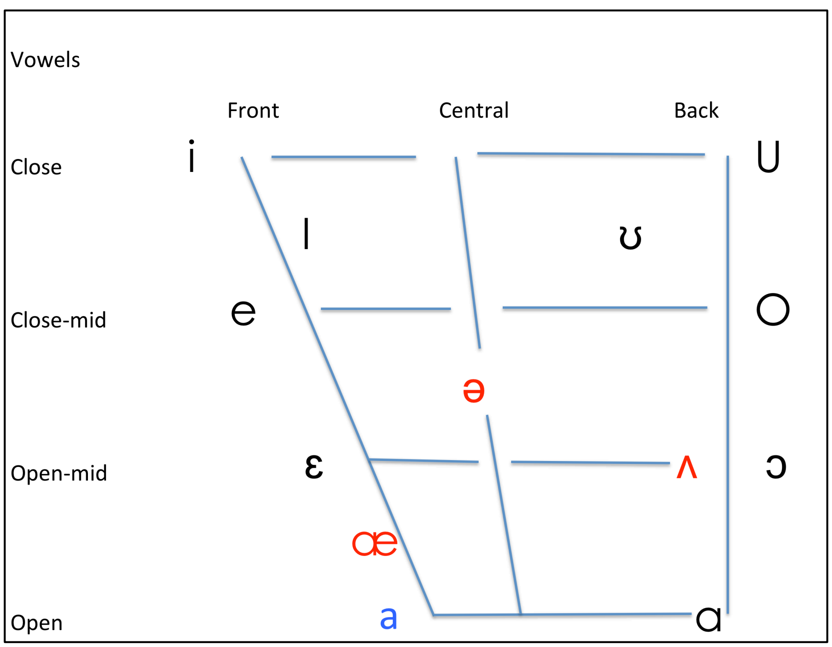


Figure 4.4: The phonetic inventory of vowels used in Dhivehi; red symbols represent vowels used in English but not in Dhivehi, and blue symbols represent vowels used in Dhivehi but not included in English pronunciation

Dhivehi-English vowels are limited to only five basic vowels each with two degrees of vowel length: /i/ and /ī/; /e/ and /ē/; /ɑ/ and /ā/; /u/ and /ū/; /o/ and /ō/ (Cain & Gair, 2000, p.9). The use of diphthongs are difficult to describe or determine in Dhivehi. Fritz (2002) states:

> In contrast to Sinhalese, Modern Dhivehi possesses true diphthongs which because of their apparently ambiguous character need a detailed examination. Basically, we have to distinguish phonemic diphthongs from numerous kinds of diphthongisations that represent the phonetic realization (i.e. pronunciation) of certain phonological structures but cannot be interpreted as diphthongs according to phonological rules.
>
> (Fritz, 2002, p. 24)

Fritz discusses that the occurrence of actual phonemic diphthongs are "restricted" (p.24) with only a few words and seem limited (at least from Fritz' descriptions) to only include /aj/ as in /skaj/ (sky), /ui/ as in /gui/ (gooey) and /ɔj/ as in /bɔj/ (boy). These actual phonemic diphthongs appear to only occur in a handful of Dhivehi words. The diphthongs /au/ and /ai/ seldom appear and are limited to only the southern dialects. From these descriptions it is clear that the inclusion of diphthongs are not prevalent features of Dhivehi pronunciation. Therefore, it is likely that Dhivehi-English pronunciation might include different vowel substitutions for diphthongs. Apart from diphthongs, during personal correspondence with the two EFL/ESL teachers from the Maldives, it was determined that many Dhivehi-English speakers have particular problems producing /æ/, as in /kæt/ (cat), and tend to substitute /æ/ with /a/. Certainly, there is more to determine concerning Dhivehi-English vowel phonology and deserves through research; however, for the purposes of constructing this research instrument, the measure of knowledge was considered adequate to draft the sentences for the Dhivehi-English speakers that might highlight unique characteristics of vowel vocalization in the accent.

Similar to how Dhivehi-English vowel usage was examined and deduced, the inventories of native Dhivehi consonants were also examined to gain insight into possible problematic features and characteristics of Dhivehi-English consonant phonology. By reviewing the relevant literature and through personal correspondence with the English language teachers from the Maldives a basic understanding was determined. Looking at Table 4.4, the most obvious missing phonemes from Dhivehi phonology that are present in English pronunciation are the dental and post alveolar fricatives and affricates. The Maldivian teachers reported 'typical' substitutions / θ / are /t/ and /d/, and /d/ substituted for /ð/. Interestingly, Cain and Gair (2000) mention that the voicing of the bilabial plosives /p/ and /b/ are contrastive in Dhivehi, as are the alveolar plosives /t/ and /d/, and velar plosives /k/ and /g/ (p. 7). Cain and Gair also describe an interesting feature of Dhivehi phonology concerning loan words and include the only descriptions of what could be considered Dhivehi-English. They explain that:

> Synchronically /p/ and /f/ contrast . . . At one point, Dhivehi had only /p/, but some time after the 1600's, word initial and intervocalic /p/ changed to /f/ . . . Subsequently, /p/ in borrowed words also appeared as /f/: /hasfatālu/ 'hospital'. Currently, however, the /p/ in newly borrowed words is retained: /ripōṱu/ 'report.'
> (Cain & Gair, 2000, p. 8)

Another interesting aspect of Dhivehi phonology is its complete lack of consonant clusters. "Dhivehi does not tolerate consonant clusters in any position", Fritz states (2002, p. 36). Though Fritz was describing the Dhivehi language, it is reasonable to believe that the many consonant clusters prevalent in English could cause problems. Through personal correspondence, it was explained that production and perception of the post-alveolar fricatives and affricates are problematic Maldivian English learners, which was predicted due to their not occurring in Dhivehi phonology and represent common consonant clusters in English. For example, Dhivehi-English speakers often have difficulty with both productive and receptive distinction with words like /ʃuz/ (shoes), / dʒuz/ (Jews) and /tʃuz/ (chews or choose). Pronunciation, it was explained, of all will often more closely resemble /tʃuz/ (chews or choose). Dhivehi phonology also lacks the velar nasal /ŋ/ that is another very common consonant cluster used in English (e.g. /rənɪŋ/ 'running').

There are, of course, other features of Dhivehi-English vowel and consonant use not described in this thesis. However, this description is, at the moment, the only descriptions of Dhivehi-English phonology. This research is not focused on the phonology of nonnative speakers. Three nonnative accents were chosen to determine if raters' differing familiarities with accents affects the scores they administer for pronunciation, and also to examine intelligibility. Though this description of Dhivehi and possibly Dhivehi-English phonology is certainly incomplete, it was deemed that an adequate knowledge of the L1 had been determined that was necessary to construct the scripts for the Dhivehi-English speakers. The sentences for each accent are described later in section 4.5.2 of this chapter.

Limitations of including additional accents

Other non-native accents were considered for inclusion in the study, but were not selected for different reasons. Mandarin-English was the first choice to include as the accent with the highest global familiarity rating; however, during the pilot study it was determined that YouTube is a blocked website in China, and is forbidden. Since all the listening samples would be accessed via embedded links to YouTube videos, Mandarin-English was eliminated from consideration, as it may have been problematic or unethical to recruit or include raters residing in China. Vietnamese-English was likewise excluded from consideration for the same reason. Mongolian-English was another accent considered for this study, but the researcher could not locate colleagues willing to assist in the recruitment of speaker participants, and feared recruiting a sufficient number of rater participants either 'very familiar' or with 'some familiarity' with Mongolian-English might prove to be too difficult.

4.3 The speaker participants

What follows is an explanation of the rigorous processes developed to recruit and select the speaker participants for the main study. Five male and five female English learners from each of the three included accents, thirty people in total, were recruited as candidates for the nonnative speaker participants. The final selection of speaker participants included thirteen people between the ages of eighteen and forty-nine (see Table 4.5). The participants included twelve nonnative speakers and one native speaker (the researcher, then 42 years-old, male from the southern United States – New Orleans, Louisiana). Four speakers from each nonnative accent (male *n*=2; female *n*=2) participated.

Table 4.5: Final speaker participant information

Speaker Order	Accent	Gender	Age	Home Country
Speaker 1	American-English	Male	42	USA
Speaker 2	Dhivehi-English	Female	23	Maldives
Speaker 3	Arabic-English	Female	18	Saudi Arabia
Speaker 4	Dhivehi-English	Female	20	Maldives
Speaker 5	Spanish-English	Male	49	Nicaragua
Speaker 6	Arabic-English	Male	18	Sudan
Speaker 7	Spanish-English	Male	36	Peru
Speaker 8	Dhivehi-English	Male	22	Maldives
Speaker 9	Spanish-English	Female	39	Cuba
Speaker 10	Arabic-English	Male	18	Palestine
Speaker 11	Spanish-English	Female	35	Mexico
Speaker 12	Arabic-English	Female	18	Oman
Speaker 13	Dhivehi-English	Male	20	Maldives

The ordering of the speakers, apart from the native speaker (Speaker 1), was decided by using an online random sequence generator[20]. Each of the twelve nonnative speakers was assigned a number one through twelve. The sequence generator provided a random sequence of numbers one through twelve, and this list was used to determine the order of the speakers. By presenting the speaker participants in a random order rather than ordering the speakers according to accent, age or gender, it was reasoned, would be reflective of how ETS delivers test-taker samples for scoring to actual raters.

[20] The online random sequence generator is available at: https://www.random.org/sequences/?mode=advanced

The speaker participants included in this study were recruited with the assistance of three colleagues that volunteered their time and effort. One assistant from the Maldives working at that time as an English lecturer at Qatar University was responsible for recruiting and recording the five female Arabic-English speakers and all ten Dhivehi-English speakers. Another research assistant working as an English lecturer at Qatar University recruited and recorded the five male Arabic-English speakers, and a third colleague that was working as an English lecturer at the College of Southern Nevada in the United States at that time recruited and recorded the ten Spanish-English speakers. In the following subsections the recruitment and selection processes for the speaker participants of each accent group are discussed separately.

4.3.1 The Spanish-English speaker participants

As stated previously, ten Spanish-English speakers (male *n*=5; female *n*=5) were recruited and recorded for possible inclusion in the study (see Table 4.6). The candidates were all native Spanish speakers from five Spanish speaking countries in Central and South America (Mexico, Peru, Columbia, Cuba, Nicaragua). All of the candidates were enrolled at the time in ESL courses at the College of Southern Nevada in the United States. The final selection of the four speaker participants to represent the Spanish-English accent (marked with an asterisk * in Table 4.6) were based on the following criteria and carried out by the researcher:

1. There must be two male and two female participants selected.
2. For a speaker to be selected their recording must include two different sentences read without any unnecessary long pauses, repeating of words or wrong words spoken (clearly mistaking one word in the script for another non-targeted word), as such problems could impact the rater transcription task.
3. When at all possible, use the first reading of the selected candidates' sentences as the sample for use in the test unless the second reading was clearly of a better quality.

4. If possible, it would be ideal if the final selected participants include speakers from different home countries, as Spanish is spoken in several countries and may demonstrate different characteristics in their pronunciation. This would decrease the likelihood of arguments made that the samples didn't encompass a wide enough range of Spanish-English accent varieties.

Table 4.6: Spanish-English speaker candidate information

Speaker	Gender	Age	No. of years studying English	Home country
1	M	31	3	Mexico
2*	M	36	14	Peru
3	F	40	3	Mexico
4	F	49	5	Colombia
5*	F	35	10	Mexico
6*	F	39	2	Cuba
7	F	22	5	Peru
8	M	33	3	Mexico
9*	M	49	10	Nicaragua
10	M	43	4	Mexico

* *Selected as a final speaker participant*

4.3.2 The Arabic-English speaker participants

Similar to the Spanish-English speaker recruitment process, ten native Arabic speakers enrolled in university EFL courses at the University of Qatar volunteered to participate in the study. Five women and five men between eighteen and nineteen years of age were recruited and recorded reading the script prepared specifically for the Arabic-English accent. The recruits' home countries were Qatar, Syria, Saudi Arabia and Oman (see table 4.7). The male and female candidate speakers were recruited and recorded separately by two of the researcher's colleagues that work as English lecturers at Qatar University. The final four candidates selected to represent Arabic-English accents in the final study were selected based on the following criteria, which is very similar to the criteria used to select Spanish-English speakers:

1. There must be two male and two female participants selected.
2. For a speaker to be selected their recording must include two different sentences read without any unnecessary long pauses, repeating of words or wrong words spoken (clearly mistaking one word in the script for another non-targeted word), as such problems could impact the rater transcription task.
3. When at all possible, use the first reading of the selected candidates' sentences as the sample for use in the test unless the second reading was clearly of a better quality.
4. If possible, it would be ideal if the final selected participants include speakers from different home countries, as Arabic is spoken in several countries and may demonstrate different characteristics in their pronunciation. This would decrease the likelihood of arguments made that the samples didn't encompass a wide enough range of Arabic-English accent or dialect varieties.

Table 4.7: Arabic-English speaker candidate information

Speaker	Gender	Age	No. of years studying English	Home country
1*	M	18	10	Sudan
2*	M	18	12	Palestine
3	M	18	12	Qatar
4	M	19	12	Qatar
5	M	18	12	Qatar
6*	F	18	3	Saudi Arabia
7	F	19	10	Syria
8	F	18	13	Oman
9*	F	18	13	Oman
10	F	18	13	Oman

* *Selected as a final speaker participant*

Since it was part of the design process to include speakers from different regions of the Arabic speaking world, some additional details concerning the four Arabic-English speakers' home countries and dialects are offered here. The dialect of the two female speaker participants selected from the group of five potential participants, one from Oman and the other from Saudi Arabia is called Peninsular Arabic. One of the male speaker participants is from Palestine, and would most likely speak in a Levantine dialect that is common in the regions of the Western Mediterranean. The remaining male speaker reported Sudan as his home country, and most likely speaks a Sudanese dialect. As stated in section 4.2.3 of this chapter, the phonological inventories of spoken Arabic[21] used as a reference to determine potential problematic phonemes claimed to include 'most dialects' of spoken Arabic. However, since this research is more concerned with the Arabic-English accent than individual dialects of the Arabic language, descriptions of Arabic-English (also discussed in section 4.2.3 of this chapter) should be sufficient and applicable to all four of the Arabic-English speaker participants.

4.3.3 The Dhivehi-English speaker participants

Four speaker participants were selected from a group of ten English learners studying at the Maldives National University to represent the Dhivehi-English accent (see table 4.8). The same colleague that recruited and recorded the female Arabic-English speakers, who is also a native Maldivian, assisted in recruiting and recording the Dhivehi-English candidates in the capital city of Malé. Since Dhivehi is spoken only in the Maldives, the selection criteria implemented to decide the four Dhivehi-English speakers for use in the research instrument is the same as those for Spanish-English and Arabic-English with exception to the fourth criteria favoring a selection of speakers from different home countries. The selection criteria applied is as follows:

1. There must be two male and two female participants selected.
2. For a speaker to be selected their recording must include two different sentences read without any unnecessary pauses, repeating of words or

[21] The phonological inventories of Arabic consonants and vowels were taken from The Speech Accent Archive website's page for the Native Phonetic Inventory: arabic [sic], which states, "These are the sounds found in most native arabic [sic] dialects". There may be missing phonemes not included in the inventories.
http://accent.gmu.edu/browse_native.php?function=detail&languageid=5

wrong words spoken (clearly mistaking one word in the script for another non-targeted word), as such problems could impact the rater transcription task.

3. When at all possible, use the first reading of the selected sentence unless the second reading is clearly of better quality.

Table 4.8: Dhivehi-English speaker candidate information

Speaker	Gender	Age	No. of years studying English	Home country
1	M	29	12	Maldives
2	F	26	12	Maldives
3	M	22	10	Maldives
4	F	23	10	Maldives
5*	M	22	11	Maldives
6*	F	23	10	Maldives
7	F	30	10	Maldives
8	M	20	12	Maldives
9*	M	20	10	Maldives
10*	F	20	12	Maldives

* *Selected as a final speaker participant*

4.4 The Rater Participants

The rater participants recruited for the main study included 190 English language teachers, researchers and academics in the fields of education, linguistics, applied linguistics and language testing. This section of the chapter provides various discussions and descriptions of the rater participants. Raters are, after all, the main subjects of this research, so this section includes more detail and discussion of the rater population than is usually included in a methodology chapter. What follows are discussions concerning characteristics of the rater population as they relate to the 'appropriateness' of the raters, and/or the development of an accent-familiarity argument. These discussions serve to clarify the measures and natures of accent-familiarity of the rater participant population, and establish the qualification of this particular rater participant population. If the rater population cannot be determined to be appropriate or qualified, then the findings of this research cannot be deemed useful. Likewise, the nature of the rater population's accent familiarities and their measures must be clear prior to making inferences from the data. For these reasons this section is designed as it is. The section concludes with an explanation concerning the reasoning for the lack of rater training included in this study. Though normally such potential limitations are included in the final chapter of a paper like this, it was deemed more beneficial to present the discussion here, as it contributes to the argument that the rater participants and their contributions are appropriate.

The rater participants reported coming from 35 different countries (see Table 4.9). After the pilot study, the question was raised whether or not it was beneficial that there be great diversity of home countries and first-language backgrounds from the rater participant population. For example, could there be any arguable benefit from including one participant from Azerbaijan or another whose first language is Lëtzebuergesch (Luxembourgish) rather than recruiting only native English speakers? The question is reasonable; however, such variety amongst the raters has value. The value is that it reveals how varied the community of educators and researchers actually are that are actively working in or interested in English language teaching and language testing. The diversity of the rater population is reflective of the diversity of English use described by Kachru (1985). The rater population was recruited via email and postings on online Language Testing bulletin boards both by the researcher and colleagues. Invitations were only sent to people actively involved in English language teaching, testing or related research. So the resulting participants are reflective of the current population of academics and educators involved in the field. It is for these reasons that there are benefits to including such participants as one rater from Azerbaijan or a native Lëtzebuergesch speaker. After all, it is also a matter that English is spoken in various accents, and unbiased testing of speakers' pronunciation that is being researched.

Table 4.9: Raters' home country (or country they were raised in)

Country	Frequency	Country	Frequency
The United Kingdom	80	New Zealand	2
United States of America	34	Norway	2
Australia	8	Azerbaijan	1
Canada	7	Denmark	1
Maldives	6	France	1
Spain	5	Italy	1
Hungary	4	Jamaica	1
Japan	4	Luxembourg	1
Brazil	3	Northern Ireland	1
Netherlands	3	Poland	1
Turkey	3	Portugal	1
Austria	2	Slovakia	1
China	2	South Africa	1
The Czech Republic	2	South Korea	1
Finland	2	Sweden	1
Germany	2	Syria	1
Greece	2	Taiwan	1
India	2		
Total:			190

Concerning the native speaker status of the rater population, most of the rater participants (n=137) reported being native English speakers, though ten of the native speakers indicated they had two native languages (see Table 4.10). The 53 nonnative English speaking rater participants included 29 different native languages (see Table 4.11). It is commonly presumed that native speakers are best suited to serve as raters of tests of speaking, though the evidence from the literature suggests differences of opinion. Brown (1995) found native English speakers tend to score more severely than non-native speakers, but that the differences found were not significant. Conversely, Fayer and Krasinski (1987) found nonnative speakers were more severe than native speakers, though their study examined speaking performance ratings of L2 Spanish. And Zhang and Elder (2011) determined that qualified nonnative English speakers can be as reliable as native speaker raters (see also Xi & Mollaun, 2009). Therefore, the inclusion of a native Lëtzebuergesch speaker participant with ESL teaching experience in the UK and Senegal is appropriate, and the ratio of native speaker to nonnative speaker rater participant was ideal.

Table 4.10: How the native English speaker rater participants reported their native language

Reported native language(s)	Frequency
English	125
British English	2
English, French	2
English, Spanish	2
English, Portuguese	1
English, German	1
English, Greek	1
Gujarati, English	1
Jamaican Creole, English	1
Dutch, English	1
Total*:*	137

Table 4.11: Reported native language(s) of nonnative English speaking rater participants

Reported native language(s)	Frequency	Reported native language(s)	Frequency
Dhivehi	6	Azeri	1
Japanese	4	Bergamasque (dialect), Italian	1
Spanish	4	Brazilian Portuguese	1
German	3	Danish	1
Hungarian	3	Hungarian, German	1
Chinese	2	Hungarian, Polish	1
Czech	2	Korean	1
Dutch	2	Lëtzebuergesch	1
Finnish	2	Mandarin	1
French	2	Polish	1
Norwegian	2	Slovakian	1
Portuguese	2	Swedish	1
Turkish	2	Tamil	1
Greek	2	Turkish	1
Arabic	1		
Total*:*			53

It is interesting to observe how some of the rater participants reported their native language. It reveals insight into how language is closely related to identity (e.g. Zuengler, 1988). In particular, how the participants that reported having more than one native language ordered the languages was revealing (whether English was listed first or second). Even among native English speakers, some participants distinguished or qualified their native language specifically (e.g. 'British English' rather than 'English', or to not be confused with 'American English'). These details concerning accent, language and identity are included here, as it provides additional understanding concerning how current English pronunciation testing includes no descriptions of a single, idealized model. Likewise, it reveals how problematic it would be for a test to insist that one standard of pronunciation is best might be received by this particular rater population.

Most of the raters (n= 134) reported they were currently teaching ESL or EFL, 44 reported they previously taught ESL or EFL and twelve had no ESL/EFL teaching experience. Prior ESL/EFL teaching was deemed an important rater characteristic to examine, as language teaching is often a requirement, or at least considered a favored characteristic, for rater recruitment on high-stakes language tests; however, it has been found to not be vital to all reliable rating (Royal-Dawson & Baird, 2009). All levels of ESL/EFL teaching were represented in the rater participants' reported experiences, and some participants listed more than one. The data revealed most (n= 143) raters' teaching experience included higher education lecturing at universities, colleges, junior colleges or technical colleges. Fifteen reported experience teaching ESL/EFL in companies or businesses. Twenty participants had experience teaching high school and or junior high school students, and ten taught at elementary schools. Eighteen participants reported they worked or had worked in private English schools. These results were considered particularly valuable toward making the claim that the rater population was 'qualified'. Language teachers of any level are accustomed to assessing students' language proficiency, and should be capable to rate speakers for pronunciation according to a rating scale.

This research theorizes that one of the means of developing accent-familiarity is by living in different places and encountering both the native languages and their accented English over time. Details were collected pertaining to the countries the raters were living in at the time of completing the test, as well as any other countries than their home countries that they had lived for one or more years. The rater population demonstrated a wide range of potential cultural and linguistic experiences with most (n= 151) stating they had lived in countries other than their home country for one or more years. It was revealed that the raters have lived in 72 different countries for one or more years (see Table 4.12). Additionally, at the time of completing the test, the participants were living in 36 different countries (see Table 4.13) with most (n= 162) stating that they had been living in that country for more than five years (see Table 4.14). This information shows the potential for the rater population to be deemed 'familiar in general' with nonnative English accents. Other more specific data was also collected concerning familiarity with particular accents.

Table 4.12: Countries other than their home country raters participants lived one or more years

Country	Frequency	Country	Frequency
The UK	22	Fiji	2
Japan	22	The Republic of the Phillippines	2
Germany	18	Canada	2
Spain	16	Hungary	2
The United States of America	14	Morocco	2
France	14	Jordan	2
Italy	11	Saudi Arabia	2
China	10	Syria	2
Thailand	9	Taiwan	2
South Korea	8	Tanzania	2
Russia	7	Botswana	1
The Czech Republic	7	Brazil	1
Malaysia	6	Brunei	1
Australia	5	Peru	1
Egypt	5	Mexico	1
India	5	Columbia	1
Greece	5	Cote d'Ivoire	1
Portugal	5	Ethiopia	1
Hong Kong	5	Nigeria	1
Argentina	4	Belgium	1
Singapore	4	Ghana	1
Kuwait	4	Oman	1
Turkey	4	Tonga	1
Poland	4	Iraq	1
Finland	3	Nepal	1
Holland	3	Cambodia	1
Israel/Palestine	3	Martinique	1
Indonesia	3	Mauritania	1
Vietnam	3	New Zealand	1
Kenya	2	Qatar	1
Sudan	2	Papua New Guinea	1
Chile	2	Denmark	1
The United Arab Emerites	2	Sweden	1
Cyprus	2	Switzerland	1
Azerbaijan	2	Bulgaria	1
Ecudaor	2	Senegal	1

Table 4.13: Countries the raters were living in at the time of completing the test

Country	Frequency	Country	Frequency
The United Kingdom	70	Norway	2
Japan	20	Sweden	2
The United States of America	17	Thailand	2
Australia	8	China	1
Canada	7	Denmark	1
Spain	7	Egypt	1
The Republic of the Maldives	5	Greece	1
The Netherlands	5	Hungary	1
Germany	4	India	1
South Korea	4	Israel	1
Turkey	4	Kuwait	1
France	3	Libya	1
Qatar	3	Luxembourg	1
Austria	2	Oman	1
Azerbaijan	2	Poland	1
Brazil	2	Portugal	1
The Czech Republic	2	The United Arab Emirates	1
Finland	2	Vietnam	1
Greece	2		
Total*:*			190

Table 4.14: Length of time in the country raters were living at the time of completing the test

Length of time	Frequency	Percent
0-6 months	4	2.1
7 months to 1 year	1	0.5
1-2 years	3	1.6
2-5 years	20	10.5
More than 5 years	162	85.3
Total	190	100

Since this research is concerned with raters' differing accent familiarities, data was collected pertaining to how familiar the rater participants were with nine different accents. The decision to include nine accents rather than only the three target accents was for two reasons: first, by doing so the accents of the speaker-participants included in the test would not be clearly identified. This was important to keep the cultural and linguistic identities of the speaker participants less predictable.

Second, it would offer insight not only into the rater participants' familiarities with individual accents, but could provide some insight into how familiar they were with nonnative speech in general. As will be discussed later in this chapter, the raters self-reported their levels of familiarity with the following nonnative English accents: Japanese-English, Korean-English, Urdu-English (Pakistan), Arabic-English, Indonesian-English, Dhivehi-English, Brazilian-English, Spanish-English and German-English. The details of the rater population's familiarity with the target languages can be seen in Tables 4.15-17 (please see the Appendix B for the familiarity reports of all the included accents).

Table 4.15: Rater participants' reported familiarity with Spanish-English

		Frequency	Percent
Valid	No Familiarity	20	10.50
	Limited Familiarity	57	30.00
	Some Familiarity	78	41.10
	Very Familiar	35	18.40
Total		190	100.00

Table 4.16: Rater participants' reported familiarity with Arabic-English

		Frequency	Percent
Valid	No Familiarity	40	21.10
	Limited Familiarity	62	32.60
	Some Familiarity	68	35.80
	Very Familiar	20	10.50
Total		190	100.00

Table 4.17: Rater participants' reported familiarity with Dhivehi-English

		Frequency	Percent
Valid	No Familiarity	165	86.80
	Limited Familiarity	17	8.90
	Some Familiarity	1	0.50
	Very Familiar	7	3.70
Total		190	100.00

The accent familiarities of the rater population for the three-targeted languages were not all ideal, but were also not unexpected. As was guessed, familiarity with Spanish-English was highest and familiarity with Dhivehi-English was lowest. This was not surprising since the numbers of speakers of and geographic sizes of the two native languages are greatly different. A goal during the data collection phase of the study was to recruit at least 30 participants either 'very familiar' or with 'some familiarity' with each target language; it was considered most ideal if 30 participants that were 'very familiar' would participate for each accent group. The Spanish-English familiarity groupings met the 'most ideal' results (see Table 4.15), and the Arabic-English familiarity results were also ideal (see Table 4.16). The numbers of raters recruited representing both 'very familiar' and 'some familiarity' with Dhivehi-English were not ideal with only 25 rater participants with any familiarity at all with the accent (see table 4.17).

Recruiting raters in the Maldives proved to be problematic. As was mentioned previously, two Maldivians contributed to the study as research assistants recruiting and recording speaker participants both in the Maldives and in Qatar. Naturally, these two people could not participate as rater participants. One of the volunteer research assistants also attempted to recruit rater participants in the Maldives first via email to English teaching faculty at the Maldives National University, and again in person while on a trip home. Email recruitment resulted with two participants. The remaining five participants 'very familiar' with Dhivehi-English were recruited on her trip home. Apart from these efforts the researcher was extended an invitation to join a teachers' group on Facebook for Maldivian teachers. Invitations to participate in the study were put on this group's message board three times with endorsements from the Maldivian research assistants, though these efforts failed to attract any volunteers. In total, eleven Maldivians logged on to the test, but only seven completed it. A possible reason for the poor completion rate of participants from the Maldives could be the length of the research instrument. Among the 190 rater participants that completed the test only four members complained that the research instrument took too long to complete, or had too many speakers; all of these complaints came from participants from the Maldives. One rater (no.189) complained simply, "too long", yet completed the test in eight minutes. The initial period of recruitment was from November 27, 2013 to April 6, 2014 where the first 185 participants completed the instrument. The recruitment period was extended due to the low number of respondents either ‘very familiar’ or ‘some familiarity’ with Dhivehi-English. Though it is unfortunate the number of rater participants with any familiarity with Dhivehi-English was low, the data collection phase needed to end.

Other biographical data collected from the raters collected included age ranges and gender. These details were collected not because it was suspected gender or ages were vital to determining the quality of the recruited rater population, but in case anyone wishes to conduct a replication study. They are included here for any interested parties. The rater participants included 113 women and 76 men. The raters’ ages are shown in Table 4.18.

Table 4.18: Age ranges of the rater participants

	Age ranges	Frequency
Valid	21-29	15
	30-39	45
	40-49	65
	50-59	36
	60 or older	29
Total		190

4.4.1 Why rater training was not employed

Before introducing the test the rater participants completed, it is necessary to first explain why rater training was not provided in this study. Rater training was not a reasonable possibility to include due to the circumstances of this study, though it was also not deemed necessary or appropriate to the study. Considering the rater participants were all uncompensated volunteers recruited from numerous locations around the world over a period of several months, it was not practical to presume an adequate number of volunteers would complete both a training session and complete the test. As was discussed in Chapter 3, the targeted group of professionals this research sought are very busy people. The decision to not include any formal or informal rater training might be considered a limitation; however, the decision was deliberate and based on established theory.

Fulcher (2003) explains that rater training should not precede a validity argument, and that including trained raters in attempting to make a validity argument is problematic. This is because rater training presumes that the validity of a test's rating scale is already established. To be precise it suggests that the scale's descriptors have already been determined to be reasonable, and that any inferences drawn from resulting scores can be theoretically and empirically defensible. This research directly questions the validity of rating scales for pronunciation (or 'delivery' on the TOEFL iBT) that rely on raters' individual measures of difficulty coping with and/or deciphering the phonological content of test candidates' utterances. Raters' differing accent familiarities and levels of familiarity, it is argued, leads to variation in the severity or leniency of raters' implementation of such rating scales. Fulcher states, "if rater scores are to be used as part of a validity argument, it is questionable whether the evidence is acceptable if the scores come from trained raters" (p. 147). It is this reasoning that rater training was not included in this study, and that the lack of rater training is not considered a limitation of this study.

4.5 The Test

A three-part test was designed specifically for the purposes of attempting to answer the research questions. For the purposes of clarification, this test was not designed or intended to be considered as a practical or ideal means of assessing the holistic speaking abilities of English learners. Rather, it was designed explicitly to attempt to answer the research questions of this study. Constructed in and administered via the online survey host, Survey Monkey, the test was open for rater participation from November 27, 2013 to August 12, 2014. Figure 4.5 illustrates the basic parts and purposes of the test.

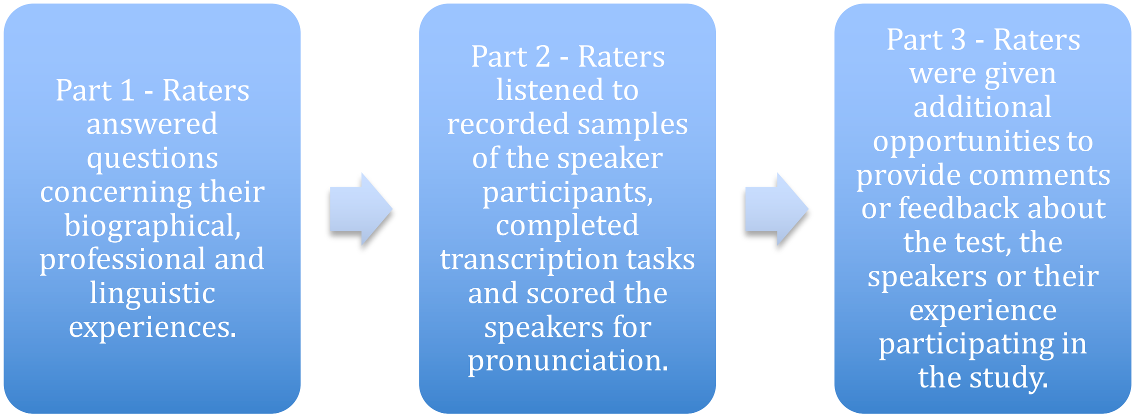


Figure 4.5: Overview of the contents of the three parts of the test

As stated, the test was designed to be able to be completed within twenty minutes. The reasoning for this design aim resulted from lessons learned during the pilot study (discussed in full in Chapter 3). It was determined that twenty minutes was the maximum, realistic amount of time uncompensated, volunteer rater participants could reasonably be expected to dedicate to completing a test. As mentioned in Chapter 3, one hundred, sixty-one potential rater participants began the online pilot study test, but only 87 completed it. Many potential rater participants simply quit after completing part of the test for whatever reasons. Time to complete the study was determined to be a major factor for incomplete tests. If the test for the main study were to have a similar or improved rate of completion, it was decided it should not require more than twenty minutes to complete. Using the time of completion data from the pilot study and consultations with colleagues, the resulting final test was determined 'manageable' to complete within the targeted timeframe. This prediction was based on rater participants being experienced English language teachers (particularly EFL and ESL teachers) and/or linguistics or applied linguistics graduate students or researchers.

This section of the chapter is divided into three subsections that describe the three parts of the test. The first part of the test focused on collecting biographical, professional and linguistic experience data from the rater participants. The second section of the test included all of the instances where the rater participants were requested to listen to audio recordings of the speaker participants, complete transcription tasks and score the speakers for pronunciation. The third and final part of the test included opportunities for the rater participants to offer comments concerning their experiences completing the test, its construction or anything else they wished to impart. A complete copy of the test can be found in Appendix B.

4.5.1 Part one of the test

The primary function of the first part of the test was to provide explanations and instructions, and to collect data concerning the professional, biographical and linguistic experiences of potential rater participants. Part one included two main pages: the overview and instructions page, and a page with questions concerning the rater participants. In the overview and instructions page rater participants were first thanked for their interest in participating in the test, and were then provided with an explanation of the research aim of the study. This was explained as, "the aim of this research is to examine the effects raters' accent-familiarities have on intelligibility success-rates and pronunciation scores on high-stakes tests of English" (see Appendix B) Rater participants were then informed of the outline of the remainder of the test and issued instructions. The instructions included that the test was designed for the average ESL/EFL teacher or researcher to complete in about twenty minutes. Rater participants were also informed that incomplete tests would not be included in the analyses. The participants were encouraged to leave comments about the test, the speakers, the recordings or any other related experiences while completing the test by using the provided comments boxes on each page of parts two and three of the test. Additionally, information pertaining to the content of the recorded samples they would encounter were provided. These were as follows:

> The recordings you will listen to and score are comprised of prepared sentences read aloud. The sentences were influenced by BKB-R materials, (Bench, Kowald & Bamford, 1979); however, the contexts may be complex. The reasoning for this choice of stimuli is to attempt to reduce the effect 'context familiarity' has on comprehension (Gass &

Varonis, 1984), and eliminate correct gap-fill answers resulting from guessing based on context. In short, the listener must be able to cope with the pronunciation of the speaker in order to accurately complete the intelligibility tasks.

Pronunciation scoring should be based solely on the quality of the pronunciation of the speaker. Again, these are not samples of spontaneous speech, so please limit your scoring to only the measure of difficulty you experienced attempting to decipher the phonetic content of the utterances.

(Full copy of the test can be found in Appendix B)

The decision to include such details concerning the aim of the research and the design of the instrument was to reduce the amount of confusion some rater participants experienced completing the pilot study test. As mentioned in the last chapter, rater participants complained that the utterances they listened to 'didn't make sense' even though explanations were provided concerning the content of the prepared sentences. The first page of Part One of the test concluded with rater participants being required to answer the consent agreement request in the affirmative in order to proceed to the next page.

The second page of Part One of the test contained eleven questions (questions 2~12) pertaining to rater participants' biographical data and linguistic experiences (see Table 4.19). The first two questions requested details concerning raters' gender and age range respectively. These questions were included mainly out of custom in research design, and not due to any inclination that they would serve as useful grouping factors to answer the research questions. These questions were not included in the pilot study, and during a Q&A session following a presentation of the results of the pilot study questions were raised as to why they were not included, and that 'some people want to know these details' was expressed. For this reason, these questions were included in the main study.

Table 4.19: Questions concerning the biographical and professional details of the rater participants from Part One of the test

No.	Question
2.	What is your gender?
3.	What is your age?
4.	What is your home country (the country you were raised in)?
5.	What is your first, or native language? (if you have more than one native language please list them according to your personal ranking order)
6.	In what country do your currently reside?
7.	How long have you lived in the country your currently reside?
8.	Do you have ESL/EFL teaching experience?
9.	If teaching, what level of education do you currently teach? (check all that apply)
10.	What is the best description of your education level?
11.	Other than your home country, please list any counties you have lived for one or more years.
12.	How would you best describe your familiarity with the following World Englishes or non-native accents?

This study, as explained in Chapters One and Two adheres to the definition of accent-familiarity described by Browne and Fulcher (in press). We explained that accent-familiarity is a speech perception benefit that is acquired through exposure to different types of linguistic experiences. These varieties of exposure include periods of emersion, or life amongst languages and accents, L2 study, the interlanguage benefits of a shared L1, media exposure and teaching experience. This study aimed to address the potential means of accent-familiarity from the perspective that there is no single or superior means of acquisition. Questions four through seven were designed to gain detailed insight into the rater participants' different means of accent-familiarity acquisition.

The questions concerned with the raters' home country or country they were raised in, and what their first language(s) is were included to provide possible insight into any instances where raters and the speaker participants shared a common L1. Additionally, the resulting data from these questions revealed the rater participants' native English speaker status. As stated, being a native English speaker was not required for this study, though it is worth knowing. As stated previously, most (n=137) of the rater participants reported being native English speakers.

Questions six, seven and eleven provided information concerning accent-familiarity gained through living amongst speakers of different accents and their respective L1s. Though this study focuses on only three (four, if Japanese-English from the pilot study is included) nonnative English accents, these questions pertaining to rater participants' life experiences abroad allow for a better understanding of the overall measure of familiarity the rater population had with nonnative English speech in general.

Questions concerned with raters' ESL/EFL teaching experience (questions eight and nine) were included to gain insight into the rater population's experience with teaching English to nonnative speakers. It was also of interest to know what level(s) of ESL/EFL the raters had. Raters could mark any number of the following answer choices for question nine: higher education (university, college, junior or technical college); business (teaching at a company or business); high school or junior high school; elementary school; no previous teaching experience; other.

It was considered ideal to have a rater population mainly consisting of professional teachers and researchers of linguistics, applied linguistics and language testing. Royal-Dawson and Baird (2009) explain that teaching experience has been deemed a vital prerequisite for rater candidacy in several countries. Though they found teaching experience as a prerequisite to be an unnecessary selection criterion, it is an aspect of the rater population's experience that was worthy of attention. Since teaching experience is considered important for determining the qualification of raters these questions were included. Additionally, since this research could not provide formal rater training (for reasons discussed earlier in this chapter), candidates with such professional and academic backgrounds were considered best suited due to their grounding in language teaching, learning and examination.

It is for these reasons that the resulting data from these questions could help establish the measure of qualification of the rater population.

The twelfth question asked the raters to self-score their measures of familiarity with nine nonnative English accents: Japanese-English; Korean-English; Urdu-English (Pakistan); Arabic-English; Dhivehi-English (Maldives); Brazilian-English; Spanish-English; German-English. To determine their personal measures of familiarity, raters chose for each accent from the scale shown in Table 4.20. Raters chose from four levels of familiarity. Studies like Winke et al. (2012), Xi and Molluan (2009) considered accent-familiarity as 'either/or' (familiar / unfamiliar), and though Carey et al. (2011) collected raters' accent-familiarity data according to four levels, in their analyses the four levels were collapsed to two levels, also 'familiar' and 'unfamiliar.' Since this study considers accent-familiarity acquisition from multiple sources and lengths of exposure, four levels seemed most appropriate. Additionally, the results of the pilot study determined there were discernable differences in both pronunciation scores and intelligibility to warrant the continued use of the four-level scale.

Table 4.20: Rubrics for raters' self-scoring of accent-familiarity

Level of Familiarity	Description
Very Familiar	You are a native speaker of the language of the country or region, have lived in the country or region for 1 or more years, and/or studied the language as a foreign or second language for 1 or more years.
Some Familiarity	You have taught students from the country or region with the accent in the last 2 years; (and/or) have visited the country or region; (and/or) have regular casual contact through encounters in your community, watching TV or movies from the region or any other personal reason to justify a feeling of "some familiarity".
Limited Familiarity	You have heard speakers of the accent but without regularity, and/or have not had students with that accent

	during the last 2 years.
No Familiarity	(no description)

4.5.2 Part two of the test

The second part of the test included all of the rating and transcription tasks the rater participants were asked to complete. This was the main part of the test. It opened with instructions, and was followed by a series of thirteen embedded videos and tasks; the individual series included one video for each speaker participant that provided the speaking sample followed first by transcription tasks and then a prompt to score that speaker. This part of the test was presented in one continuous web page allowing the rater participants the ability to review or revisit the instructions, any of the speakers' audio samples and/or the scores or answers they entered in the transcription exercises without being required to leave that page. What follows are descriptions of different aspects of this part of the test. These subsections include explanations of the instructions to the test, how different elements of part two were constructed, as well as descriptions of the reasoning and motivations of their construction. The section concludes with an explanation of a limitation of the test as a result of using the online survey host Survey Monkey that could not be avoided.

Instructions

Section two began with the instructions to the rater participants concerning how to complete the necessary tasks. The instructions noted that the section included thirteen speakers, and that for each speaker an embedded video was included. The rater participants were informed that they would complete two tasks following the listening to of each speaker's recording: first, completing an incomplete transcript of the speaker's utterance; and second, scoring that speaker for pronunciation by means of selection from the provided pronunciation scale. The pronunciation scale implemented in the test was then presented followed by a recommendation that the test be completed in a quiet room and using headphones. Additionally, the rater participants were informed that guesses were acceptable for the gap-fill tasks, and that it was also acceptable to leave any gap-fill items unanswered if the rater participant had no idea what the answer was. The instructions concluded with the message that the rater participants were required to score each speaker for pronunciation. Figure 4.6 is a screen shot taken of the opening of part two of the test.

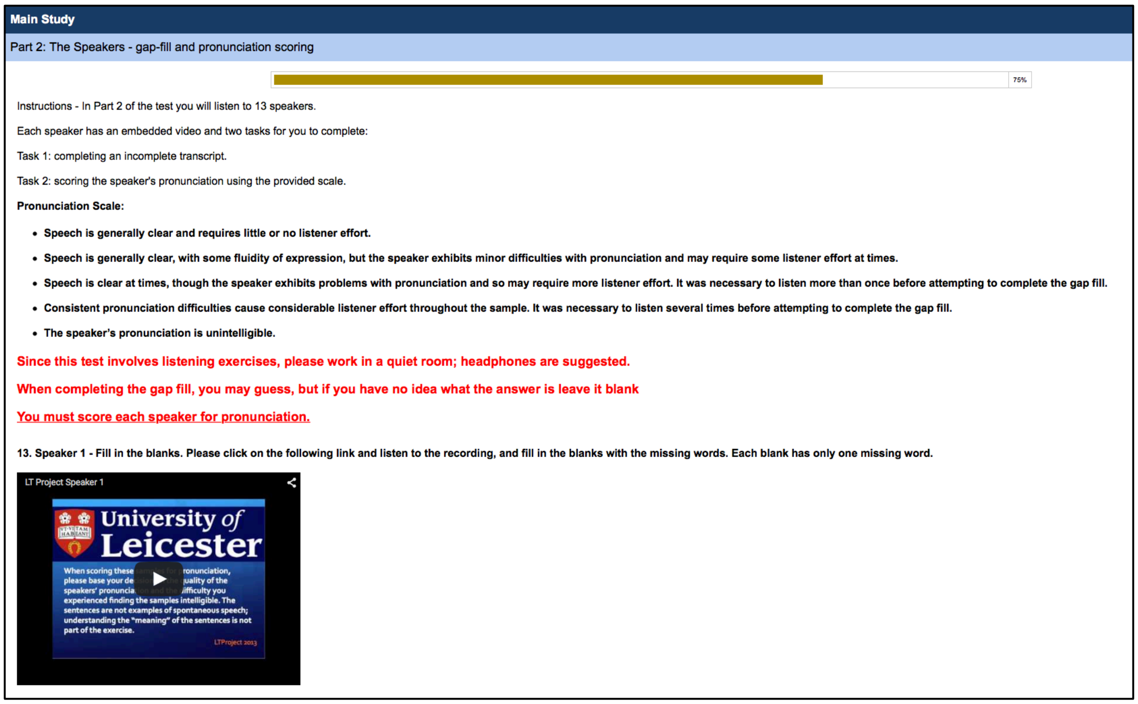


Figure 4.6: Screenshot of the opening and instructions taken from Part 2 of the test

The audio recordings

Similar to the audio recordings of the speaker participants used in the pilot study, all of the recordings were provided using embedded YouTube videos. Each audio file consisted of one speaker participant reading two prepared sentences. Thirteen audio files in all were included and were constructed into videos in a three-phase process.

The first step of the process was the actual recording of the speaker participants in their respective home countries or institutions. All of the recordings were completed using Olympus LS-7 linear Pulse Code Modulation digital recorders in 24 bit/ 96 khz sound quality. Identical recorders were sent to the research assistant(s) at each location for use. The recordings were all completed in either a campus-recording studio (female Arabic-English voices) or in quiet classrooms with only a speaker and a research assistant in the room. The recording of Speaker 1 (the researcher) was conducted by the researcher personally; all other recordings were supervised by research assistants.

The same process was applied for the recording of all samples from the speaker participants from all accent groups. The general aim of the procedures included the following: The speakers from each accent group read the same script designed for their accent. Each script began with the same series of biographical prompts for the candidates to first state their name, home country and age. The scripts then instructed the candidates to read a list of ten sentences designed specifically for their accent. Each script included their respective list of ten sentences two times, so every sentence was read twice by each speaker. This method was selected to prevent any candidate's recording from being excluded due to technical problems, coughs or other unintended deviations from the script. It was also considered a potential means to reduce potential stress and/or anxiety candidates might experience if they were only given one chance to read each sentence aloud. The research assistants managing the recording sessions provided no explanation or instruction concerning proper pronunciation or meanings of the sentences. All of the research assistants that oversaw the recording sessions reported that all of these steps were followed.

As stated, the researcher was not present at the recordings of the nonnative speaker candidates, so could not make determinations of the quality of each recording on sight. All of the original recordings were reviewed by a group of four university EFL lecturers or associate professors. The recordings were all considered to be of a high enough audio quality to be considered acceptable for use in the final test. It was also confirmed that all of the speaker candidates' pronunciation were deemed to be representative of non-native English speech. These research assistants that confirmed the quality of the recordings reported being 'very familiar' or having 'some familiarity' with Spanish-English; 'some familiarity' with Arabic-English; and none had any familiarity with Dhivehi-English.

The second phase of the process to develop the audio samples into videos for use in the test included editing the original audio files. From the resulting thirty files, twelve recordings were selected for editing. These included four files from each of the three accents (two male and two female) and the recording of the researcher (Speaker 1). The editing process was performed in two stages. The first stage employed the Apple GarageBand (version '11) software to select the two sentences selected for each speaker, and remove the remaining sections of the recordings. The resulting edited files were each between 4.3 and seven seconds and included the two target sentences with a 1.5 second pause between. The second stage of the editing process was performed using the Apple iMovie (version '11) software. In this stage the newly edited audio files were converted into video files and normalized for volume. In this step the image that rater participants would see on the embedded videos was added. The image (shown at the bottom of Figure 4.6) included the name and seal of the University of Leicester and the following message to the rater participants:

> When scoring these samples for pronunciation, please base your decision on the quality of the speakers' pronunciation and the difficulty you experienced finding the samples intelligible. The sentences are not examples of spontaneous speech; understanding the "meaning" of the sentences is not part of the exercise.

In order to attempt to prevent rater participants from needing to make volume adjustments for each embedded video while completing the test the audio tracks of each video was edited so all would deliver a consistent measure of volume. The final thirteen videos were played for two colleagues, and deemed to be of an acceptable audio quality delivered in matching volumes.

The final stage of the process of producing the embedded videos involved uploading the iMovie files into YouTube. Once uploaded, the videos could be embedded into the test. Using HTML code in the embedding process the videos were delivered to the rater candidates free from advertising, so no unnecessary distractions would occur during completion of the test.

The sentences

As stated previously, the sentences the speaker participants read were specifically designed for this test. The choice to have the speakers read prepared sentences was made for two reasons. First, extemporaneous speech would make limiting the lengths of the recordings to less than ten seconds but contain a similar amount of speech problematic. Second, the researcher wanted to feature specific characteristics of the three chosen accents that could only be possible through prescribed readings.

Some may argue that the choice to use recorded samples of the speaker participants reading prepared scripts rather extemporaneous speech is inappropriate for this kind of test. It has been argued that reading produces fewer segmental errors than extemporaneous speech (Dickerson & Dickerson, 1977; Wenk, 1979); however, Oyama (1982) found reading to produce more accented speech than extemporaneous speech. And Munro and Derwing (1994) observed no differences of global foreign accent ratings between readings and extemporaneous speech utterances. Additionally, Luoma (2004, p.50) states that reading aloud tests tend to focus on pronunciation, and pronunciation is the focus of this test.

Three different sets of ten sentences were constructed for each of the three included nonnative accents. The different groups of potential rater participants were recorded reading the list of sentences drafted for their accent group two times. From each list of ten sentences, eight were included in the final test resulting in 24 sentences read by nonnative speaker participants. Two sentences were constructed for the native speaker participant. In total, the final version of the test included 26 sentences. There were four main goals concerning the construction of the sentences for the three nonnative accent groups: 1. Similar sentence lengths; 2. Lexical level similarity; 3. Include consonants and/or vowels present in English phonology found to be missing from the target accents' L1; 4. The contexts of the sentences should be complex, or not easily predictable. Table 4.21 shows the full list of sentences in the order they appeared in the test.

Table 4.21: The complete sentence list included in Part 2 of the test

No.	Sentences with intelligibility items underlined
1.	He started with his train.
2.	The ice cream was electric.
3.	The judge laughed at his mother.
4.	The cat parted with his stick.
5.	The tea cloth is quite wet.
6.	The jury learned the law.
7.	The creature travelled quietly.
8.	The passenger stood in a bath.
9.	The jug stood on the shelf.
10.	The washing function broke.
11.	He paid to change his server.
12.	The credit was quite usual.
13.	The machine was awkwardly noisy.
14.	They're staying for improvement.
15.	The old hag challenged the theater.
16.	The huge girl is shouting.
17.	The change demanded some strength.
18.	The agency is transferring along.
19.	The blend is rather different.
20.	The children feared the bridge.
21.	The soldiers swept into the attack.
22.	The patient suffered a seizure.
23.	Playing involved her hands.
24.	The boy wore a patch.
25.	Father looked at the chief.
26.	The farmer managed the bulls.

First, the sentences were to all be between four and six words in length, and comprised of between six and ten syllables. Like the sentences constructed for the pilot study, the primary influence of the design of these sentences was the BKB-R sentence lists (Bench, Kowald & Bamford, 1979) for partially hearing children. The BKB-R sentences also included a similar word length of between three and six words totaling four to six syllables, and were constructed to maintain a similar lexical level of difficulty reflective of the children they aimed to test, namely, children aged eight to fifteen years. This study also aimed for similar lexical levels of difficulty in the sentence construction. The aim was to match the expected levels of English proficiency of the speaker participants, so that the lexical content of the sentences would be familiar enough to the speakers, and not require instruction concerning meaning or pronunciation. This method, it was reasoned would also guarantee that all of the rater participants should be familiar with all of the words included in the test if they were capable of finding the speech intelligible. Additionally, by attempting to maintain a similar level of lexical difficulty of all of the sentences, it was reasoned the difficulty level of the different intelligibility tasks should not be greatly affected simply by word-choice, but due mainly to the quality of pronunciation by the speakers. To maintain the lexical level of all sentences, apart from pronouns, articles and prepositions all of the word choices were selected to be within the top 7,000 most frequently used English words on the Corpus of Contemporary American English. Two exceptions to this rule were included in the test were the words 'jug' /dʒəg/ and 'seizure' /ˈsiːʒər/. 'Jug' and its accompanying sentence "The jug stood on the shelf" was included because it was included in the original BKB-R sentence lists, and it included the voiced postalveolar affricate /dʒ/, which was a targeted phoneme for Spanish-English. 'Seizure' was included because though it was not within the top 7,000 words the fact it contains both a voiced lingua-palatal fricative /ʒ/ and a schwa with voiced postalveolar fricative /ər/ an exception was made (the full sentence was, 'The patient suffered a seizure'). In the case of both sentences that included these exceptions, the contexts of the sentences were less complicated and more predictable than the other sentences in order to attempt to offset the differences in lexical complexity caused by vocabulary choices made outside of the 7,000 most frequently used words. The Corpus of Contemporary American English was chosen because this research refers often to the TOEFL iBT as the main model of testing that is in question, so selecting this corpus seemed most appropriate.

Some researchers may argue that modeling this research instrument, even partly, based on the BKB-R sentence list is inappropriate because of the differences in test subjects and purposes. There are, however, arguments to suggest that important similarities exist that make such an approach appropriate. Those that might oppose this approach may consider it inappropriate to use a test design for measuring sensorineural hearing loss in partial-hearing children to examine pronunciation scoring and intelligibility differences between raters due to differences in levels of accent-familiarity; however, there are similarities between the impairment of sensorineural hearing loss and the benefits of accent-familiarity. For example, sensorineural hearing loss is not an affliction that affects how loudly the listener receives sound, but rather affects the clarity of utterances, and impairs the ability to successfully process speech. According to the American Speech-Language-Hearing Association, sensorineural hearing loss, "reduces the ability to hear faint sounds. Even when speech is loud enough to hear, it may still be unclear or sound muffled"[22]. The positive speech perception benefits of accent-familiarity this research suggests increases the familiar listener's ability to process utterances spoken in familiar accents. In short, sensorineural hearing loss decreases the clarity of utterances, negatively affecting speech processing, and accent-familiarity increases the clarity of utterances, positively affecting how speech is processed and perceived. The theory of the Perceptual Magnet and The Exemplar theory offer evidence to suggest that accent-familiarity trains the listener to more effectively process speech successfully in that accent that might be unintelligible to a listener unfamiliar with that particular accent. To the familiar listener, the content of the speech is clear even if the speaker's delivery

---

[22] Taken from the American Speech-Language-Hearing Association's webpage for Sensorineural Hearing Loss. http://www.asha.org/public/hearing/sensorineural-hearing-loss/

is not executed clearly; whereas, with sensorineural hearing loss a speaker's pronunciation might be clear but the listener's ability to process the speech is impaired. The BKB-R sentence lists were constructed and used to measure this degree of sensorineural hearing loss by observing utterance transcription success rates. Intelligibility tasks of both single words spoken in isolation (e.g. Skinner & Miller, 1983) and intelligibility of complete sentences (e.g. Boothroyd, 1984) have been conducted to measure sensorineural hearing loss. This research, too, utilizes utterance transcription success rates to measure, or determine, if accent-familiarity benefits speech perception processing. Additionally, sensorineural hearing loss is a condition that can be measured according to its severity, similar to how this research suggests accent-familiarity is not necessarily an 'either/or' rater characteristic, but develops and increases with prolonged and varied exposure to the accent.

The sentences constructed for each of the three included accents were also designed to include elements that might effectively reflect pronunciation difficulties for each group of speakers. As stated earlier, the three accents' phonetic inventories were investigated, and missing vowel and consonant phonemes from each that are present in English phonetic inventories were identified. Tables 4.22, 4.23 and 4.24 show the complete sets of sentences constructed for each accent with details of the potentially problematic phonemes. All of the sentences were transcribed into broad phonetic transcriptions in the International Phonetic Alphabet (IPA) using online tools in order to confirm target phonemes present in the sentences. IPA transcriptions were conducted in the following accents: Standard American English (General American)[23] and the Received Pronunciation (RP)[24](IPA transcriptions of each sentence can be found in Appendix B). The reasoning for including both RP and General American accents was to determine any differences between the two most well known accents of native English concerning the targeted phonemes. Since this study is largely influenced by the TOEFL iBT, the General American English

[23] The Standard American English transcriptions were conducted using the online site EasyPronunciation; http://easypronunciation.com/en/english-phonetic-transcription-converter
[24] The Received Pronunciation transcriptions were conducted using the online site PhoTransEdit; http://www.photransedit.com/online/text2phonetics.aspx

transcriptions were consulted as the primary resource. The sentences marked with an asterisk (*)were not included in the final test. The decision to exclude these sentences from use in the final test was not due to any shortcomings in their design, but was based on the quality of the recordings of the speakers.

Table 4.22: Sentences constructed for Spanish-English speaker participants

Sentences	Descriptions of targeted problematic phonemes
The jug stood on the shelf.	Voiced post alveolar affricate /ˈdʒəg/ (jug); /ˈstʊd/ (stood); voiceless lingua-palatal fricative /ˈʃelf/ (shelf)
* The shoes were very yellow.	Voiceless lingua-palatal fricative/ˈʃuːz/(shoes); /ˈveriː/ (very); diphthong approximant /ˈjeloʊ/ (yellow)
The machine was awkwardly noisy.	Voiceless lingua-palatal fricative /məˈʃiːn/ (machine); back, open-mid vowel/ˈɔkˈwɔrdˈliː/ (awkwardly); back rising diphthong; voiceless alveolar fricative /ˈnɔɪziː/ (noisy)
The change demanded some strength.	/ˈtʃeɪndʒ/ (change); cluster of voiced lingua-alveolar stops (/did/) /dɪˈmændɪd/ (demanded); velar nasal & voiceless lingua-dental fricative /ˈstreŋkθ/ (strength)
They're staying for improvement.	voiceless dental fricative /ˈðer/ (they're); velar nasal /ˈsteɪɪŋ/ (staying)
The agency is transferring along.	Front rising diphthong; voiceless post alveolar affricate /ˈeɪdʒənsi/ (agency); front open-mid-open vowel /trænsˈfərɪŋ/ (transferring); velar nasal /əˈlɔŋ/ (along)
*She threw her toy.	Voiceless lingua-palatal fricative/ˈʃiː/(she); voiceless dental fricative/ˈθruː/(threw) /; diphthong from rounded to unrounded, from low-mid back to mid or high front position /ˈtɔɪ/(toy)
The soldiers swept into the attack.	Voiced post alveolar affricate /ˈsoʊldʒərz/ (soldiers); low, front, lax, unrounded with voiceless lingua-alveolar stop (/pt/) /ˈswept/ (swept); low, front, lax, unrounded vowel with voiceless lingua-velar stop /əˈtæk/ (attack)
The washing function broke.	Velar nasal /ˈwɔʃɪŋ/(washing); velar nasal /ˈfəŋ(k)ʃən/(function);
The patient suffered a seizure.	Voiceless lingua-palatal fricative /ˈpeɪʃənt/ (patient); schwa & schwa with voiced post alveolar fricative (/ər/) /ˈsəfərd/ (suffered); voiced lingua-palatal fricative (/ʒ/) schwa with voiced postalveolar fricative (/ər/) /ˈsiːʒər/ (seizure)

Table 4.23: Sentences constructed for Dhivehi-English speaker participants

Sentences	Descriptions of targeted problematic phonemes
The judge laughed at his mother.	Voiced post alveolar affricate, mid-central vowel, voiced post alveolar affricate /ˈdʒədʒ/ (judge); near-open front unrounded vowel /ˈlæft/(laughed); open-mid back unrounded vowel, voiced dental fricative /ˈmʌðr̩ / (mother)
Father looked at the chief.	Voiced dental fricative /ˈfɑːðr̩/ (father); near-close, near-back rounded vowel; /kt/ consonant cluster /ˈlʊkt/ (looked); voiceless post alveolar affricate /ˈtʃiːf/ (chief)
The farmer managed the bulls.	Mid-central vowel /ˈfɑːrmər/ (farmer); near-open front unrounded vowel, mid-central vowel,
*Some structures were under the tree.	Consonant cluster /str/, open-mid back unrounded vowel, voiceless post alveolar affricate, mid-central vowel /ˈstrʌktʃərz/; open-mid back unrounded vowel /ˈʌndr / (under)
The creature travelled quietly.	Consonant cluster /kr/, voiceless post alveolar affricate /ˈkriːtʃr/ (creature); consonant cluster /tr/, near-open front unrounded vowel /ˈtrævld / (travelled); consonant cluster /kw/, near-close near-front unrounded vowel, mid-central vowel /ˈkwaɪətli/ (quietly)
The cat parted with his stick.	Near-open front unrounded vowel /kæt/ (cat); mid-central vowel /ˈpɑːrtəd/ (parted); near-close near-front unrounded vowel /ˈstɪk/ (stick)
*They’re minding society.	Voiced dental fricative /ˈðer/ (they're); near-close near-front unrounded vowel, velar nasal /ˈmaɪndɪŋ/ (minding); mid-central vowel, near-close near-front unrounded vowel, mid-central vowel /səˈsaɪəti/ (society)
The passenger stood in a bath.	Near-open front unrounded vowel, mid-central vowel, voiced post alveolar affricate, mid-central vowel /ˈpæsəndʒər/ (passenger); consonant cluster /st/, near-close near-back rounded vowel /ˈstʊd/ (stood); near-open front unrounded vowel, voiceless dental fricative /ˈbæθ/ (bath)
The old hag challenged the theater.	Near-open front unrounded vowel /ˈhæg/ (hag); voiceless post alveolar affricate, near-open front unrounded vowel, voiced post alveolar affricate /ˈtʃælədʒd/ (challenged); voiceless dental fricative, mid-central vowel / ˈθiːətər/ (theater)
The huge girl is shouting.	Voiced post alveolar affricate /ˈhjuːdʒ/ (huge); voiceless post alveolar fricative, near-close, near-back rounded vowel, near-close near-front unrounded vowel, velar nasal /ˈʃaʊtɪŋ/ (shouting)

Table 4.24: Sentences constructed for Arabic-English speaker participants

Sentences	Descriptions of targeted problematic phonemes
The tea cloth is quite wet.	Mid, central, lax, unrounded vowel /ðə/ (the); initial voiceless stop /ti/ (tea); initial voiceless stop /klɒθ/ (cloth); word-final voiceless stop /kwajt/ (quite) and /wɛt/ (wet); open-mid front rounded vowel /wɛt/
*They're shopping for cheese.	Open-mid front rounded vowel; aveolar approximant /ðɛɹ/ (they're); low, back, tense, unrounded vowel; velar nalsal /ʃɑpɪŋ/ (shopping); postalveolar lateral fricative /tʃiz/ (cheese)
The jury learned the law.	Mid, central, lax, unrounded vowel /ləɹnd/ (learned); postalveolar affricate; high-mid, back, lax, rounded vowel; Aveolar approximant [dʒʊɹi] (jury)
Playing involved her hands.	Bilabial plosive; Velar nalsal /pleɪŋ/ (playing); labidental fricative /ɪnvɑlvd/ (involved); aveolar approximant /həɹ/ (her)
The blend is rather different.	Mid, central, lax, unrounded vowel /ðə/ (the); open-mid front rounded vowel /blɛnd/; aveolar approximant /ɹæðəɹ/ (rather), /dɪfəɹənt/ (different)
He paid to change his server.	Bilabial plosive /peɪd/ (paid); postalveolar lateral fricative; postalveolar fricative /tʃendʒ/ (change); mid, central, lax, unrounded vowel; aveolar approximant /səɹvəɹ/ (server)
*The group dreamt of treasure.	Mid, central, lax, unrounded vowel /ðə/ (the); Aveolar approximant; bilabial plosive /gɹup/ (group); open-mid front rounded vowel /dɹɛmt/ (dreamt); postalveolar fricative /tɹɛʒəɹ/ (treasure)
The boy wore a patch.	Mid, central, lax, unrounded vowel /ðə/ (the); initial voiced stop; back, rising diphthong /bɔi/ (boy); bilabial plosive; postalveolar lateral fricative /pætʃ/ (patch)
The credit was quite usual.	Mid, central, lax, unrounded vowel /ðə/ (the); aveolar approximant; open-mid front rounded vowel /kɹɛdət/ (credit); postalveolar fricative /juʒəwəl/ (usual)
The children feared the bridge.	Mid, central, lax, unrounded vowel /ðə/ (the); postalveolar lateral fricative /tʃɪldɹən/ (children); aveolar approximant /tʃɪldɹən/ (children), /fɪɹd/ (feared), /bɹɪdʒ/ (bridge); postalveolar fricative /bɹɪdʒ/ (bridge)

The pronunciation rating scale

The scale used for pronunciation (see Table 4.25) was similar to the scale used in the pilot study discussed in Chapter 3. Two main differences were made from the pronunciation score descriptors from the pilot test to produce the scale for the main test. First, the choice of wording in the lowest score descriptions were altered to not include the word 'comprehend.' In the pilot study, the description, "cannot comprehend at all" was used. The term 'comprehend' proved problematic for some rater participants, as it often refers to notions of locutionary and/or illocutionary force. Though it was explained that 'comprehension' was interpreted in the study according to Munro, Derwing and Morton's (2006) version of comprehensibility it was decided that it would be better to use 'unintelligible', which is more commonly applied for word recognition. Secondly, the pilot study included details pertaining to the number of times the rater needed to listen to each recording before scoring. This was removed, as the TOEFL iBT descriptors do not include such details, and since raters were free to replay the recordings at their own discretion, multiple playbacks should not necessarily require delivering lower scores.

Table 4.25: The pronunciation score descriptors for the main test

Score	Description
5	Speech is generally clear and requires little or no listener effort.
4	Speech is generally clear, with some fluidity of expression, but the speaker exhibits minor difficulties with pronunciation and may require some listener effort at times.
3	Speech is clear at times, though the speaker exhibits problems with pronunciation and so may require more listener effort. It was necessary to listen more than once before attempting to complete the gap fill.
2	Consistent pronunciation difficulties cause considerable listener effort throughout the sample. It was necessary to listen several times before attempting to complete the gap fill.
1	The speaker's pronunciation is unintelligible.

A limitation of the design

An unfortunate, yet unavoidable limitation of the design of part two of the test was that all rater participants completed the same version of the test. All of the items and speakers were presented in the same order; any possible impact of order effect could not be eliminated. This occurred mainly due to the fact that the online survey host, Survey Monkey, could not successfully randomize the 78 intelligibility gap-fill items and thirteen pronunciation score questions included in part two of the test without corrupting the necessary order and groupings of items allocated for each speaker. The test included thirteen sets of items with one set for each speaker that required a specific item order. Survey Monkey's item randomizing function could not be controlled to successfully randomize only each speaker participant's set of items. The only reasonable means to deliver the test with the item order altered would be to construct multiple versions of the test in Survey Monkey. However, alternative versions of the test could not reasonably be constructed for different groups of rater participants to complete. It was not feasible because rater participant recruitment was conducted via email and other web-based invitations. One invitation was sent to all potential participants with no certainty how many people invited would complete the test. Though the data collection period lasted nine months, the responses were not very spread out. One hundred, forty-one of the total 190 participants completed the test in the first month (November 27, 2013 to December 27, 2013) with 134 of those participants completing the test over a period of just twelve days (December 11 ~ December 23). By that point, implementing multiple versions of the test was not logical. The only reason the data collection period remained open for nine months was in order to recruit more rater participants either very or somewhat familiar with Dhivehi English. Finally, there was no available budget to financially compensate rater participants, which would have made possible the prior division of the rater population into uniform groups for the implementation of multiple versions of the test. While the use of a single version of the test may be considered a shortcoming of the research instrument, it does not diminish the findings or conclusions of this study enough to be considered detrimental. Other similar studies concerned with measuring the effects of rater characteristics on test scores have also used only one set of stimuli presented in the same order to all of their rater participants (e.g. Carey et al.,2011; Xi and Mollaun, 2009). While the order effect could possibly impact the pronunciation scores, it should not account for differences in transcription accuracy, which serves as supporting evidence for the raters' score justification comments. Nevertheless, it would have been preferred to have the rater participants divided into groups of equal sizes and familiarity with the target accents to complete different versions of the test, but this was simply not possible under the circumstances of this study.

4.5.3 Part three of the test

The third part of the test included additional opportunities for the rater participants to provide comments. Though there were optional comment spaces following the gap-fill and pronunciation score tasks for each speaker, it was decided that offering additional opportunities for raters' comments before final submission would be best. These were not considered an unnecessary redundancy, but rather opportunities to gain potential insight into the raters' experiences and opinions in hindsight of their completing the test.

The section included two questions that did not require answers. The questions requested any additional comments concerning the research instrument or their experiences participating in the study. Finally, the rater participants were requested to voluntarily share their email address if they were willing to participate in future research. This marked the end of the test.

4.6 Analyses

Overview

This section of the chapter describes the different analyses procedures conducted. It is separated into two main divisions. First, are descriptions of the Many Facetes Rasch Measurements analyses performed followed by explanations of the other analyses included in the study. The MFRM analyses section is subdivided into two groups: pronunciation score analyses and analyses of the intelligibility task outcome data. As was explained in Chapter 3, the MFRM software, Facets (version 3.71), that was employed in this study is capable of conducting both the pronunciation score and intelligibility data analyses simultaneously; however, when the combined data were examined using Facets, the fit statistics, which provide a kind of quality control measure (Green, 2013) were corrupted. As a result of this corruption, it was advised by Mike Linacre (personal correspondence) that the MFRM analyses be done separately. Linacre advised, “The two analyses do tell two different, but related stories, so my choice would be to keep them separate”. This test was designed to measure raters and rater behavior. Though the speaker participants were scored for pronunciation in a similar manner to regular test candidates on high-stakes tests of spoken English are scored, the focus of this study was to observe how the different rater participants scored the same speaker participants. Scoring speakers for pronunciation is considered a ‘typical rater task’. It is not usual or customary, however, that raters ever be required or requested to complete intelligibility tasks of test candidates’ utterances in actual high-stakes, semi-direct speaking tests. It is this fundamentally different nature of the tasks the rater participants were asked to perform that breaks from any standard model for ‘fit’ in a standard Rasch model of rater performance that makes analysing the two data sets separately more appropriate. The non MFRM analyses were included to compliment the MFRM analyses and provide additional insight into the rater population and their backgrounds.

4.6.1 Many-Facets Rasch Measurement analyses

Many-Facets Rasch Measurement (MFRM) offers a robust insight into dimensions of language tests beyond test candidate performance and item difficulty. MFRM is an expansion of Gorg Rasch’s model developed in the 1950s known as ‘the Rasch model’ (Rasch, 1980). The Rasch model is a dichotomous mathematical model that can be applied to ‘yes/no’, ‘right/wrong’ or other similar type of items. How the Rasch model can be applied to cope with polytomous models like rating scales and also analyse multiple facets will be discussed later in this section, but it is necessary to first explain what the Rasch model is and how it operates in order to understand how it applies to this study.

The Rasch model differs from most statistical modeling. Most statistical models aim to fit the data, but the Rasch model attempts to fit the data to the model. Linacre (2012b) explains:

> Descriptive statistics are based on summarizing the data efficiently and parsimoniously. The data are considered to be given (*Latin* "datum") truth. The statistical model (regression, ANOVA, etc.) is intended to describe the dataset. So a good descriptive statistical model is one which fits the data. If the model misfits the data, then try a different **descriptive** model. Rasch is a **prescriptive** statistical method. The Rasch model gives us what we want (additive measures in a unidimensional framework), so it is our "truth" . . . If the data don't fit the model usefully, then the dataset as a whole doesn't support unidimensional measurement.
>
> (Linacre, 2012b, p. 10)

Generally speaking, the Rash model operates to predict how test-takers will perform on test items based on the test-taker's ability and the level of difficulty of the item. This is referred to as 'fit', and pertains to how the data fits the Rasch model. Bachman (2004) suggests the intent of Rasch analysis, "is to identify individual items and raters that do not fit the model, so that the test developer can then decide on appropriate action" (p. 147). When the Rasch model is illustrated for a test item, the model provides locations of a continuous latent variable of the proficiency of test-takers, and the likelihood, or possibility of success on the item (see figure 4.1). The model utilizes log odds, or 'logits' to predict outcomes. The logit scale calculates success of an item as "1" and failure as "0", and can be seen in the y-axis of Figure 4.7. The x-axis represents the ability of test-takers; "0" reflects the mean ability level of a given test-taker population with abler test-takers receiving positive logit scores and less able test-takers reflected with negative logit scores. The latent variable is expressed as the curving blue line. How the Rasch model works, or what it attempts to demonstrate, is how to predict test-taker performance on any given item. Figure 4.7 shows that a test-taker with a "0" logit ability rating will have "0.5" probability of success, or a 50% chance of success; however, a test-taker with a "1.1" logit ability rating will have a 75% chance of success. The Rasch dichotomous model, "specifies the probability, P, that person n of ability B_n succeeds on item I of difficulty D_i" (Linacre, 2012a, p. 17). Linacre writes the model as:

$$log_e\left(\frac{P_{ni1}}{1 - P_{ni}}\right) = B_n - D_i \tag{1}$$

In short, when the ability measure of a test-taker and the difficulty level of a test item are equal, the odds of the test-taker getting the correct answer is even, or "0.5." Likewise, when the ability of the test-taker and the difficulty of the test item do not match, for example if the item is more difficult than the ability level of the test-taker, the model predicts the test-taker will have a less than even chance, and is reflected as <0.5 probability. Bond and Fox state that the aim of the Rasch approach is:

> to develop fundamental measures that can be used across similar appropriate measurement situations, not merely to describe the data produced by administering Test *a* to Sample *b* on Day *c.* Rasch modeling addresses itself to estimating properties of persons and tests that go beyond the particular observations made during any testing situation" (2007, p.143).

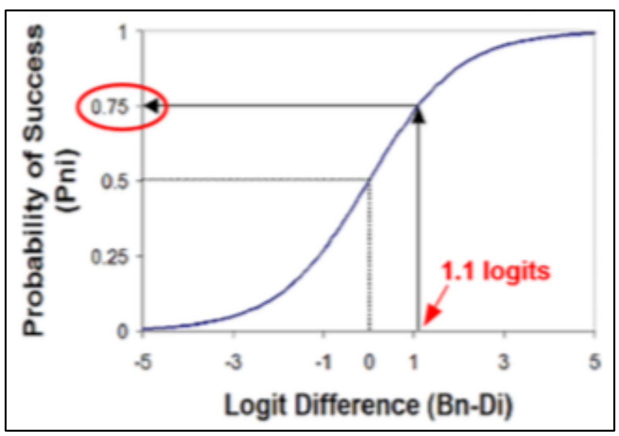


Figure 4.7: A visualization of the Rasch Model. From Many-Facet Rasch Measurement: Facets Tutorial by M. Linacre, 2012a

Though the original Rasch model was created as a dichotomous model, it can be applied to polytomous models. Polytomous models are those like Likert scales and the pronunciation rating scales applied in this research. Known as the "Rasch-Andrich Rating Scale Model" (Linacre, 2012a, p. 19), it estimates the probability (P_{nij}) that a test-taker (*n*) of a certain proficiency level (B_n) will receive an appropriate level score (*j*) on an item (*i*) based on the item's level of difficulty (D_i) rather than the probability ($P_{ni(j-1)}$) of recieving an inappropriate level score (*j-1*). Linacre writes the model as:

$$log_e\left(\frac{P_{nij}}{P_{ni(j-1)}}\right) = B_n - D_i - F_j \qquad (2)$$

The rating scale $\{F_j\}$ is the same for every item

This polytomous model employed by the Facets software can deliver insight into both how the well the raters fit the Rasch model when scoring the speakers for pronunciation, and also whether or not the structure of the rating scales designed for the test aligned with item difficulty appropriately.

With some of the basics of MFRM explained, the different procedures performed using the Facets software can now be detailed. The MFRM procedures included analysing both the pronunciation score data and the intelligibility task data separately. As explained, it was recommended that the analyses of the pronunciation data and intelligibility data be conducted separately (Linacre, personal correspondence). Test raters of speaking tests are not generally required or requested to complete gap-fill intelligibility tasks; such tasks are usually test-taker tasks. And scoring pronunciation is a more 'typical' rater task. If the intelligibility data were included in a single Facets analysis, it would corrupt the raters' fit statistics. The pronunciation score data and its analyses were included to provide details of how the rater participants judged the different speakers' pronunciation abilities. These findings can reveal whether or not raters' differing familiarity levels with the three nonnative accents resulted in construct-irrelevant pronunciation score variance.

Winke et al. (2011, 2013) also used MFRM analyses to determine score variances between raters familiar and unfamiliar with different accents. In that study they examined only TOEFL iBT speaking test scores. Though they determined raters' accent familiarities to affect outcomes, their research did not seek to uncover why accent-familiarity affects trained raters to score familiar accents differently than unfamiliar accents. By gaining insight into intelligibility success differences between accent-familiarity groups could serve to explain why score differences occur. MFRM analyses of the intelligibility task data might provide both additional evidence and an explanation why this threat exists. As stated, this aspect of the design was conceived from personal experiences. These experiences made it evident that accented speech is often more intelligible in familiar accents than unfamiliar accents. Likewise, it seems evident that intelligible speech will be rated for pronunciation more highly than speech that is not. The cause for the differences in scores that have been observed in other studies due to accent-familiarity (e.g. Carey et al., 2011; Winke et al., 2011, 2013; Xi & Mollaun, 2009) may very well be due to intelligibility. This study, however, is the first of its kind to consider and include intelligibility as a factor.

The Pronunciation score MFRM analyses

MFRM analyses of the pronunciation scores the speaker participants received were conducted using the Facets software. Similar to the pilot study, the scores given to the native English speaker participant (Speaker One) were not included in these analyses. Similar to the pilot study, three facets were examined in the analyses, though since there were three different accents to examine, Linacre (personal correspondence) suggested three separate analyses be conducted. In each analysis two facets - the raters and the speakers, and one grouping facet- the raters' familiarity levels of Spanish-English, Arabic-English and Dhivehi-English were examined. These analyses were primarily conducted to gain insight into how the individual raters and groups of raters when divided according to their levels of familiarity with the target nonnative accents 'fit' within a Rasch model. Bachman describes 'fit' as:

> With MFRM, data-to-model fit refers to how closely the scores of test-takers for a particular task rated by a particular rater are to the scores that are predicted by the model – by the parameters for difficulty and severity that are estimated by the model. In other words, in a Rasch analysis, the intent is to identify items and raters that do not fit the model.
>
> (Bachman, 2004, p. 147)

In the case of the MFRM pronunciation score analyses in this study, the 'fit' statistics serve to identify unexpected rater behavior. If, for example, the model predicts a test-taker of a particular proficiency level should receive a particular score (in the case of this study – based on the scores from all of the rater participants) but does not, it is possible the rater will not 'fit' the model. A rater's 'fit' status in this study would likely not be significantly affected if the rater scores only one of the speakers either too harshly or too leniently, but is very effective at identifying raters whose behavior patterns do not fit the model. The analyses output from the Facets software however, provides details of all instances of unexpected ratings.

Interestingly, the Rasch Model does not expect all raters to score all test-takers exactly the same. Within the model, according to Green (2013), "raters are expected to act as independent experts and not be coerced into agreeing with each other" (p. 224). The Facets software can investigate the extent of inter-rater agreement. It provides what are called 'exact agreement statistics' and 'expected agreement statistics.' When a rater population is working as 'independent experts' the actual agreement statistic will be the same or slightly higher than the expected agreement statistic; however, when the exact agreement approaches 100% Linacre (2012b) states, "the raters are behaving the same way as optical scanners do for "bubble sheets". The Raters have become part of the data-collection mechanism, they are no longer a facet of the measurement situation". Facets will determine whether or not the rater participants are acting as independent experts or as "rating machines" (p.26).

Also included in the MFRM analyses of the pronunciation data were bias iteration reports. These analyses measure and determine any bias interaction occurring between individual raters or groups of raters toward any of the speaker participants. In the case of this study, the bias iteration reports were utilized to determine if raters, when divided according to familiarity level of each included accent, demonstrated either positive or negative bias toward speakers of that accent. This research argues that MFRM analyses is the most comprehensive means of analysing the pronunciation score data available, and is most capable of reliably addressing the research questions of this study.

The Intelligibility task MFRM analyses

MFRM analyses of the intelligibility data were also examined using the Facets software. The dichotomous data was collected by means of the transcription tasks rater participants completed in Part 2 of the test, and were included in order to gain insight into how well the rater population coped with the pronunciation of the speaker participants. Each item was scored as either correct or incorrect. The Facets software utilizes the Rasch dichotomous model (see equation 3). Similar to the explanation of the Rasch polytomous model, the dichotomous model functions to predict success or failure on items by different test-takers, or in the case of this study, rater participants based on the estimated difficulty level of the item and the perceived measure of ability of the test-taker. Linacre (2012a) explains,

> Success is a score of "1", and failure is a score of "0" on an item. Then the Rasch dichotomous model specifies the probability, that person n of ability succeeds (scores 1) on an item i of difficulty and similarly is the probability of failure (scores 0).
>
> (Linacre, 2012a, p.17)

In the case of this research, the model can help identify instances when raters or groups of raters are performing outside of what the model predicts. In short, if increased accent-familiarity enhances the ability of raters to successfully transcribe items the model predicts they should not be able to transcribe MFRM analyses will reveal such instances.

$$log_e\left(\frac{P_{ni1}}{1-P_{ni0}}\right) = B_n - D_i \tag{3}$$

addition and subtraction are "additive"

Six separate MFRM analyses were conducted in two stages with the intelligibility data. All analyses investigated three facets similar to the pronunciation score analyses, which included the raters and the items as separate facets, and the familiarity levels of the raters with each of the three nonnative accents as grouping facets. Also similar to the pronunciation score analyses, separate MFRM calculations were conducted for each of the accents. In the first stage, the results from all 79 intelligibility items were examined three times, one analysis for each accent-familiarity grouping. This set of analyses was conducted primarily to observe whether or not differences between raters' familiarity levels with one particular accent reveal significant differences in intelligibility success when coping with a group of speakers from a variety of nonnative accents. The second stage of the analyses examined the items from each nonnative accent group separately. For example, the 24 items from the Dhivehi-English speaker participants were analysed using the raters' familiarity levels with Dhivehi-English as the grouping facet, and likewise for the Spanish-English and Arabic-English items. These analyses were conducted to observe whether or not familiarity with a specific accent significantly benefits raters' abilities to process that particular accent.

4.6.2 Other analyses

Other analyses were conducted to complement the MFRM analyses. These included IPA phonetic transcriptions of the speakers' utterances, and other statistical analyses. These analyses were included to offer additional insight and details into the findings of this study.

Firstly, broad phonetic transcriptions using the International Phonetic Alphabet (IPA) of all speaker participants' utterances were completed. Two colleagues, both with more than ten years of university level ESL/EFL teaching experience, and had completed a graduate level Phonetics and Phonology course at the University of Melbourne volunteered to complete the transcriptions. These two transcriptions were completed independently by the volunteers. The final submissions were not edited by the researcher, and no discussions were held with the two volunteers to recommend changes or explanations for the decisions they made, as this research predicts speech perception is likely not equal between all listeners.

The other statistical analyses included were conducted using SPSS Statistics (version 20.0) to compliment the Facets analyses. These included conducting analyses of variance (ANOVA) and t-tests. ANOVA were conducted to investigate how the raters when divided according to their accent-familiarity levels with each of the three included accents differed scoring each speaker-participant and completing each intelligibility gap-fill item. T-tests were conducted in order to measure differences in pronunciation scores and intelligibility between raters ‘very familiar’ and ‘all other raters’. Correlations were also calculated using SPSS. The results of these analyses and all others described in this chapter will be presented and discussed in the next chapter.

Chapter 5 : Findings and Discussions

In this chapter the results from the research instrument and analyses described in Chapter 4 are presented and discussed. Rather than separating the findings and discussions into different chapters, it was decided that this approach would better serve to provide a smoother narrative of the outcomes. Final conclusions and implications will not be included in this chapter, and will instead be reserved for the next and final chapter. Divided into eight sections, the chapter begins by first repeating the research questions from Chapter 1, as this chapter will flow according to the order of the research questions as they are presented. Section 5.2 defines the alpha level for significance used in the analyses and justifies not including Bonferroni adjustments to the multiple statistical comparisons incorporated in the thesis. Section 5.3 presents the results to establish that the test and rater population performed appropriately. Next, the main research question is answered in section 5.3; namely, did accent-familiarity as a rater characteristic cause significant differences of pronunciation scores? The results from the pronunciation scores of each accent group are presented and discussed separately. Also addressed is whether or not the raters demonstrated bias toward speakers of familiar accents. Section 5.4 is concerned with whether or not accent-familiarity affected intelligibility. These results and discussions are also separated according to each of the three accents examined. The chapter continues by answering the sub-questions concerning whether or not correlations exist between both pronunciation scores and raters' accent-familiarity levels, and intelligibility success with raters' accent-familiarity levels in Section 5.5. This is followed by Section 5.6 that aimed to determine if the effects of raters' differing accent-familiarities were equal among the three accents investigated. Finally, the accent-familiarity levels reported by the rater-participants with nine different World Englishes are examined alongside the global population sizes of each accent in Section 5.7 in order to determine if accent-population can be used for predicting which accents raters are most likely to be familiar with.

5.1 The research questions

Here the research questions are repeated, and presented in the order they are discussed in this chapter:

- Do the raters behave like independent experts?
- Do the raters have the same leniency/severity?
- Does accent-familiarity as a rater characteristic cause significant differences of pronunciation scores?
- Does accent-familiarity cause a positive bias in pronunciation scores?
- Do raters 'very familiar' with a particular accent tend to show the greatest leniency scoring pronunciation in that accent?
- Does accent-familiarity as a rater characteristic cause significant differences in intelligibility success-rates?
- Do raters 'very familiar' with a particular accent tend to show the greatest intelligibility success coping with speech in that accent?
- Is there a correlation between raters' levels of accent-familiarity and pronunciation scores?
- Is there a correlation between raters' levels of accent-familiarity and intelligibility?
- Is the rater accent-familiarity effect more prevalent with some accents than others?
- Do the population sizes of test candidates' L1 affect the likelihood of raters having familiarity with their accents?

5.2 The alpha level and the inclusion of multiple statistical comparisons

Included in this thesis are interpretations of multiple statistical comparisons. As a result, it is necessary to first clarify the alpha level applied for determining significance, and provide some explanation concerning the lack of Bonferroni adjustments. Although, various *p* values will be presented, an alpha level of .05 was applied for all statistical tests. As stated, Bonferroni adjustments were not calculated for all t-tests and ANOVA included in this thesis, so it is possible that one or more of the results of these tests were spuriously significant (see Brown, 2001). However, there are differences of view among social scientists as to their necessity when discussing multiple statistical comparisons (see Perneger, 1998; Siegel, 1990). The lack of Bonferroni adjustments should not represent major challenges to the findings of significance in this thesis, as the t-tests and ANOVA are supported by the FACETS analyses that should reduce any inclinations that the results are spurious.

5.3 Determining the appropriateness of the test and rater population

Before addressing the main research question, it is best to first examine how the test and rater population operated and performed according to their fit to the Rasch Model, and clearly establish what the alpha level was for determining significance. Similar to the pilot study, the pronunciation score data and intelligibility data were anaylsed separately using the Facets MFRM software. Also included in theWhat follows are the findings that establish that the both the test and raters functioned and performed according to the design and within the Rasch model.

5.3.1 The test

In order to determine whether or not the test fit the Rasch model it was necessary to examine the Infit and Outfit mean-squares from different aspects of the test. Infit scores are sensitive to unexpected response patterns from inliers, and Outfit scores likewise sensitive to response patterns from outliers (Green, 2013). Researchers do not all agree concerning what the acceptable values of 'Infit' and 'Outfit scores' are. Green suggests mean-squares between .5 to 1.5 are acceptable; however, McNamarra (1996) advises .75 to 1.3, and Wright and Linacre (1994) recommend .4 to 1.2. The fit statistics of both the test and raters largely fell within the recommendations of Wright and Linacre. The few exceptions that did occur where test items or raters' fit results did not fall within the recommended measures will be shown to be either potentially beneficial or not detrimental to the study.

Since the test was designed to measure how a group of raters score the same group of speakers, the intelligibility items and pronunciation scoring opportunities of the speaker participants' recordings can be treated as test items. Table 5.1 shows the final Facets analyses conducted of the pronunciation scoring items listed from lowest to highest according to their logit measures. Only one of the items, Speaker 8 from the Maldives with an Infit score of 1.56, fell outside of Green's (2013) recommended .5-1.5 measures. Outfit scores were not of real consequence in this test, as outlying raters were expected. What can be determined from speaker 8's Infit results is that inter-rater reliability was very low, as is evident by his ZSTD score of 4.90. ZSTD scores above 2.0 indicate greater than expected variance, and scores below -2.0 indicate less than expected variance, though ZSTD scores can be ignored if Infit mean-squares are acceptable (Green, 2013, p.171). The information in table 5.1, however, does not reveal why such variance was observed. What is clear is that something about the pronunciation or speech performance of Speaker 8 was not equally agreed upon by the raters; such details can possibly be determined by the bias analyses and intelligibility task data. The fact that only one of the speaker's performances failed to produce a pronunciation scoring item to fit the Rasch model is not problematic. Part of the test design was to include different accents and speakers that might cause such problems, as it could be useful and informative to investigate if accent-familiarity might be a reason why inter-rater reliability could be challenged. For this reason, the pronunciation score data for Speaker 8 were not removed.

Table 5.1: Pronunciation scoring item/speaker facet summary statistics

Spkr No.	Accent	Gender	Total Score	Obs. Avrg.	Measure in Logits	Model S.E.	Infit MnSq	ZStd	Est. Discrm.	Corr. PtMea
5	S.E.	M	421	2.24	-1.14	0.11	0.77	-2.40	1.27	0.53
3	A.E.	F	434	2.31	-1.00	0.10	1.39	3.50	0.54	0.35
9	S.E.	F	468	2.49	-0.64	0.10	1.04	0.30	0.93	0.44
2	D.E.	F	469	2.49	-0.63	0.10	1.26	2.40	0.72	0.42
7	S.E.	M	488	2.60	-0.43	0.10	0.82	-1.80	1.19	0.54
4	D.E.	F	495	2.63	-0.36	0.10	0.79	-2.20	1.24	0.54
8	D.E.	M	520	2.77	-0.11	0.10	1.56	4.90	0.37	0.60
6	A.E.	M	581	3.09	0.50	0.10	0.85	-1.50	1.17	0.66
12	A.E.	F	599	3.19	0.67	0.10	1.03	0.30	0.98	0.64
13	D.E.	M	608	3.23	0.76	0.10	0.78	-2.30	1.24	0.66
10	A.E.	M	651	3.46	1.19	0.10	0.75	-2.70	1.27	0.62
11	S.E.	F	651	3.46	1.19	0.10	0.91	-0.90	1.11	0.58
		Mean	532.10	2.83	0.00	0.10	1.00	-0.20		0.55
		S.D.	78.80	0.42	0.79	0.00	0.26	2.50		0.09

Also shown in Table 5.1 are the results of how the speaker participants performed. The table shows from top to bottom the lowest scoring speaker (Speaker 5) at the top to the highest scoring speaker (Speaker 11) at the bottom. The speakers' performance can be determined by both the observed average scores they received, and from their respective logit scores. Based on the quality of the Infit scores of the Speaker-Participants samples used to examine the rater-participants' pronunciation scores, it was determined that the test to measure pronunciation score variance operated as it was intended to operate.

Next it was necessary to determine that the 73 intelligibility items from the twelve nonnative speakers' utterances also fit the Rasch model. The the fit statistics from the Facets analyses shown in Table 5.2 are listed from most difficult (top) to easiest. Like in the pilot study, the six items from speaker 1, the native speaker, were not included in the Facets analyses. The results determine that all but two items (26, 33) were within Wright and Linacre's (1994) .4-1.2 recommend Infit range. Items 26-'jug' (Speaker 5) and 33-'change' (Speaker 6) failed to receive any fit measures because one was not answered correctly by any raters (33), and the other was answered correctly by all raters (26). While both items failed to fit the Rasch model, they were not detrimental to the findings, as their results yielded no data. Section 5.4 of this chapter will include additional descriptions and discussions of the intelligibility items based on the findings of separate MFRM analyses of the items from each accent. The results from the separate analyses' Logit measures for each item may differ from the results shown in Table 5.2, as what is shown in the table presents all 73 items analysed as one test. Based on these findings, the test designed to measure intelligibility operated as intended, but it must next be determined that the rater population was appropriate before attempting to apply the data collected from the test to answer the remaining research questions.

Table 5.2: All intelligibility item Facet summary statistics

Item No.	Spkr No.	Accent	Item	Total Count	Total Score	Obs Avrg	Difficulty Measure in Logits	Model S.E.	Infit MnSq	ZStd
26	5	SE	jug	189	0	0.00	-8.09	1.83	**	
30	10	SE	function	189	15	0.08	-4.05	0.27	1.07	0.3
60	2	AE	feared	189	16	0.09	-3.98	0.27	1.03	0.2
10	3	DE	cat	189	18	0.11	-3.84	0.25	1.10	0.5
18	5	AE	learned	189	21	0.08	-3.66	0.24	0.97	-0.1
25	4	DE	bath	189	25	0.13	-3.45	0.22	1.01	0.1
54	4	SE	transferring	189	27	0.17	-3.36	0.21	1.11	0.7
23	9	DE	passenger	189	33	0.14	-3.10	0.20	1.16	1.2
32	6	AE	paid	189	40	0.21	-2.84	0.19	1.10	0.9
39	7	SE	awkwardly	185	41	0.22	-2.79	0.18	1.11	1.0
55	9	SE	along	189	45	0.24	-2.68	0.18	1.11	1.1
50	10	SE	change	184	54	0.33	-2.38	0.17	0.97	-0.3
41	13	SE	they're	189	58	0.37	-2.29	0.17	0.98	-0.2
56	9	AE	blend	189	62	0.29	-2.18	0.16	1.18	2.8
42	7	SE	staying	182	66	0.31	-2.02	0.16	1.07	1.1
76	7	DE	chief	189	69	0.36	-2.00	0.16	1.05	0.8
7	2	DE	judge	189	85	0.45	-1.60	0.16	1.01	0.1
66	12	SE	suffered	189	89	0.47	-1.50	0.16	0.97	-0.6
73	2	AE	patch	189	89	0.54	-1.50	0.16	0.95	-0.9
12	11	DE	stick	189	103	0.47	-1.16	0.16	1.08	1.4
19	3	AE	law	186	105	0.57	-1.08	0.16	0.99	-0.1
68	8	AE	playing	189	114	0.61	-0.88	0.16	0.95	-0.8
48	12	DE	girl	189	115	0.61	-0.85	0.16	0.84	-2.7
78	13	DE	managed	189	117	0.62	-0.80	0.16	0.99	0.0
65	8	SE	patient	189	118	0.63	-0.78	0.16	1.22	3.2
20	4	DE	creature	189	120	0.64	-0.72	0.16	1.11	1.6
46	11	DE	theatre	189	120	0.62	-0.72	0.16	1.04	0.5
67	6	SE	seizure	189	130	0.69	-0.45	0.17	0.94	-0.8
36	10	AE	quite	188	130	0.71	-0.42	0.17	1.13	1.6
27	8	SE	stood	189	133	0.72	-0.36	0.17	1.02	0.2
45	3	DE	challenged	189	135	0.72	-0.30	0.17	0.88	-1.4
61	3	AE	bridge	187	134	0.74	-0.30	0.17	1.06	0.6
14	11	AE	cloth	189	136	0.69	-0.27	0.17	0.84	-1.8
64	8	SE	attack	189	136	0.76	-0.27	0.17	0.99	0.0
53	13	SE	agency	189	137	0.76	-0.24	0.18	0.97	-0.2
16	5	AE	wet	186	137	0.70	-0.19	0.18	0.79	-2.4
44	3	DE	hag	189	143	0.77	-0.05	0.18	0.76	-2.6
79	11	DE	bulls	189	143	0.72	-0.05	0.18	1.08	0.8
13	9	AE	tea	189	145	0.72	0.02	0.18	0.91	-0.8
17	3	AE	jury	189	149	0.79	0.16	0.19	0.95	-0.3
8	2	DE	laughed	189	150	0.79	0.20	0.19	0.87	-1.1
11	2	DE	parted	189	152	0.81	0.27	0.20	1.04	0.3

Table 5.2 (continued): All intelligibility Facet summary statistics

Item No.	Spkr No.	Accent	Item	Total Count	Total Score	Obs Avrg	Difficulty Measure in Logits	Model S.E.	Infit MnSq	ZStd
71	12	AE	boy	189	152	0.81	0.27	0.20	0.99	0.0
28	3	SE	shelf	189	154	0.83	0.35	0.20	1.07	0.5
15	6	AE	quite	186	155	0.84	0.47	0.21	0.83	-1.3
62	5	SE	soldiers	189	157	0.81	0.47	0.21	0.99	0.0
40	8	SE	noisy	189	158	0.86	0.52	0.21	0.98	-0.1
37	11	AE	usual	188	158	0.83	0.56	0.21	0.90	-0.7
47	12	DE	huge	189	163	0.86	0.75	0.22	1.01	0.1
72	7	AE	wore	189	163	0.84	0.75	0.22	0.84	-1.0
52	8	SE	strength	189	164	0.89	0.80	0.23	1.00	0.0
29	9	SE	washing	189	166	0.87	0.91	0.23	0.95	-0.2
49	4	DE	shouting	189	168	0.90	1.02	0.24	0.91	-0.4
24	12	DE	stood	189	170	0.90	1.15	0.25	0.90	-0.4
69	5	AE	involved	189	170	0.88	1.15	0.25	0.85	-0.8
51	12	SE	demanded	189	171	0.91	1.21	0.26	0.99	0.0
70	13	AE	hands	189	171	0.91	1.21	0.26	1.06	0.3
75	2	DE	looked	189	172	0.92	1.28	0.27	1.12	0.6
35	6	AE	credit	188	172	0.92	1.35	0.27	1.00	0.0
9	9	DE	mother	189	173	0.90	1.35	0.27	1.05	0.3
63	6	SE	swept	189	174	0.93	1.43	0.28	0.98	0.0
34	11	AE	server	189	175	0.92	1.51	0.29	0.94	-0.2
31	5	SE	broke	189	179	0.95	1.89	0.33	0.96	0.0
21	4	DE	travelled	189	183	0.97	2.45	0.42	0.99	0.0
43	13	SE	improvement	189	184	0.98	2.65	0.46	0.91	0.0
57	10	AE	rather	189	185	0.98	2.88	0.51	0.95	0.0
58	10	AE	different	189	185	0.98	2.88	0.51	1.01	0.1
59	10	AE	children	189	185	0.98	2.88	0.51	0.98	0.1
77	7	DE	farmer	189	185	0.97	2.88	0.51	0.99	0.1
74	13	DE	father	189	186	0.98	3.18	0.59	1.02	0.2
22	4	DE	quietly	189	187	0.99	3.60	0.72	0.98	0.1
38	7	SE	machine	189	187	0.99	3.60	0.72	0.97	0.1
33	6	AE	change	189	189	1.00	5.52	1.83	***	0.2
		Mean		188.6	129.7	0.69	0.04	0.32	0.99	0.1
		S.D.		1.2	54.2	0.29	2.32	0.34	0.09	0.1

*Note: **= minimum possible score; ***=maximum possible score*

5.3.2 The rater population

This section attempts to answer two of the sub-research questions concerned with the performance of the rater-participants. Namely, did the raters behave like independent experts, and did they have the same measures of severity and leniency?

To answer these questions, the results from the 'rater' facet examined in the MFRM analyses are presented and discussed. Since the data included in the three separate MFRM analyses examining pronunciation scoring for the 'rater' facet were identical, it is not necessary to present the three separate results to answer these two research questions. The results from the analyses of the 'rater' facet were arguably identical in all three analyses with only minor differences in individual measures usually occurring at the .01 or .001 level. The results used in this section were taken from the Spanish-English (see Appendix B for all of the output files from the three Facets analyses). As explained in Chapter 4, separate Facets analyses were conducted to measure differences in the grouping facets – one for each of the three accents included in the study.

The analyses of the 'rater facet' revealed evidence to suggest that the rater participants did act like independent experts and not rating machines, which answers the first question. This is demonstrated by the *inter-rater agreement opportunity outcomes* that can be used to determine such rater behavior. According to Green:

> The Rasch model wants raters to be independent experts. In such cases the exact agreement statistics is expected to be same or slightly more than the expected agreement statistic.
>
> (Green, 2013, p.224).

The results revealed that 'exact agreements' (n=21841; 34.5%) were higher than 'expected agreements' (n=20764.2; 32.2%), though whether or not they could be deemed as only 'slightly higher' could be debated. Nevertheless, these were untrained raters, and yet their agreement statistics reflect a group that agreed with each other more than the model expected. A common aim of rater training is to achieve higher instances of rater agreement and consistency (see McNamara, 2000; Luoma, 2004; Bachman and Palmer, 1996; 2010). As a result, I feel that the rater population did act as independent experts, and that their lack of rater training should not be considered as detrimental to the reliability of the results based on rater-agreement arguments.

The results determined, however that the rater participants did not all share the same measures of leniency when scoring the speakers' pronunciation. This finding was expected. Table 5.3 provides the results concerning leniency measures from the first Facets analyses conducted with the rater participants' pronunciation scores. The table shows a sampling of the fifteen most severe raters at the top and fifteen most lenient raters toward the bottom (separated by vertical lines |||) (See the full Facets report in Appendix B). Of the 190 rater participants, 22 demonstrated rating behavior outside of Wright and Linacre's (1994) .4-1.2 recommended fit statistics. In a normal testing situation, any such raters and their scores might be eliminated from the data; however, since this test was not designed to make validity arguments of inferences from the pronunciation scores, and is instead specifically interested in the instances of variance in the data, only two of the 22 were removed.

Table 5.3: Rasch analyses of all rater participants' pronunciation ratings

Rater No.	Total Score	Obs. Avrg	Logit Meas.	Model S.E.	Infit MnSq	ZStd	Outfit MnSq	ZStd	Estim. Discrm	Obs. Agrement%	Exact Agrement%
138	20	1.67	-2.64	0.47	0.37	-1.9	0.43	-1.7	1.70	22.2	21.7
24	20	1.67	-2.38	0.47	0.57	-1.1	0.59	-1.0	1.51	25.4	26.9
145	22	1.83	-2.35	0.44	0.31	-2.3	0.35	-2.1	1.79	21.1	22.0
183	22	1.83	-2.35	0.44	1.10	0.3	1.18	0.5	0.80	23.3	22.0
75	23	1.92	-2.15	0.43	2.25	2.4	2.30	2.6	-0.69	25.1	24.0
65	24	2.00	-1.96	0.43	0.44	-1.6	0.45	-1.6	1.65	27.3	25.6
72	22	1.83	-1.96	0.44	2.08	2.2	2.10	2.2	-0.34	18.0	29.8
162	22	1.83	-1.96	0.44	1.00	0.1	1.03	0.1	0.97	28.5	29.8
147	24	2.00	-1.84	0.43	0.80	-0.4	0.77	-0.5	1.21	29.0	28.4
111	25	2.08	-1.67	0.42	0.87	-0.2	0.94	0.0	1.11	27.5	29.8
5	26	2.17	-1.62	0.41	1.56	1.3	1.65	1.5	0.21	27.7	27.9
53	26	2.17	-1.61	0.41	0.93	0.0	0.94	0.0	1.03	29.1	28.6
151	26	2.17	-1.61	0.41	1.04	0.2	1.03	0.1	0.97	31.0	28.6
180	26	2.17	-1.61	0.41	0.47	-1.5	0.47	-1.5	1.61	31.2	28.6
40	24	2.00	-1.59	0.43	1.08	0.3	1.04	0.2	0.97	30.7	31.9
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
123	42	3.50	0.79	0.38	0.85	-0.2	0.85	-0.2	1.16	33.7	30.0
149	42	3.50	0.79	0.38	0.96	0.0	0.96	0.0	1.07	33.5	30.0
179	43	3.58	0.94	0.39	0.97	0.0	0.96	0.0	1.04	24.8	28.1
157	43	3.58	1.06	0.39	0.17	-3.3	0.16	-3.4	1.96	29.0	26.2
21	44	3.67	1.09	0.39	1.00	0.1	0.98	0.0	1.01	31.0	27.5
177	42	3.50	1.17	0.38	1.27	0.7	1.24	0.7	0.74	19.7	22.5
31	44	3.67	1.21	0.39	1.39	1.0	1.32	0.8	0.55	23.5	24.8
187	44	3.67	1.21	0.39	2.04	2.2	1.99	2.1	-0.23	27.7	24.8
152	45	3.75	1.24	0.39	0.78	-0.4	0.82	-0.3	1.19	27.0	25.6
109	45	3.75	1.24	0.39	0.43	-1.7	0.42	-1.8	1.66	31.1	26.2
158	46	3.83	1.40	0.40	3.03	3.5	2.98	3.5	-1.36	14.7	24.3
175	44	3.67	1.47	0.39	1.10	0.3	1.05	0.2	0.92	19.7	19.5
97	47	3.92	1.56	0.41	1.45	1.1	1.40	1.0	0.62	19.4	23.2
171	49	4.08	1.90	0.42	2.54	2.9	2.42	2.7	-0.67	17.4	19.8
22	49	4.08	2.02	0.42	1.86	1.8	1.71	1.6	0.21	16.2	17.1
Mean	34	2.83	-0.32	0.39	1.0	-0.20	1.0	-0.2			
S.D.	5.6	0.47	0.83	0.02	0.7	1.40	0.7	1.4			

Table 5.4 shows the 22 outlying raters from the first analyses. As stated earlier, additional detail can be gained concerning rater behavior from ZSTD values. Values above 2.0 indicate greater than expected variance (unpredictable behavior), and negative values below -2.0 suggest less than expected variance (predictable behavior) (Green, 2013). Ultimately raters 188 and 189 had to be removed from the data set for pronunciation score analyses. While this was very unfortunate because both raters were 'very familiar' with Dhivehi-English (only seven participants in total were 'very

familiar' with Dhivehi-English), it was necessary because the scores they delivered caused too great an impact to the mean scores of their already small Divehi-English familiarity rater subgroup (particularly concerning bias toward Dhivehi-English speakers' pronunciation scores). Both raters are, however, included in discussions later in this chapter, and their data from the the intelligibility task analyses were not removed. Their results from the intelligibility analyses (included in section 5.4.3 of this chapter) provide telling details not only concerning why their rater fit statistics were as they were, and present compelling evidence supporting the hypotheses of this research. Also included in Table 5.4 are details concerning the outlying raters' linguistic, geographic and ESL teaching experience. This information is included in the table, as the findings illustrate that all raters, irrespective of background, evidence variation.

Table 5.4: Outlying rater participants' details (before removing the outliers)

Rater No.	Birth Country	First Language	Country of residence	ESL/EFL teaching experience	Obs. Avrg	Logit Meas.	Model S.E.	Infit MnSq	ZStd	Estim. Discrim	Obs. Agremnt %	Exact Agremnt %
138	Japan	Japanese	Japan	Uni	1.67	-2.64	0.47	0.37	-1.9	1.70	22.2	21.7
145	Spain	Spanish	Spain	Uni & HS	1.83	-2.35	0.44	0.31	-2.3	1.79	21.1	22.0
75	Taiwan	Mandarin	USA	NO ESL	1.92	-2.15	0.43	2.25	2.4	-0.69	25.1	24.0
72	India	English	France	Uni	1.83	-1.96	0.44	2.08	2.2	-0.34	18.0	29.8
5	US	English	Japan	Uni	2.17	-1.62	0.41	1.56	1.3	0.21	27.7	27.9
33	UK	English	UK	Uni	2.25	-1.45	0.41	0.18	-3.1	1.94	36.5	29.8
39	UK	English	Scotland	NO ESL	2.42	-1.13	0.40	0.28	-2.4	1.85	36.5	31.1
118	Australia	English	Australia	Uni	2.50	-0.97	0.39	1.61	1.4	0.33	31.7	32.8
82	UK	English	UK	Uni	2.50	-0.85	0.39	0.19	-3.0	1.87	40.6	34.1
173	US	English	Spain	HS	2.67	-0.67	0.39	2.39	2.7	-0.57	31.6	32.9
188	Maldives	Dhivehi	Maldives	Ele	2.83	-0.25	0.38	5.85	6.2	-4.42	9.7	34.4
189	Maldives	Dhivehi	Maldives	HS&JHS	2.83	-0.25	0.38	5.62	6.1	-4.13	16.8	34.4
129	Spain	French	Spain	Uni	3.00	-0.08	0.38	0.38	-1.9	1.70	41.9	34.3
124	UK	English	UK	NO ESL	3.00	0.04	0.38	0.33	-2.2	1.76	39.4	33.5
107	UK	English	UK	Uni	3.17	0.20	0.38	0.38	-1.9	1.66	36.3	32.4
130	S. Africa	English	UK	Uni	3.25	0.35	0.38	0.39	-1.9	1.67	40.7	32.8
70	UK	English	UK	Uni	3.50	0.79	0.38	0.37	-2.0	1.76	39.0	30.0
157	Australia	English	England	Uni	3.58	1.06	0.39	0.17	-3.3	1.96	29.0	26.2
187	Maldives	Dhivehi	Maldives	Ele	3.67	1.21	0.39	2.04	2.2	-0.23	27.7	24.8
158	UK	English	UK	Uni	3.83	1.40	0.40	3.03	3.5	-1.36	14.7	24.3
171	Spain	Spanish	Spain	HS&JHS	4.08	1.90	0.42	2.54	2.9	-0.67	17.4	19.8
22	UK	English	England	NO ESL	4.08	2.02	0.42	1.86	1.8	0.21	16.2	17.1

Key: Uni.= University; HS= High School; JHS= Junior High School; Ele= elementary school; NO ESL= No ESL teaching experience

After the two outlying raters were removed all of the Facets analyses of the pronunciation scores were conducted a second time. The removal of 188 and 189 did slightly adjust the final fit statistics of the remaining rater participants, though no additional raters were considered for removal. From these new analyses additional information concerning severity/leniency was gained.

The results of a Pearson's chi-square included in the Facets analyses rejected the null-hypothesis that all raters would demonstrate the same measure of severity. It was determined that at the least, the most lenient and most severe raters' pronunciation scores were significantly different ($\chi^2(187)=842.4$, $p=.00$). The results also included a rater separation index of 1.96 (G), which as mentioned in Chapter 3, can be applied to the formula (4G+1)/3 to determine a strata index of 2.95 occurred. These results provide additional detail to the answer to the research question concerning whether or not the raters evidenced the same measure of severity. The findings suggest that not only did the raters demonstrate severity variance, but that the rater-participants could be divided into three distinct levels of severity. These calculations also included a reliability measure of the separation index to be .79. A reliability score of 1.0 would suggest that all raters demonstrated non-interchangeable severity, and a score of zero that all of the raters had interchangeable severity. The .79 indicates a high level of reliability that rater severity variance occurred.

The next section of this chapter will attempt to answer the main research question, and determine if raters' differing accent-familiarity levels resulted in significant differences of pronunciation scores. It will also address whether or not the different accent-familiarity groups of raters demonstrated bias toward speakers with accents they were familiar with, and if the 'very familiar' measure of accent-familiarity resulted in the highest measures of rater-leniency.

5.4 Raters' accent familiarities and pronunciation scores

Now that it has been established that the test and raters functioned and operated as intended, three of the research questions concerning how accent-familiarity levels impacted pronunciation scores. The section is divided into three sub-sections – one addressing each accent researched. First addressed in each subsection is the main research question of this study concerning whether or not accent-familiarity level, as a rater characteristic, caused significant differences of pronunciation scores. Also addressed in this section is whether or not raters' accent familiarities lead to bias, and the research question asking if the raters 'very familiar' with each accent evidenced the greatest leniency scoring pronunciation.

It will be revealed that significant differences of pronunciation scores did occur between rater subgroups for different speakers from each accent. These findings indicate that raters' differing levels of accent-familiarity can cause a significant effect to pronunciation scores. Evidence will be presented suggesting that raters with higher levels of accent-familiarity often leads to both greater measures of leniency and bias scoring the pronunciation of speakers of familiar accents. The 'very familiar' level of accent-familiarity will also be shown to be the most impactful level toward leniency and positive bias, though the instances of significant bias determined were few.

In order to answer the main research question, results from one-way analyses of variance (ANOVA) are presented first to determine if significant differences of mean scores occurred between the four rater subgroups' pronunciation scores for each speaker when divided according to accent-familiarity. Next, the results from independent t-tests performed with each speaker-participant's pronunciation scores are presented that determine if significant variance of scores were observed between raters 'very familiar' with each accent and all other raters. These are followed by the results of the Facets bias interaction reports included to determine if bias occurred.

5.4.1 Spanish-English pronunciation scores

In this section the results from the analyses of the pronunciation scores concerning the Spanish-English speakers are presented and discussed in order to first answer the main research question, and then determine if bias occurred. As explained, the first analyses used to determine the main research question for Spanish-English are ANOVA tests conducted to examine pronunciation score differences between the four accent-familiarity rater-subgroups. Shapiro-Wilk tests revealed the data were not normally distributed, and were positively skewed for speakers 7 and 11 and negatively skewed for speakers 5 and 9. Table 5.5 shows the results of the ANOVA tests, and reveals that significant variance of scores occurred for three of the four speakers, which positively answers the main research question – significant differences of scores did occur, though the effect size was small as indicated by the Eta squared results. Only Speaker 5, the lowest scoring participant of all twelve speakers, failed to show significant variance of scores. The speaker facet summary statistics for Speaker 5 (see Table 5.1 again) showed a ZSTD result of -2.4, which suggests greater than usual rater-agreement occurred (see Green, 2013). This finding could indicate that accent-familiarity levels may have little or less impact on pronunciation scores for low level speakers. Nevertheless, the results of the ANOVA do reject the null hypothesis that accent-familiarity level will not affect scores, at least for Spanish-English.

Table 5.5: One-way analysis of variance (ANOVA) of Spanish-English speakers' pronunciation scores by familiarity level with Spanish-English

Spkr No.		Sum of Squares	*df*	Mean Square	*F*	*p*	Eta Squared
5	Between Groups	2.68	3	0.89	1.79	0.152	0.028
	Within Groups	91.49	183	0.50			
	Total	94.17	186				
7	Between Groups	7.11	3	2.37	4.09	0.008	0.063
	Within Groups	106.00	183	0.58			
	Total	113.11	186				
9	Between Groups	7.12	3	2.37	3.89	0.01	0.060
	Within Groups	111.63	183	0.61			
	Total	118.75	186				
11	Between Groups	10.50	3	3.50	5.16	0.002	0.078
	Within Groups	124.03	183	0.68			
	Total	134.52	186				

While the results of the ANOVA rejected the null-hypothesis, they did not reveal whether or not increased accent-familiarity lead to increased leniency. As explained, this research hypothesizes that the greater the exposure and familiarity with an accent a rater has, the more lenient the rater will be when scoring speakers of that accent for pronunciation. Table 5.6 shows the results of independent t-tests of the Spanish-English speakers' scores between raters 'very familiar' and raters 'less familiar'. All four speakers' mean scores from the 'very familiar' raters were higher than those from the other raters; two speakers' scores were significantly higher. Though the effect sizes were small to medium the findings indicate that at least for Speaker 7, there is a 65% chance that the speaker would receive a higher score if rated by a 'very familiar' rater than a rater with less familiarity, which does represent a significant impact to scores based on who conducts the rating. With this evidence the hypothesis that raters 'very familiar' with a particular accent will demonstrate the greatest leniency seems plausible, at least concerning Spanish-English.

Table 5.6: Independent t-test results examining pronunciation score variance between raters 'very familiar' with Spanish-English and all other raters

	"Very familiar" with Spanish-English			Not "very familiar" with Spanish-English						
	N	*Mean*	*S.D.*	*N*	*Mean*	*S.D.*	*t*	*df*	*d*	
Speaker 5	35	2.37	0.81	152	2.21	0.69	-1.21	185	0.213	
Speaker 7	35	2.97	0.95	152	2.51	0.71	-3.26	185	0.549	**
Speaker 9	35	2.63	0.88	152	2.47	0.78	-1.08	185	0.192	
Speaker 11	35	3.74	0.89	152	3.40	0.83	-2.16	185	0.395	*

*Note: *= p ≤ .05, ** = p ≤ .001*

The question of whether or not bias was observed is next determined. The results of the Facets bias interaction reports are shown Table 5.7. The results indicate what measures of bias were observed from each accent-familiarity rater-subgroup for each of the four Spanish-English speakers. The 'observed scores' column present the total score delivered for each speaker given by all raters of that subgroup. For example, Speaker 5 received an observed score of 83 from the 35 raters in the 'very familiar' subgroup. From that observed score an average score of 2.37 can be calculated, which matches the mean score for the 'very familiar' rater-group shown in Table 5.6. The 'expected score' column in Table 5.7 shows what the Rasch model predicts the speaker's observed score should be. The 'observed expected average' column shows the calculations of the observed scores less the expected scores divided by observed counts revealing positive (when results are positive) and negative (with negative results) bias in terms of the response metric (Linacre, 2012b). Also included is the calculated bias size that determines the measure of positive or negative bias demonstrated by each rater-subgroup for each speaker, as well as a significance factor. Again using the results for Speaker 5 from the 'very familiar' raters, the observed score (81.24) was 1.76 higher than the expected score, revealing that 'very familiar' raters demonstrated some positive bias. The calculated bias size for Speaker 5 is 0.1; p=0.672, which is not a significant measure of bias. Significant bias was observed (p=0.024) only for Speaker 7 by the 'very familiar' rater-subgroup. The mean scores at the bottom of the table also reveal that the 'very familiar' raters demonstrated at least some positive bias scoring the Spanish-English speakers' as a whole, and the other familiarity groupings of raters showed some negative bias scoring Spanish-English pronunciation. Though significant bias was observed toward only one speaker, the results provide some evidence to support the hypothesis that the 'very familiar' level of accent-familiarity were more positively biased than the other rater-subgroups, at least concerning scoring Spanish-English pronunciation.

Table 5.7: Facets bias interaction results for Spanish-English speakers and rater Spanish-English familiarity subgroups

Spkr No.	Spanish English Familiarity level	Obs Score	Expctd Score	Obsrvd Count	Obs-Exp Avrg	Bias Size	Model S.E.	*t*	*d.f.*	*p*	
5	Very	83	81.24	35	0.05	0.10	0.24	0.43	34	0.672	
7	Very	104	93.86	35	0.29	0.55	0.23	2.36	34	0.024	*
9	Very	92	90.11	35	0.05	0.10	0.23	0.45	34	0.658	
11	Very	131	124.01	35	0.20	0.40	0.24	1.64	34	0.110	
5	Some	181	179.97	78	0.01	0.03	0.16	0.17	77	0.869	
7	Some	200	208.48	78	-0.11	-0.21	0.16	-1.32	77	0.191	
9	Some	208	200	78	0.1	0.2	0.16	1.25	77	0.214	
11	Some	268	276.75	78	-0.11	-0.21	0.15	-1.37	77	0.176	
5	Limited	115	120.31	55	-0.1	-0.21	0.2	-1.05	54	0.301	
7	Limited	137	139.66	55	-0.05	-0.09	0.19	-0.5	54	0.619	
9	Limited	124	133.88	55	-0.18	-0.37	0.19	-1.88	54	0.066	
11	Limited	195	187.41	55	0.14	0.26	0.19	1.4	54	0.167	
5	No	42	39.56	20	0.12	0.28	0.33	0.84	19	0.413	
7	No	47	46.04	20	0.05	0.1	0.32	0.31	19	0.760	
9	No	44	44.08	20	0	-0.01	0.33	-0.03	19	0.979	
11	No	57	62.78	20	-0.29	-0.54	0.31	-1.77	19	0.094	
	(all) Mean	126.75	126.76	47.00	0.01	0.02	0.23	0.06			
	S.D.	67.67	68.73	22.50	0.15	0.29	0.06	1.28			
	(Very)*Mean*	102.50	97.31	35.00	0.15	0.29	0.24	1.22			
	S.D.	20.86	18.57	0.00	0.12	0.23	0.01	0.95			
	(Some) *Mean*	214.25	216.30	78.00	-0.03	-0.05	0.16	-0.32			
	S.D.	37.58	42.04	0.00	0.10	0.20	0.01	1.27			
	(Limited) *Mean*	142.75	145.32	55.00	-0.05	-0.10	0.19	-0.51			
	S.D.	35.98	29.21	0.00	0.14	0.27	0.01	1.39			
	(No) *Mean*	47.50	48.12	20.00	-0.03	-0.04	0.32	-0.16			
	S.D.	6.66	10.15	0.00	0.18	0.35	0.01	1.13			

*Note: *p<.05*

5.4.2 Arabic-English pronunciation scores

This section presents the results from the analyses of the pronunciation scores for the Arabic-English speakers, and is structured in the same manner as the previous section. Shapiro-Wilk results determined the data were not normally distributed and were positively skewed for all but Speaker 3 that were negatively skewed. Table 5.8 shows the results of the ANOVA tests used to measure the pronunciation score differences between the four rater-subgroups for each of the four Arabic-English speaker-participants. The results show that between the four rater-subgroups significantly different pronunciation scores were given to two of the four Arabic-Speaker participants', though the effect sizes were smaller than those observed with Spanish-English. These results do indicate that the answer to the main research question concerning whether or not accent-familiarity can cause significant score variance for Arabic-English speakers is, 'yes'. However, the differences in results from the Spanish-English results suggest the effects of accent-familiarity levels may not be generalizable concerning all accents, but accent-dependent. Again, these findings do not provide the details needed to determine if raters with more familiarity scored the speakers more leniently than raters with less familiarity, and only identify that score variance occurred between familiarity level rater-subgroups.

Table 5.8: One-way analysis of variance of Arabic-English speakers' pronunciation scores by familiarity level with Arabic-English

Spkr No.		Sum of Squares	*df*	Mean Square	*F*	*p*	Eta Squared
3	Between Groups	5.88	3	1.96	2.67	0.049	0.042
	Within Groups	134.13	183	0.73			
	Total	140.01	186				
6	Between Groups	7.80	3	2.60	3.39	0.019	0.053
	Within Groups	140.47	183	0.77			
	Total	148.27	186				
10	Between Groups	3.57	3	1.19	1.86	0.137	0.030
	Within Groups	116.95	183	0.64			
	Total	120.52	186				
12	Between Groups	6.59	3	2.20	2.43	0.066	0.038
	Within Groups	165.23	183	0.90			
	Total	171.82	186				

The results from independent t-tests examining the pronunciation scores from the 'very familiar' with Arabic-English raters and all other raters are presented in Table 5.9. Significant differences of pronunciation scores were observed between rater groups only for Speaker 3. This finding does confirm that the 'very familiar' accent-familiarity level concerning Arabic-English did demonstrate significantly higher pronunciation scores, but the results were less than those observed from the Spanish-English findings. Nevertheless, the findings do suggest that the hypothesis concerning raters 'very familiar' with Arabic-English will demonstrate the most leniency toward scoring speakers of the accent's pronunciation is plausible.

Table 5.9: Independent t-test results examining pronunciation score variance between raters 'very familiar' with Arabic-English and all other raters

	"Very familiar" with Arabic-English			Not "very familiar" with Arabic-English						
	N	*Mean*	*S.D.*	*N*	*Mean*	*S.D.*	*t*	*df*	*d*	
Speaker 3	20	2.75	0.97	167	2.26	0.84	-2.43	185	0.54	*
Speaker 6	20	3.15	0.88	167	3.09	0.90	-0.28	185	0.07	
Speaker 10	20	3.50	0.69	167	3.46	0.82	-0.20	185	0.05	
Speaker 12	20	3.30	0.92	167	3.17	0.97	-0.58	185	0.14	

*Note: *= $p \leq .05$*

Concerning whether or not bias was observed by any of the rater-subgroups, the results of the Facets bias interaction analyses are presented in Table 5.10. Though the results do not include any instances of statistically significant bias occurring, the measure of bias raters 'very familiar' with Arabic-English showed for Speaker 3 was very close to being significant (p=0.059). Slightly negative measures of bias were observed for the other three speakers from the 'very familiar' raters due to even greater bias revealed from the raters with 'some familiarity', which fails to confirm the findings from the Spanish-English analyses. The mean scores from the 'some familiarity' rater-subgroups show overall more bias was demonstrated toward the pronunciation scoring of Arabic-English speakers than from the 'very familiar' raters. These findings suggest that less exposure than what determines a rater to be 'very familiar' with an accent can result in positive bias. Of course, the measures of bias determined from these results were not significant, but they do demonstrate that raters' accent-familiarity differences should be considered a threat to pronunciation score reliability by test developers. The effects of raters' accent-familiarity level differences observed toward scoring the pronunciation of Arabic-English and Spanish-

English were not equal. Next the findings from the analyses of the Dhivehi-English speakers' pronunciation scores are presented.

Table 5.10: Facets bias interaction results for Arabic-English speakers and rater Arabic-English familiarity subgroup

Spkr No.	Arabic English Familiarity level	Obs Score	Expctd Score	Obsrvd Count	Obs-Exp Avrg	Bias Size	Model S.E.	*t*	*d.f.*	*p*
3	Very	55	48.62	20	0.32	0.61	0.31	2.00	19	0.059
6	Very	63	64.58	20	-0.08	-0.14	0.30	-0.48	19	0.637
10	Very	70	72.01	20	-0.10	-0.19	0.31	-0.62	19	0.541
12	Very	66	66.51	20	-0.03	-0.05	0.30	-0.15	19	0.880
3	Some	159	163.07	68	-0.06	-0.12	0.17	-0.70	67	0.488
6	Some	227	217.01	68	0.15	0.27	0.17	1.64	67	0.105
10	Some	247	242.17	68	0.07	0.14	0.17	0.81	67	0.420
12	Some	231	223.52	68	0.11	0.21	0.17	1.24	67	0.221
3	Limited	130	135.84	61	-0.10	-0.21	0.19	-1.09	60	0.282
6	Limited	182	182.73	61	-0.01	-0.02	0.17	-0.13	60	0.899
10	Limited	206	205.48	61	0.01	0.02	0.18	0.09	60	0.928
12	Limited	186	188.55	61	-0.04	-0.08	0.17	-0.44	60	0.658
3	No	90	86.55	39	0.09	0.18	0.23	0.80	38	0.428
6	No	109	116.66	39	-0.20	-0.36	0.22	-1.67	38	0.104
10	No	128	131.28	39	-0.08	-0.16	0.22	-0.72	38	0.477
12	No	116	120.4	39	-0.11	-0.21	0.22	-0.96	38	0.345
	(all) Mean	141.56	141.56	47.00	0.00	-0.01	0.22	-0.02		
	S.D.	65.15	63.38	19.53	0.13	0.24	0.06	1.04		
	(Very)*Mean*	63.50	62.93	20.00	0.03	0.06	0.31	0.19		
	S.D.	6.35	10.05	0.00	0.20	0.37	0.01	1.22		
	(Some) *Mean*	216.00	211.44	68.00	0.07	0.13	0.17	0.75		
	S.D.	38.97	33.97	0.00	0.09	0.17	0.00	1.02		
	(Limited) *Mean*	176.00	178.15	61.00	-0.04	-0.07	0.18	-0.39		
	S.D.	32.41	29.81	0.00	0.05	0.10	0.01	0.51		
	(No) *Mean*	110.75	113.72	39.00	-0.08	-0.14	0.22	-0.64		
	S.D.	15.90	19.15	0.00	0.12	0.23	0.01	1.04		

5.4.3 Dhivehi-English pronunciation scores

This section presents the results from the analyses of the the Dhivehi-English speakers' pronunciation scores. The section is also structured like the two previous sections, and begins with the results of ANOVA tests conducted to determine if significant variance of pronunciation scores occurred between the four rater-subgroups for the four Dhivehi-English speakers. Shapiro-Wilkes results determined the data were not normally distributed, and were negatively skewed for all but Speaker 13 that was positively skewed. Table 5.11 shows the results of the ANOVA. Significant differences of scores were determined for three of the four speakers; however, the sizes of two of the four rater-subgroups were very small ('some familiarity', *n*=1; 'very familiar', *n*=5), so the reliability of the ANOVA results is not high. Even so, accent-familiarity was determined to be a significant factor concerning pronunciation score outcomes for the Dhivehi-English speakers, and provides additional empirical evidence rejecting the null hypothesis that raters' differing accent-familiarity levels will not impact pronunciation scores.

Table 5.11: One-way analysis of variance of Dhivehi-English speakers' pronunciation scores by familiarity level with Dhivehi-English

Spkr No.		Sum of Squares	*df*	Mean Square	*F*	*p*	Eta Square
2	Between Groups	15.168	3	5.056	7.278	0.000	0.106
	Within Groups	127.827	184	0.695			
	Total	142.995	187				
4	Between Groups	7.464	3	2.488	4.479	0.005	0.068
	Within Groups	102.211	184	0.555			
	Total	109.676	187				
8	Between Groups	8.762	3	2.921	2.389	0.070	0.037
	Within Groups	224.94	184	1.223			
	Total	233.702	187				
13	Between Groups	10.364	3	3.455	4.992	0.002	0.075
	Within Groups	127.338	184	0.692			
	Total	137.702	187				

In order to determine the outcome of the research question concerning whether or not the raters 'very familiar' with Dhivehi-English showed significant leniency scoring the pronunciation of those speakers than all other raters independent t-tests were conducted. Table 5.12 shows the results, and reveals that significant score variances occurred for all four Dhivehi-English speakers. The effect sizes further suggest that the impact to scores between groups were significant. Though there were only five raters in the 'very familiar' subgroup, which was unfortunate, the findings do suggest that the hypothesis is arguable that raters 'very familiar' with Dhivehi-English will demonstrate the highest measures of leniency scoring the pronunciation of speakers of that accent. The results of the bias interaction analyses provide additional evidence to support this claim.

Table 5.12: Independent t-test results examining pronunciation score variance between raters 'very familiar' with Dhivehi-English and all other raters

	"Very familiar" with Dhivehi-English			Not "very familiar" with Dhivehi-English						
	N	*Mean*	*S.D.*	*N*	*Mean*	*S.D.*	*t*	*df*	*d*	
Speaker 2	5	4.20	0.84	183	2.45	0.83	-4.66	186	2.10	**
Speaker 4	5	3.80	0.84	183	2.60	0.74	-3.56	186	1.52	**
Speaker 8	5	4.00	1.00	183	2.73	1.10	-2.54	186	1.21	*
Speaker 13	5	4.40	0.55	183	3.20	0.84	-3.15	186	1.69	*

*Note: *= p ≤ .05, ** = p ≤ .001*

Table 5.13 shows the bias interaction results from the MFRM analyses for Dhivehi-English familiarity rater-subgroups and the Dhivehi-English speakers. Significant positive bias was determined from the 'very familiar' rater subgroup for Speaker 2, and some positive bias for the other three speakers as evidenced by the observed-expected averages and bias size calculations. The positive bias observed from the 'very familiar' raters toward Dhivehi-English speakers' pronunciation was greater than the results observed from the other two accents. While these results may seem excessively high and suggestive that there must be something wrong with the five 'very familiar' raters with Dhivehi-English, the problem may be due to relying on 'listener effort' measures to score pronunciation. It is possible that these raters were truly not experiencing as much difficulty coping with the pronunciation of the Dhivehi-Speakers than the other raters. After all, all five reported being Dhivehi L1 speakers, which makes this particular group of 'very familiar' raters different from the other two 'very familiar' rater-subgroups included in this study. The results of the intelligibility analyses of the Dhivehi-English items in section 5.4.3 will show that perhaps the 'very familiar' raters were not unjustified for their pronunciation scores.

Table 5.13: Facets bias interaction results for Dhivehi-English speakers and rater Dhivehi-English familiarity subgroups

Spkr No.	Dhivehi English Familiarity level	Obs Score	Expctd Score	Obsrvd Count	Obs-Exp Avrg	Bias Size	Model S.E.	*t*	*d.f.*	*p*
2	Very	21	14.39	5	1.32	2.59	0.69	3.74	4	0.020 *
4	Very	19	15.12	5	0.78	1.45	0.63	2.30	4	0.083
8	Very	20	15.81	5	0.84	1.62	0.66	2.46	4	0.070
13	Very	22	18.15	5	0.77	1.72	0.75	2.29	4	0.084
2	Some	2	2.01	1	-0.01	-0.03	1.49	-0.02	1	0.989
4	Some	2	2.14	1	-0.14	-0.29	1.49	-0.20	1	0.877
8	Some	2	2.26	1	-0.26	-0.54	1.49	-0.36	1	0.778
13	Some	2	2.71	1	-0.71	-1.41	1.49	-0.94	1	0.518
2	Limited	41	43.79	17	-0.16	-0.32	0.34	-0.93	16	0.364
4	Limited	43	46.21	17	-0.19	-0.36	0.34	-1.06	16	0.304
8	Limited	49	48.53	17	0.03	0.05	0.33	0.16	16	0.878
13	Limited	60	56.58	17	0.20	0.38	0.33	1.13	16	0.275
2	No	405	408.83	165	-0.02	-0.05	0.11	-0.42	164	0.676
4	No	431	431.55	165	0.00	-0.01	0.11	-0.06	164	0.953
8	No	449	453.41	165	-0.03	-0.05	0.11	-0.47	164	0.638
13	No	524	530.55	165	-0.04	-0.07	0.11	-0.69	164	0.488
	(all) *Mean*	130.75	130.75	47.00	0.15	0.29	0.65	0.43		
	S.D.	193.85	196.22	70.62	0.52	1.03	0.54	1.48		
	(Very)*Mean*	20.50	15.87	5.00	0.93	1.85	0.68	2.70		
	S.D.	1.29	1.63	0.00	0.26	0.51	0.05	0.70		
	(Some) *Mean*	2.00	2.28	1.00	-0.28	-0.57	1.49	-0.38		
	S.D.	0.00	0.30	0.00	0.30	0.60	0.00	0.40		
	(Limited) *Mean*	48.25	48.78	17.00	-0.03	-0.06	0.34	-0.18		
	S.D.	8.54	5.55	0.00	0.18	0.35	0.01	1.03		
	(No) *Mean*	452.25	456.09	165.00	-0.02	-0.05	0.11	-0.41		
	S.D.	51.13	52.87	0.00	0.02	0.03	0.00	0.26		

*Note: *p<.05*

The results of this section concerning the analyses of the pronunciation scores data has determined that raters' accent-familiarity differences, in all three examined accents, do appear to have an impact on test scores. Therefore, the answer of the main research question, 'does accent-familiarity as a rater characteristic cause significant differences of pronunciation scores?', is 'yes'; raters' accent-familiarity levels can cause significant differences of pronunciation scores.

The results from the analyses also determined the answer to the sub-research question, “does accent-familiarity cause a positive bias in pronunciation scores?” Bias does often occur as a result of increased familiarity, though instances of significant bias were few. It was also determined that it is reasonable to suggest that the ‘very familiar’ accent-familiarity level appears to be the most reliably impactful measure of accent-familiarity toward both leniency and positive bias.

What remains unknown from only examining the pronunciation scores data is why increased accent-familiarity affected scores and lead to bias. The results in the next section are concerned with intelligibility, and reveal for the first time evidence that offers an explanation why raters’ accent-familiarity levels affect pronunciation scores.

5.5 Raters’ accent familiarities and intelligibility

In this section of the chapter, the research questions concerning intelligibility will be addressed and answered. What will be revealed is that raters’ differing accent familiarities caused significant differences in intelligibility success with each accent. It is also determined that raters ‘very familiar’ with each accent demonstrated the greatest benefits to intelligibility among the four familiarity level groups of raters.

Similar to the previous section, the results from the analyses of the items from the three accents will be introduced and discussed separately. The analyses were conducted separately in order to determine what effects raters’ accent familiarities had on each accent. Like with the effects determined from the pronunciation scores, it was considered possible that the accent-familiarity benefits to intelligibility might not be identical for all accents. For this reason, the section is divided into three sub-sections – one for each accent. Conclusions concerning how the findings from all three sub-sections work together to form generalizable statements and implications of how raters’ accent familiarity differences affect intelligibility will be included in the next chapter.

Each sub-section begins by first answering whether or not accent-familiarity as a rater characteristic caused significant differences to intelligibility. In order to answer this question as thoroughly as possible, the results of two different analyses are presented. First, the results from the Facets analyses are provided. These results are useful for determining numerous details about the test and how the raters performed. They also reveal whether or not significantly different averages of intelligibility success occurred between the four accent-familiarity rater-subgroups. To investigate more deeply into how the rater-subgroups' intelligibility differed, ANOVA tests were performed to examine each intelligibility item. From these results all significant differences of means between accent-familiarity rater-subgroups that occurred on any intelligibility items can be determined.

The question concerning whether or not raters 'very familiar' with a particular accent tended to show the greatest intelligibility success coping with speech in that accent is next addressed for each accent. To answer this question, the results of two analyses are presented. First addressed are the results of independent t-tests that were conducted for each intelligibility item from each of the three accents. Specifically, the t-tests measured the intelligibility success differences between the 'very familiar' raters and all other raters to determine if significant differences in intelligibility occurred between the two rater groups. To compliment the t-test findings, the results from the two International Phonetic Alphabet (IPA) transcriptions that were conducted are also presented and discussed. The results of the IPA transcriptions provide visual representations not only of the production of the speaker-participants' speech, but show how differing accent-familiarity levels affected speech perception. It was fortunate that for each of the three accents one or the other research assistant that completed the IPA transcriptions was 'very familiar' with the accent, and the other was not. These phonetic transcriptions effectively highlight numerous instances where significant differences in speech perception were revealed from the other analyses. The combination of results provides compelling evidence not only that raters' differing levels of accent-familiarity affects intelligibility-success rates, but that raters 'very familiar' with a particular accent are more capable of finding speech in that accent intelligible than raters with less familiarity. In many cases, the differences that will be shown were significant.

5.5.1 Spanish-English familiarity and intelligibility

In this section, the results from the analyses of the Spanish-English intelligibility gap-fill items' will be discussed. As explained, the question concerning whether or not raters' accent-familiarity levels with Spanish-English affected intelligibility success will be discussed first, followed by results and discussions concerning how the 'very familiar' group of raters compared with the rest of the raters completing the intelligibility items. The results will show that significant differences in intelligibility occurred between rater-subgroups divided according to accent-familiarity levels with Spanish-English, and that the 'very familiar' raters demonstrated the greatest intelligibility success of any group.

Figure 5.1 shows the variable map from the Facets analyses of the Spanish-English intelligibility items. As explained in the previous chapter, the Facets analyses of the intelligibility items examined three facets – the raters and the test items, and one grouping facet – the raters' familiarity levels with Spanish-English. The variable map is separated into four columns that present the following results and information:

Column 1. The Logit scale is shown in this column, and ranges from -4~5. Again, the scale can be employed when examining the contents of the other columns and noting their horizontal alignments on the logit scale to determine their logit measures.

Column 2. This column shows the results of the rater facet analysis. The 190 raters are represented in the column with the most capable raters appearing toward the top of the scale, and the less capable below. 'Capability', in this regard refers to the raters' ability to successfully transcribe the gap-fill items from the Spanish-English speakers' utterances.

Column 3. The third column shows the 'grouping facet' examined in the analyses. Like in the second column, ability levels are shown from top to bottom. The most notable details in this column are that the 'very familiar' sub-group was the most capable, and the 'no familiarity' sub-group demonstrated the least ability.

Column 4. Here the gap-fill items are shown from easiest at the top, to most difficult at the bottom. Each item is identified in the column with both the target word and which of the four speakers uttered the item. Table 5.14 shows the full report of each intelligibility item from the Facets analyses of the Spanish-English speakers to compliment the variable map. The findings reveal the total rater population experienced a mean intelligibility rate of 61%.

```
+-------------------------------------------------------------------------------------------------+
|Measr|+rater     |+Spanish English Familiarity Level         |+Item                              |
|     |           |(most instances of intelligibility-top)    |(easiest - top)                    |
|-----+-----------+-------------------------------------------+-----------------------------------|
|   5 +           +                                           +                                   |
|     |           |                                           | S7 machine                        |
|     |           |                                           |                                   |
|     |           |                                           |                                   |
|   4 +           +                                           +                                   |
|     |           |                                           |                                   |
|     | .         |                                           |                                   |
|     | .         |                                           | S7 improvement                    |
|     | *         |                                           |                                   |
|   3 +           +                                           +                                   |
|     | .         |                                           |                                   |
|     | **        |                                           |                                   |
|     | *         |                                           | S5 broke                          |
|     | **.       |                                           |                                   |
|   2 + ***       +                                           + S11 swept                         |
|     | ******    |                                           |                                   |
|     | ****.     |                                           | S9 demanded                       |
|     | **.       |                                           | S5 washing                        |
|     | *****.    |                                           | S9 strength                       |
|   1 + ***.      +                                           + S7 noisy                          |
|     | ********. |                                           | S11 soldiers     S5 shelf         |
|     | *         |                                           |                                   |
|     | *****.    | Very Familiar                             |                                   |
|     | ***.      |                                           | S11 attack       S9 agency        |
*   0 * **.       * Limited Familiarity  Some Familiarity     * S11 seizure      S5 stood         *
|     | ***.      |                                           |                                   |
|     | *.        | No Familiarity                            | S11 patient                       |
|     | .         |                                           |                                   |
|     | *.        |                                           |                                   |
|  -1 +           +                                           +                                   |
|     |           |                                           | S11 suffered                      |
|     | .         |                                           |                                   |
|     |           |                                           |                                   |
|     | .         |                                           | S7 staying                        |
|  -2 +           +                                           + S7 they're                        |
|     |           |                                           | S9 change                         |
|     |           |                                           | S9 along                          |
|     |           |                                           | S7 awkwardly                      |
|     |           |                                           |                                   |
|  -3 + .         +                                           +                                   |
|     |           |                                           | S9 transferring                   |
|     |           |                                           |                                   |
|     |           |                                           |                                   |
|  -4 + .         +                                           + S5 function      S5 jug           |
|-----+-----------+-------------------------------------------+-----------------------------------|
|Measr| * = 3     |+Spanish English Familiarity Level         |+Item                              |
|     |           |(least instances of intelligibility-bottom|(most difficult - bottom)          |
+-------------------------------------------------------------------------------------------------+
```

Figure 5.1: The Facets variable map from the analyses of the Spanish-English intelligibility gap-fill items

Table 5.14: Facets Spanish-English intelligibility item measurement report

Item No.	Spkr No.	Accent	Item	Total Count	Total Score	Obs Avrg	Difficulty Measure in Logits	Model S.E.	Infit MnSq	ZStd
26	5	SE	jug	189	0	0.00	-8.00	1.83	**	
30	5	SE	function	189	15	0.08	-3.90	0.28	1.09	0.4
54	9	SE	transferring	189	27	0.14	-3.17	0.22	1.10	0.7
39	7	SE	awkwardly	185	41	0.22	-2.58	0.19	1.13	1.2
55	9	SE	along	189	45	0.24	-2.45	0.19	1.19	1.8
50	9	SE	change	184	54	0.29	-2.15	0.18	1.02	0.2
41	7	SE	they're	189	58	0.31	-2.04	0.17	0.85	-2.0
42	7	SE	staying	182	66	0.36	-1.76	0.17	0.94	-0.8
66	11	SE	suffered	189	89	0.47	-1.19	0.16	0.79	-3.7
65	11	SE	patient	189	118	0.62	-0.42	0.17	1.00	0.0
67	11	SE	seizure	189	130	0.69	-0.08	0.17	0.87	-1.6
27	5	SE	stood	189	133	0.70	0.02	0.18	1.09	1.1
64	11	SE	attack	189	136	0.72	0.11	0.18	0.94	-0.6
53	9	SE	agency	189	137	0.72	0.14	0.18	0.97	-0.3
28	5	SE	shelf	189	154	0.81	0.76	0.21	1.07	0.6
62	11	SE	soldiers	189	157	0.83	0.89	0.21	1.09	0.7
40	7	SE	noisy	189	158	0.84	0.94	0.22	1.04	0.3
52	9	SE	strength	189	164	0.87	1.24	0.23	1.05	0.3
29	5	SE	washing	189	166	0.88	1.36	0.24	0.93	-0.3
51	9	SE	demanded	189	171	0.90	1.69	0.27	0.97	-0.1
63	11	SE	swept	189	174	0.92	1.93	0.30	0.97	0.0
31	5	SE	broke	189	179	0.95	2.45	0.36	1.09	0.4
43	7	SE	improvement	189	184	0.97	3.37	0.52	0.76	-0.4
38	7	SE	machine	189	187	0.99	4.84	1.02	0.76	0.0
		Mean		188.3	114.3	0.61	-0.33	0.33	0.99	-0.1
		S.D.		1.8	59.0	0.31	2.63	0.36	0.12	1.2

*Note: **=maximum possible score*

The variable map shows that the greatest differences in intelligibility occurred between the 'very familiar' and 'no familiarity' groups of raters. Table 5.15 shows the results from the accent-familiarity level measurement report, and reveals that overall the 'very familiar' raters' observed averages were 15% higher than the raters with 'no familiarity'. A Pearson's Chi-square determined that the variance between the groups was significant ($\chi^2(3)=17.3$, p=.00). From these findings alone, the answer to the research question concerning whether or not raters' accent-familiarity differences with Spanish-English would cause significant differences in intelligibility success is answered affirmatively. Accent-familiarity differences did result in significant differences of intelligibility success. The details of which items and from which speakers the most significant differences occurred were determined through other analyses.

Table 5.15: Facets Spanish-English familiarity level report measures

Familarity Level	Total Score	Total Count	Obsvd Avrg	Logit Measur	Model S.E.	Infit MnSq	ZStd
No	245	474	0.52	-0.36	0.13	1.07	1.0
Limited	790	1359	0.58	-0.03	0.08	0.98	-0.4
Some	1144	1847	0.62	0.08	0.07	0.96	-1.0
Very	564	840	0.67	0.31	0.10	1.05	0.8
Mean	685.8	1,130.0	0.60	0.00	0.09	1.01	0.1
S.D.	327.9	519.8	0.06	0.24	0.02	0.04	0.9

ANOVA tests were performed for each intelligibility item in order to determine precisely which items resulted in significant differences, and from which speakers. Table 5.16 shows the nine items that resulted in significant differences from the ANOVA tests. Each of the four speakers' utterances included at least two items that resulted in significantly different intelligibility success rates. The effect sizes determined by the Eta square values suggest that the impact was medium to large (see Cohen, 1988). These findings suggest that the significant intelligibility variance determined by the Chi-square results from the Facets analyses may not have been due to the pronunciation of only one or two of the Spanish-English speakers, but may have been more associated with differences in the abilities of the rater groups to process the Spanish-English accent in general. For these reasons the null hypothesis that raters' differing accent-familiarity levels will not affect intelligibility is rejected.

Table 5.16: Significant results from one-way analyses of variance of the Spanish-English intelligibility items by familiarity level with Spanish-English

Item	Spkr No.		Sum of Squares	*df*	Mean Square	*F*	*p*	Eta Squared
washing	5	Between Groups	1.03	3	0.35	3.47	0.017	0.053
		Within Groups	18.41	185	0.10			
		Total	19.44	188				
broke	5	Between Groups	0.41	3	0.14	2.81	0.041	0.044
		Within Groups	9.06	185	0.05			
		Total	9.47	188				
they're	7	Between Groups	6.46	3	2.15	11.81	0.000	0.161
		Within Groups	33.74	185	0.18			
		Total	40.20	188				
improvement	7	Between Groups	0.20	3	0.07	2.66	0.050	0.041
		Within Groups	4.67	185	0.03			
		Total	4.87	188				
demanded	9	Between Groups	0.81	3	0.27	3.23	0.024	0.050
		Within Groups	15.48	185	0.08			
		Total	16.29	188				
agency	9	Between Groups	2.40	3	0.80	4.19	0.007	0.064
		Within Groups	35.30	185	0.19			
		Total	37.69	188				
soldiers	11	Between Groups	1.21	3	0.40	2.94	0.034	0.046
		Within Groups	25.37	185	0.14			
		Total	26.58	188				
attack	11	Between Groups	2.23	3	0.74	3.87	0.010	0.059
		Within Groups	35.47	185	0.19			
		Total	37.69	188				
suffered	11	Between Groups	1.95	3	0.65	2.65	0.050	0.041
		Within Groups	45.20	185	0.24			
		Total	47.14	188				

The next research question to be considered concerns whether or not the raters 'very familiar' with Spanish-English demonstrated greater intelligibility success than all other raters. To answer this question, the results from independent t-tests are examined that measured the intelligibility differences between the 'very familiar' raters and all other raters for each Spanish-English intelligibility item. Also presented to support the t-test findings in answering the research question are the results of the two IPA transcriptions completed of the Spanish-English utterances. The findings from both the t-tests and the IPA transcriptions show that the 'very familiar' rater-subgroup did experience significantly greater intelligibility success than all other rater groups combined.

The t-tests confirmed that the 'very familiar' raters experienced more intelligibility with nineteen of the 24 Spanish-English. Table 5.17 shows the six items that caused significant intelligibility success variance between rater groups. A surprising finding was that one of the significant items, 'along' – Speaker 9, was significantly more intelligible to all raters not 'very familiar' with Spanish-English; among all 73 items included in this study, this was the only item where the 'very familiar' raters showed significantly less intelligibility than all other raters. One item, 'jug' from Speaker 5, as mentioned earlier in this chapter, was not accurately transcribed by any of the 190 rater-participants. These findings that determine that the raters 'very familiar' with Spanish-English were more successful than all other raters completing 79% of the Spanish-English intelligibility items, and 17% of the items significantly better is convincing evidence to suggest that the 'very familiar' level of familiarity with Spanish-English experiences the greatest intelligibility benefits of the four levels of accent-familiarity included in this study.

Table 5.17: Significant results from independent t-tests measuring the intelligibility differences between the raters 'very familiar' with Spanish-English and all other raters

Item	Spkr. No.	"Very familiar" with Spanish-English			Not "very familiar" with Spanish-English						
		N	*Mean*	*S.D.*	*N*	*Mean*	*S.D.*	*t*	*df*	*d*	
they're	7	35	0.69	0.47	155	0.22	0.42	-5.40	46.7	1.05	**
improvement	7	35	1.00	0.00	155	0.97	0.18	-2.27	154.0	0.24	*
demanded	9	35	0.97	0.17	155	0.89	0.31	-2.13	94.7	0.32	*
agency	9	35	0.94	0.24	155	0.68	0.47	-4.84	103.8	0.70	**
along	9	35	0.09	0.28	155	0.27	0.45	3.09	77.1	-0.48	*
attack	11	35	0.89	0.32	155	0.68	0.47	-3.05	70.2	0.52	*

Note: $*= p \leq .05$, $** = p \leq .001$

The two IPA transcriptions of the Spanish-English speaker-participants' utterances shown in Table 5.17 provide additional evidence supporting the finding that raters 'very familiar' with Spanish-English demonstrated greater intelligibility success than all other raters. Transcription 1 was completed by a research assistant with 'no familiarity' with Spanish-English, and Transcription 2 completed by an assistant 'very familiar' with the accent. Examples best reflecting the intelligibility differences between the two transcriptions include those found in the results for the first sentence from Speaker 9. Not only was the item, 'demanded', perceived as 'the mandate'(/ðə mændeɪt/) in the first transcription, but other curious interpretations of that utterance were perceived by the research assistant with no familiarity with Spanish-English. The second transcription of the same utterance shows that not only did the assistant 'very familiar' with Spanish-English accurately identify all target words, but the results suggest that the 'very familiar' research assistant found Speaker 9's pronunciation highly intelligible. Both assistants listened to the same recordings, but speech-perception was clearly different. It is possible that for the second assistant, The Perceptual Magnet Effect occurred, resulting in a warping of her perception of the utterances to more identifiably accurate interpretations of the target words. Other notable examples of the accent-familiarity benefit visible in the IPA transcriptions are those for: 'function'- sentence two, Speaker 5; 'noisy'- sentence one, Speaker 7; and 'attack'- sentence one, Speaker 11. In all of these examples again the research assistant 'very familiar' with Spanish-English correctly identified the target words, and the speech perception abilities of the assistant with 'no familiarity' were very different. The answer to the research question based on these findings and those of the t-tests determine that the 'very familiar' level of accent-familiarity does demonstrate better intelligibility success than all other levels of familiarity, at least concerning Spanish-English.

Table 5.18: The results of the two IPA transcriptions of the Spanish-English speakers' utterances

Spkr No.	Sentences with intelligibility items underlined	Transcription No. 1	Transcription No. 2
5	The jug stood on the shelf.	/ðə jæps stu:d ɔn ðə ʃelf/	/ðə jɑt stul on ðə ʃelf/
5	The washing function broke.	/ðə wʌʃɪŋ ma:ʃən broʊk /	/ðə wɒʃiŋ fuʃən brəʊk/
7	The machine was awkwardly noisy.	/ðə məʃi:n wɔ:z ɔ:kwəgli: naɪs/	/ðe məʃin wəz aʊʔwɜ˞tli nɔɪsi/
7	They're staying for improvement.	/ðeɪə speɪjɪŋ fɔ impru:vmənt /	/ðei ɑr ə stin fəʊr ɪmpruvmɪn/
9	The change demanded some strength.	/ðeɪ ʃeɪndʒ ðə mændeɪt sɔ:ŋs steŋθ /	/deɪ tʃeɪŋz dɪmɑndɪd sʌm strenθ/
9	The agency is transferring along.	/ðɪ ædʒensi: ɪz tra:sfərɪŋ ʌ laʊn /	/di aɪdʒensi is traʊferiŋ ə ləʊn/
11	The soldiers swept into the attack.	/ðə sɔ:ldʒɜ:ʳz swept ɪntʊ ðeə tænk /	/ðə səʊldʒɜ˞s swep ɪntu ði ətæk/
11	The patient suffered a seizure.	/ðə peɪtənt sɜ:fɜ:d ʌ si:sɜ:/	/ðə peɪʃən sʌfə˞ ə sisɜ˞/

5.5.2 Arabic-English familiarity and intelligibility

In this section, the results from the intelligibility gap-fill items' analyses from the Arabic-English speakers will be discussed. Like in the previous section, the first research question addressed will be whether or not raters' accent-familiarity levels with Arabic-English affected intelligibility success. This will be followed by results and discussions concerning how the 'very familiar' group of raters compared with the rest of the raters completing the items. Significant variance in intelligibility between rater-subgroups was observed, but were limited to only certain items. It will be shown that the measures of intelligibility differences between accent-familiarity levels determined for Arabic-English were less than those observed for Spanish-English. Evidence is shown, however, supporting the notion that the raters 'very familiar' with Arabic-English showed the greatest success with intelligibility of any rater-subgroup, which was similar to the Spanish-English findings.

In order to address whether or not the null-hypothesis can be rejected that raters' accent-familiarity levels with Arabic-English will not affect intelligibility success, the results of the Facets analyses are examined. The variable map from the analyses are shown in Figure 5.2. As explained in the previous section, the Facets analyses of the intelligibility items investigated three facets: the raters, the items and a grouping facet comprised of the raters grouped according to their respective levels of accent-familiarity with Arabic-English. The same descriptions of the columns apply from the previous section.

```
+-----------------------------------------------------------------------------------------------------------------------------+
|Measr|+rater    |+Arabic English Familiarity Level                              |+Item                                        |
|     |          |(most instances of intelligibility-top)                        |(easiest - top)                              |
|-----+----------+---------------------------------------------------------------+---------------------------------------------|
|   5 +          +                                                               + S6 change                                   |
|     | .        |                                                               |                                             |
|     |          |                                                               |                                             |
|     |          |                                                               |                                             |
|     |          |                                                               |                                             |
|   4 + *        +                                                               +                                             |
|     | .        |                                                               |                                             |
|     |          |                                                               |                                             |
|     |          |                                                               |                                             |
|     | ****     |                                                               | S10 children   S10 different   S10 rather   |
|   3 +          +                                                               +                                             |
|     | **.      |                                                               |                                             |
|     | *.       |                                                               |                                             |
|     | **.      |                                                               |                                             |
|     | *****.   |                                                               |                                             |
|   2 + ******.  +                                                               +                                             |
|     | ***.     |                                                               | S6 server                                   |
|     | ***      |                                                               | S6 credit                                   |
|     | ****.    |                                                               | S12 hands      S12 involved                 |
|     | *******. |                                                               |                                             |
|   1 + **.      +                                                               + S12 wore                                    |
|     | *.       |                                                               |                                             |
|     | ****     |                                                               | S3 quite       S6 usual                     |
|     | *.       |                                                               | S12 boy                                     |
|     | *.       | Very Familiar                                                 | S3 jury        S3 tea                       |
*   0 * **.      * Limited Familiarity  No Familiarity       Some Familiarity    *                                             *
|     | **.      |                                                               | S10 bridge     S3 cloth        S3 wet       |
|     | .        |                                                               | S6 quite                                    |
|     | .        |                                                               |                                             |
|     | **       |                                                               | S12 playing                                 |
|  -1 +          +                                                               +                                             |
|     | *        |                                                               | S3 law                                      |
|     | .        |                                                               |                                             |
|     |          |                                                               | S12 patch                                   |
|     |          |                                                               |                                             |
|  -2 + .        +                                                               +                                             |
|     |          |                                                               |                                             |
|     |          |                                                               | S10 blend                                   |
|     |          |                                                               |                                             |
|     |          |                                                               |                                             |
|  -3 +          +                                                               + S6 paid                                     |
|     |          |                                                               |                                             |
|     |          |                                                               |                                             |
|     |          |                                                               |                                             |
|     |          |                                                               |                                             |
|  -4 +          +                                                               + S3 learned                                  |
|     |          |                                                               |                                             |
|     |          |                                                               | S10 feared                                  |
|     |          |                                                               |                                             |
|     |          |                                                               |                                             |
|  -5 +          +                                                               +                                             |
|-----+----------+---------------------------------------------------------------+---------------------------------------------|
|Measr| * = 3    |+Arabic English Familiarity Level                              |+Item                                        |
|     |          |(least instances of intelligibility-bottom)                    |(most difficult - bottom)                    |
+-----------------------------------------------------------------------------------------------------------------------------+
```

Figure 5.2: The Facets variable map from the analyses of the Arabic-English intelligibility gap-fill items

The Facets item measurement reports shown in Table 5.19 provide additional detail to the findings shown in column four of the variable map. The items are listed from most difficult at the top ('feared'; Speaker 10) to the easiest item at the bottom ('change'; Speaker 6). Similar to the Spanish-English item measurement results, each of the four Arabic-English speakers produced items of various levels of difficulty. Speaker 10 received the overall highest average pronunciation scores of all four Arabic-English speakers, yet produced the most difficult intelligibility item. Additionally, Speaker 3 performed least-well in average pronunciation scores, but produced the most consistent measures of intelligibility difficulty scores. Like with the Spanish-English speaker-participants, the difficulty measures of the intelligibility items resulting from the Arabic-English speakers' respective utterances can not be predicted based on the pronunciation scores the speakers received. The first finding that may explain why less intelligibility variance was observed for Arabic-English than with Spanish-English (or Dhivehi-English) shown in Table 5.19 is that raters' overall intelligibility success average was higher (71%) than that of Spanish-English (61%; and Dhivehi-English 69%). This finding also contributes to the hypothesis suggesting that increased intelligibility leads to higher pronunciation scores because apart from Speaker 3, the other Arabic-English speaker-participants scored higher average pronunciation scores than the other three accents (see again Table 5.1).

Table 5.19: Facets Arabic-English intelligibility item measurement report

Item No.	Spkr No.	Accent	Item	Total Count	Total Score	Obs Avrg	Difficulty Measure in Logits	Model S.E.	Infit MnSq	ZStd
60	10	AE	feared	190	16	0.08	-4.33	0.28	1.08	0.4
18	3	AE	learned	190	21	0.11	-3.99	0.25	0.92	-0.4
32	6	AE	paid	190	40	0.21	-3.09	0.19	1.14	1.3
56	10	AE	blend	190	62	0.33	-2.36	0.17	1.3	3.7
73	12	AE	patch	190	89	0.47	-1.61	0.16	1.09	1.3
19	3	AE	law	187	106	0.57	-1.12	0.17	0.95	-0.6
68	12	AE	playing	190	115	0.61	-0.89	0.17	1.09	1.2
36	6	AE	quite	189	131	0.69	-0.39	0.18	1.16	1.7
61	10	AE	bridge	188	134	0.71	-0.28	0.18	1.16	1.6
14	3	AE	cloth	190	137	0.72	-0.22	0.18	0.73	-3.2
16	3	AE	wet	187	138	0.74	-0.14	0.19	0.68	-3.6
13	3	AE	tea	190	146	0.77	0.10	0.19	0.84	-1.5
17	3	AE	jury	190	150	0.79	0.26	0.20	0.91	-0.7
71	12	AE	boy	190	153	0.81	0.38	0.21	1.09	0.7
15	3	AE	quite	187	156	0.83	0.59	0.22	0.73	-2.2
37	6	AE	usual	189	159	0.84	0.69	0.22	1.01	0.0
72	12	AE	wore	190	164	0.86	0.91	0.23	0.85	-1.0
69	12	AE	involved	190	171	0.90	1.33	0.26	0.81	-1.0
70	12	AE	hands	190	172	0.91	1.40	0.27	1.15	0.8
35	6	AE	credit	189	173	0.92	1.55	0.28	1.11	0.5
34	6	AE	server	190	176	0.93	1.72	0.30	0.97	0.0
57	10	AE	rather	190	186	0.98	3.16	0.52	0.87	-0.1
58	10	AE	different	190	186	0.98	3.16	0.52	1.01	0.1
59	10	AE	children	190	186	0.98	3.16	0.52	1.03	0.2
33	6	AE	change	190	190	1.00	5.83	1.83	***	0.2
		Mean		189.4	134.3	0.71	0.23	0.32	0.99	0.0
		S.D.		1.0	50.6	0.27	2.26	0.33	0.16	1.6

*Note: ***=maximum possible score*

Perhaps the most noticeable difference between the variable map for Spanish-English intelligibility and the Arabic-English map, is that in column three it appears as though the raters from the three lower levels of Arabic-English familiarity all performed identically. The variable map shows that the 'very familiar' raters performed the best; however, Table 5.20 reveals exactly how similarly the 'no familiarity' and 'limited familiarity' groups performed, as well as the 'some familiarity' and 'very familiar' groups. Their similarities are such that the two pairs of rater-subgroups' observed averages were identical. The Pearson's chi-square included in the Facets analyses failed to reveal an overall significant difference of intelligibility between the four rater-subgroups ($\chi^2(3)=4.5$, $p=.21$). These results are different from the findings for Spanish-English, as they fail to determine any overall significant differences between rater-subgroups completing the 25 items. This does not, however, mean that there were no individual items that resulted in significant differences of intelligibility between groups.

Table 5.20: Facets Arabic-English familiarity level report measures

Familiarity Level	Total Score	Total Count	Obsvd Avrg	Logit Measure	Model S.E.	Infit MnSq	ZStd
No	683	990	0.79	-0.10	0.09	1.16	2.9
Limited	1065	1546	0.79	-0.10	0.08	0.99	-0.2
Some	1238	1700	0.82	0.08	0.07	0.88	-2.9
Very	371	500	0.82	0.12	0.14	1.07	0.8
Mean	839.3	1184.0	0.81	0.00	0.10	1.02	0.2
S.D.	336.8	475.1	0.02	0.10	0.03	0.10	2.1

In order to determine if any of the individual Arabic-English intelligibility items resulted in significant intelligibility variance between groups the results of the ANOVA tests conducted for each item were examined. Table 5.21 shows the five items that caused significant intelligibility variance between the four rater-subgroups. Of the four Arabic-English speaker-participants only Speaker 3 failed produce any utterances resulting in significantly different intelligibility success. As reflected in Table 5.1, Speaker 3 received the second-lowest average pronunciation scores from all raters. It is possible that raters' accent-familiarity benefits to intelligibility are not as impactful processing lower-level speakers' speech, or it is possible that the accent-familiarity benefits to intelligibility are accent-specific. The greatest differences in intelligibility success between rater-subgroups occurred with the utterances from Speaker 6. The average pronunciation scores determined for Speaker 6 (3.09; also in Table 5.1) suggest he was an intermediate-level speaker, so it is possible the intelligibility related benefits to accent-familiarity with Arabic-English are more impactful at that level.

Though the results of the Chi-square in the Facets analyses failed to reveal significant variance in overall intelligibility between the four rater-subgroups coping with all 25 items, the five items identified in the ANOVA do serve to reject the null-hypothesis. While the total number of items that caused significant variance in intelligibility was fewer than those determined for Spanish-English, significant variance in intelligibility between groups with 20% of the items does represent enough of an impact to intelligibility that could explain why pronunciation score variance was also observed for the Arabic-English speakers between groups. The results concerning how the raters 'very familiar' with Arabic-English differed with all other raters' intelligibility success also demonstrate that the findings for Spanish-English and Arabic-English were not equal.

Table 5.21: Significant results from one-way analyses of variance of the Arabic-English intelligibility items by familiarity level with Arabic-English

Item	Spkr No.		Sum of Squares	*df*	Mean Square	*F*	*p*	Eta Squared
server	6	Between Groups	0.88	3	0.29	4.46	0.005	0.068
		Within Groups	12.08	184	0.07			
		Total	12.96	187				
quite	6	Between Groups	2.77	3	0.92	4.54	0.004	0.069
		Within Groups	37.34	184	0.20			
		Total	40.11	187				
usual	6	Between Groups	1.11	3	0.37	2.82	0.040	0.044
		Within Groups	24.11	184	0.13			
		Total	25.21	187				
blend	10	Between Groups	1.97	3	0.66	3.05	0.030	0.047
		Within Groups	39.59	184	0.22			
		Total	41.55	187				
playing	12	Between Groups	4.13	3	1.38	6.21	0.000	0.092
		Within Groups	40.75	184	0.22			
		Total	44.87	187				

The first findings presented to determine whether or not the raters 'very familiar' with Arabic-English experienced significantly better intelligibility success than all other raters are the results of the t-tests conducted with each Arabic-English intelligibility item. The results determined that the 'very familiar' raters' mean intelligibility success rates were higher than all other raters for 44% of the items (n=11), though significant variance was observed in only three items shown in Table 5.22. The effect sizes of the three significant items were small. Based on these findings alone it is possible to make the argument that the raters 'very familiar' with Arabic-English did demonstrate significantly better intelligibility completing 12% of the intelligibility items, though such an argument would not be very strong. The measure of the claim made concerning the intelligibility benefits of the 'very familiar' level determined by the Spanish-English analyses are not at all equal. Among the other findings from the t-tests were that five items (20%) were determined to be more intelligible to the raters not 'very familiar', and four items (16%) were equally intelligible between groups. In order to gain a better understanding of the possible intelligibility differences between 'very familiar' raters and all other raters the results of the IPA transcriptions were examined.

Table 5.22: Significant results from independent t-tests measuring the intelligibility differences between the raters 'very familiar' with Arabic-English and all other raters

Item	Spkr. No.	"Very familiar" with Arabic-English			Not "very familiar" with Arabic-English						
		N	*Mean*	*S.D.*	*N*	*Mean*	*S.D.*	*t*	*df*	*d*	
law	3	20	0.80	0.41	167	0.54	0.50	-2.62	26.3	0.569	*
quite	6	20	0.95	0.22	169	0.66	0.47	-4.64	43.2	0.790	**
usual	6	20	1.00	0.00	169	0.82	0.38	-6.02	168.0	0.670	**

*Note: *= p ≤ .05, ** = p ≤ .001*

The results of the IPA transcriptions shown in Table 5.23 also reveal very little differences in how the two transcription-participants processed the four Arabic-English speaker-participants' utterances. The assistant that completed transcription 1 reported being 'very familiar' with Arabic-English, and the second participant had 'some familiarity'. Perhaps the most notable difference between the two transcriptions is with the item, 'quite' in the second sentence from Speaker 6. The item is perceived accurately in the first transcription, /kwaɪt/, whereas the second participant perceived the utterance as ‘quiet’ as reflected in her transcription, /kwɑɪjet/. While this result is reflective of the significant variance determined in the t-test for the item, there are not enough examples of speech perception variation reflected in the IPA transcriptions to make the same claims made concerning Spanish-English.

Table 5.23: The results of the two IPA transcriptions of the Arabic-English speakers’ utterances

Spkr No.	Sentences with intelligibility items underlined	Transcription No. 1	Transcription No. 2
3	The tea cloth is quite wet.	/ðə ti: klɔð ɪz kwaɪt wet/	/ðə ti kləʊð ɪz kwɑɪt wet/
3	The jury learned the law.	/ðə dʒu:rɪ len ðə lɔ:/	/ðə dʒɝri lɪn ðə ləʊ/
6	He paid to change his server.	/hi: bed tʊ ʃeɪndʒ hɪz sɜ:vɜ:/	/hi bɪt tu tʃeɪndʒ hɪz sɝrvɝ/
6	The credit was quite usual.	/ðə kredit wɔ:z kwaɪt u:ʒʊəl/	/ðə kredɪt wəz kwɑɪjet juʒuəl/
10	The blend is rather different.	/ðə blend ɪz ra:ðə dɪfərənt/	/ðə blend iz ræðə dɪfɝent/
10	The children feared the bridge.	/ðə tʃɪldrən fɪə ðə brɪdʒ/	/ðə tʃɪldren fiɾ ðə brɪdʒ/
12	Playing involved her hands.	/pleɪjɪŋ ɪnvɔlved hɜ: hænd/	/laɪiŋ ɪnvɒlvɪd hɚ hænz/
12	The boy wore a patch.	/ðə bɔɪ wɔ: ʌ bʌdʒ/	/ðə bɔɪ wəʊɝ ə bætʃ/

The findings concerning the impact of accent-familiarity level with Arabic-English upon intelligibility are different from those determined for Spanish-English and what will be shown for Dhivehi-English. The results of the ANOVA did determine that significant intelligibility variance did occur between levels with 20% of the items, which does serve to reject the null hypothesis. But the findings examining how the ‘very familiar’ raters’ intelligibility differed from all other raters did not determine explicitly that being ‘very familiar’ with Arabic-English yields raters with a significantly better intelligibility competence processing speech in that accent. The results shown in Table 5.21 do, however, suggest that intelligibility differences occurred between the raters with ‘limited familiarity’ and ‘some familiarity’ demonstrating, though not significantly, that increased familiarity leads to greater intelligibility.

5.5.3 Dhivehi-English familiarity and intelligibility

In this section the research questions concerning how raters' differing levels of familiarity with Dhivehi-English affected the outcomes of the intelligibility gap-fill items will be addressed. The structure of this section is like those of the preceding two sections. It begins by first answering whether or not significant differences in overall intelligibility success occurred between rater-subgroups, followed by closer examinations of how the four groups differed with each Dhivehi-English test item. The section concludes after determining how the 'very familiar' rater-subgroup compared with all other raters' intelligibility success. It will be shown that the effects of raters' accent-familiarity differences with Dhivehi-English on intelligibility were different from those determined in both the results of the Arabic-English and Spanish-English analyses. It will also be shown that raters 'very familiar' with Dhivehi-English did have greater intelligibility success – significantly better with many items – than all other raters.

Figure 5.3 shows the Facets variable map of the Dhivehi-English intelligibility item data analyses. Like with the two other accents' analyses, the same three facets were investigated, and the contents of the columns' described concerning Figure 5.1 are also applicable to Figure 5.3. Looking at the third column, it is clear that the 'very familiar'(n=7) and 'some familiarity'(n=1) raters' intelligibility success were highest among the rater-subgroups. This result was expected, and though significance cannot be determined only from the variable map, it offers the first evidence suggesting the potential rejection of the null hypothesis. The Facets item measurement report presented in Table 5.24 includes the item difficulty details of the contents shown in the fourth column of the variable map. It reveals that similar to both the results from the Spanish-English and Arabic-English item measurement reports that the four Dhivehi-English speakers produced a range of easier and more difficult items from their utterances. What is gained from this information is simply additional evidence to suggest that test candidates with differing levels of pronunciation proficiency can produce utterances with contents that are both more easily intelligible and more difficult to find intelligible. This is important because it suggests that the potential for intelligibility failure, which could result in reduced pronunciation scores, are possible not only for lower level speakers, but higher level speakers as well depending on who is rating their speech. Also determined from the item measurement report is that the overall intelligibility rate for the Dhivehi-English items was 69%. Even though Dhivehi-English was determined to be least familiar overall to the rater-participants, the Dhivehi-English speakers' utterances achieved a greater overall intelligibility success rate than determined for the Spanish-English utterances (61%), which was most familiar to the rater-population.

```
+-------------------------------------------------------------------------------------------------------+
|Measr|+rater     |+Dhivehi English Familiarity Level         |+Item                                     |
|     |           |(most instances of intelligibility-top)   |(easiest - top)                           |
|-----+-----------+-------------------------------------------+------------------------------------------|
|   6 + .         +                                           +                                          |
|     |           |                                           |                                          |
|     |           |                                           |                                          |
|     | .         |                                           |                                          |
|   5 +           +                                           +                                          |
|     | .         |                                           |                                          |
|     | .         |                                           |                                          |
|     |           |                                           |                                          |
|   4 +           +                                           + S4 quietly                               |
|     | **.       |                                           |                                          |
|     | .         |                                           | S13 father                               |
|     | ****.     |                                           | S13 farmer                               |
|   3 + *.        +                                           +                                          |
|     | ********** |                                          | S4 travelled                             |
|     | .         |                                           |                                          |
|     | ******.   |                                           |                                          |
|   2 + ********** +                                          +                                          |
|     | *         |                                           |                                          |
|     | ****.     |                                           | S13 looked     S2 mother                 |
|     | *****.    |                                           | S4 stood       S8 shouting               |
|   1 + ****.     +                                           + S8 huge                                  |
|     | **.       |                                           |                                          |
|     | **.       |                                           |                                          |
|     |           | Some Familiarity     Very Familiar         | S2 laughed     S2 parted                 |
*   0 * *.        *                                           * S13 bulls      S8 hag                    *
|     | *.        | Limited Familiarity                       | S8 challenged                            |
|     | .         | No Familiarity                            |                                          |
|     | *.        |                                           | S13 managed    S4 creature    S8 theatre |
|  -1 + *         +                                           + S8 girl                                  |
|     |           |                                           | S2 stick                                 |
|     | .         |                                           |                                          |
|     |           |                                           | S2 judge                                 |
|  -2 +           +                                           +                                          |
|     |           |                                           | S13 chief                                |
|     |           |                                           |                                          |
|     |           |                                           |                                          |
|  -3 +           +                                           +                                          |
|     |           |                                           |                                          |
|     |           |                                           | S4 passenger                             |
|     |           |                                           | S4 bath                                  |
|  -4 +           +                                           +                                          |
|     |           |                                           | S2 cat                                   |
|     |           |                                           |                                          |
|     |           |                                           |                                          |
|  -5 +           +                                           +                                          |
|-----+-----------+-------------------------------------------+------------------------------------------|
|Measr| * = 3     |+Dhivehi English Familiarity Level         |+Item                                     |
|     |           |(least instances of intelligibility-bottom|(most difficult - bottom)                 |
+-------------------------------------------------------------------------------------------------------+
```

Figure 5.3: The Facets variable map from the analyses of the Dhivehi-English intelligibility gap-fill items

Table 5.24: Facets Dhivehi-English intelligibility item measurement report

Item No.	Spkr No.	Accent	Item	Total Count	Total Score	Obs Avrg	Difficulty Measure in Logits	Model S.E.	Infit MnSq	ZStd
10	2	DE	cat	190	18	0.09	-4.30	0.28	0.96	-0.1
25	4	DE	bath	190	25	0.13	-3.84	0.24	1.14	0.8
23	4	DE	passenger	190	33	0.17	-3.43	0.21	1.10	0.7
76	13	DE	chief	190	70	0.37	-2.15	0.17	1.02	0.3
7	2	DE	judge	190	86	0.45	-1.71	0.16	0.91	-1.3
12	2	DE	stick	190	103	0.54	-1.25	0.17	1.09	1.3
48	8	DE	girl	190	116	0.61	-0.89	0.17	0.76	-3.7
78	13	DE	managed	190	118	0.62	-0.83	0.17	1.01	0.1
20	4	DE	creature	190	120	0.63	-0.77	0.17	1.07	0.9
46	8	DE	theatre	190	121	0.64	-0.74	0.17	1.14	1.8
45	8	DE	challenged	190	136	0.72	-0.27	0.18	0.85	-1.7
44	8	DE	hag	190	144	0.76	0.00	0.19	0.72	-2.9
79	13	DE	bulls	190	144	0.76	0.00	0.19	1.15	1.4
8	2	DE	laughed	190	151	0.79	0.27	0.20	0.90	-0.8
11	2	DE	parted	190	153	0.81	0.36	0.21	1.05	0.4
47	8	DE	huge	190	164	0.86	0.88	0.23	1.03	0.2
49	8	DE	shouting	190	169	0.89	1.18	0.25	0.86	-0.8
24	4	DE	stood	190	171	0.90	1.31	0.26	0.95	-0.1
75	13	DE	looked	190	173	0.91	1.46	0.28	1.25	1.2
9	2	DE	mother	190	174	0.92	1.53	0.28	1.10	0.5
21	4	DE	travelled	190	184	0.97	2.70	0.43	1.01	0.1
77	13	DE	farmer	190	186	0.98	3.15	0.52	0.98	0.1
74	13	DE	father	190	187	0.98	3.45	0.59	1.01	0.2
22	4	DE	quietly	190	188	0.99	3.88	0.72	0.95	0.1
		Mean		190.0	130.6	0.69	0.00	0.27	1.00	0.0
		S.D.		0.0	50.8	0.27	2.14	0.14	0.12	1.3

As with the other two accents, the results of the Facets analyses are first considered in order to determine the answer to the research question of whether or not significant differences in intelligibility occurred between the rater-subgroups. Table 5.25 shows the results from the familiarity level report measures, and provides the details of the results shown in column three of the variable map. The table shows a 21 percent difference in observed averages between the 'very familiar' and 'no familiarity' rater-subgroups; however, the overall differences in intelligibility between the four subgroups were only 'almost significant' according to the Pearson's chi-square included in the analyses ($\chi^2(3)=7.6$, $p=.06$). Like with the other two accents investigated in this study, it is necessary to examine how the four rater-groups' intelligibility success differed with each item before determining if the null hypothesis can be rejected or not.

Table 5.25: Facets Dhivehi-English familiarity level report measures

Familiarity Level	Total Score	Total Count	Obsvd Avrg	Logit Measure	Model S.E.	Infit MnSq	ZStd
No	2670	3960	0.67	-0.42	0.05	0.99	-0.2
Limited	297	408	0.73	-0.18	0.15	0.92	-0.9
Some	20	24	0.83	0.33	0.71	0.71	-0.5
Very	147	168	0.88	0.27	0.32	1.43	2.1
Mean	783.5	1140.0	0.78	0.00	0.31	1.01	0.1
S.D.	1093.6	1633.9	0.08	0.31	0.25	0.26	1.2

In order to better understand what the differences in intelligibility were that occurred between the rater-subgroups beyond overall averages, the results of ANOVA tests conducted for each intelligibility item were examined. Table 5.26 shows the significant results from those tests. Of the 24 total items uttered by the Dhivehi-English speakers, surprisingly only three resulted in significant differences between the four rater-subgroups. It is surprising because it was suspected that because the 'very familiar' and 'some familiarity' groups were so small, and that the differences in observed averages between groups determined by the familiarity level reports that more items would have resulted in significant intelligibility variance between the four groups. This was not the case; however, the overall high measure of intelligibility success rates among all raters transcribing the Dhivehi-English speaker-participants may have contributed to the low number of significant items. Based on these results alone, it is difficult to reliably claim that the answer to the research question concerning whether or not raters' accent-familiarity differences resulted in significant variance in intelligibility success, is 'yes'. It will be shown, however, that accent-familiarity level with Dhivehi-English did significantly affect intelligibility when examining how the ‘very familiar’ and all other raters’ intelligibility success compared with each item.

Table 5.26: Significant results from one-way analyses of variance of the Dhivehi-English intelligibility items by familiarity level with Dhivehi-English

Item	Spkr No.		Sum of Squares	*df*	Mean Square	*F*	*p*	Eta Squared
cat	2	Between Groups	4.28	3	1.43	22.08	0.000	0.263
		Within Groups	12.02	186	0.07			
		Total	16.30	189				
passenger	4	Between Groups	3.43	3	1.14	8.93	0.000	0.126
		Within Groups	23.84	186	0.13			
		Total	27.27	189				
chief	13	Between Groups	2.39	3	0.80	3.55	0.016	0.054
		Within Groups	41.82	186	0.23			
		Total	44.21	189				

Table 5.27 shows the results of the t-tests conducted to determine if the ‘very familiar’ measure of accent-familiarity with Dhivehi-English was significantly more successful completing the Dhivehi-English gap-fill items. The results show just how impactful the ‘very familiar’ accent-familiarity level was to intelligibility concerning Dhivehi-English. Five of the nine items were answered correctly by all ‘very familiar’ raters, and the other four were missed by only one rater each. With Dhivehi-English the greatest differences of mean scores determined by the t-tests were observed. The item, ‘cat’- Speaker 2, was answered incorrectly by only one of the seven ‘very familiar’ raters, but only twelve of the 183 other raters answered it correctly – an 80% difference in intelligibility. In total, three of the items (passenger – Speaker 4; ‘chief’ – Speaker 13; and ‘cat’) resulted in mean score differences greater than 50%. Neither of the two other accents included in this study produced such drastic differences from any items.

Table 5.27: Significant results from independent t-tests measuring the intelligibility differences between the raters ‘very familiar’ with Dhivehi-English and all other raters

Item	Spkr. No.	"Very familiar" with Dhivehi-English			Not "very familiar" with Dhivehi-English						
		N	*Mean*	*S.D.*	*N*	*Mean*	*S.D.*	*t*	*df*	*d*	
judge	2	7	0.86	0.38	183	0.44	0.50	-2.85	6.8	0.95	*
cat	2	7	0.86	0.38	183	0.07	0.25	-8.11	188.0	2.46	**
parted	2	7	1.00	0.00	183	0.80	0.40	-6.79	182.0	0.71	**
passenger	4	7	0.86	0.38	183	0.15	0.36	-5.17	188.0	1.92	**
stood	4	7	1.00	0.00	183	0.90	0.31	-4.59	182.0	0.46	**
huge	8	7	1.00	0.00	183	0.86	0.35	-5.49	182.0	0.57	**
shouting	8	7	1.00	0.00	183	0.89	0.32	-4.86	182.0	0.49	**
chief	13	7	0.86	0.38	183	0.35	0.48	-3.45	6.8	1.18	*
bulls	13	7	1.00	0.00	183	0.75	0.43	-7.82	182.0	0.82	**

*Note: *= p ≤ .05, ** = p ≤ .001*

The results of the two IPA transcriptions shown in Table 5.28 also show some the same differences in intelligibility determined by the t-tests. The first transcription was completed by a native Maldivian and Dhivehi-English L2 speaker. The assistant that completed the second transcription reported having ‘limited familiarity’ with the accent. In all sentences, transcription 1 demonstrates complete intelligibility occurred; however, transcription 2 shows several instances where speech perception differed, and intelligibility was not successful. Examples illustrating how intelligibility was affected by level of familiarity in the transcriptions include the second sentence from Speaker 2, which includes two instances concerning how ‘cat’ and ‘stick’ were interpreted. The second sentence from Speaker 4 also included two troublesome items, ‘passenger’ and ‘bath’; and two items from the second sentence from Speaker 13, ‘managed’ and ‘bulls’ showed intelligibility did not occur in the second transcription. These examples from the IPA transcriptions provide evidence suggesting that the assistant ‘very familiar’ with Dhivehi-English was more capable successfully transcribing the speech in that accent than the assistant with limited familiarity.

Table 5.28: The results of the two IPA transcriptions of the Dhivehi-English speakers' utterances

Spkr No.	Sentences with intelligibility items underlined	Transcription No. 1	Transcription No. 2
2	The judge laughed at his mother.	/ðə dʒʊdʒ la:fd æt hɪz mʌd̠ɜ:/	/ðə dʒʌdʒ lɑft aʔ hɪz mɑdə/
2	The cat parted with his stick.	/ðə ket pa:təd wit̠ hɪz stɪk/	/ðə keɪk pɑtɪd wɪt hɪs tɪk/
4	The creature travelled quietly.	/ðə krɪeɪtʃɜ: trævəld kwaɪᵊtli:/	/ðə krieɪtʃət trævəld kwaɪɪtli/
4	The passenger stood in a bath.	/ðə pæsendʒɜ: stu:d ɪn ə ba:t̠/	/ðə prisentə stud ɪn ə bʌt/
8	The old hag challenged the theater.	/ðə ɔ:ld hæg tʃæləndʒ ðə t̠ɪeɪtɜ:/	/ðə əʊl hæg tʃælendz ðə tiɝtə/
8	The huge girl is shouting.	/ðə hɪu:dʒ gɜ:l ɪz ʃaʊtɪŋ/	/ðə hudʒ gɜl iz ʃəʊtiŋ/
13	Father looked at the chief.	/fa:d̠ɜ: lʊk æt ðə tʃi:fs/	/fɑdə lʌkt æt ðə tʃifs/
13	The farmer managed the bulls.	/ðə fa:mɜ: mæneɪdʒ ðə bʊls/	/ðə fɑmɚ mænɪ ðə buts/

Based on the evidence from the ANOVA, t-tests and IPA transcriptions it is possible to reject the null hypothesis, and suggest that accent-familiarity level did affect intelligibility. The findings are not the same as those determined for Spanish-English nor those determined for Arabic-English, but the null hypothesis can confidently be rejected. The answer to the research question concerning if the raters 'very familiar' with Dhivehi-English demonstrated greater intelligibility benefits is also 'yes'. Among the three accents, Dhivehi-English showed the greatest differences in intelligibility success from the 'very familiar' rater-subgroup on individual items. True, the overall intelligibility differences determined by the Chi-square between the four subgroups were not significant (p=.06), but when examining the results from the t-tests, the fact that raters 'very familiar' did have significant advantages is clear. Ten of the 24 total items from the Dhivehi-English speakers were answered correctly by all 'very familiar' raters; all of the raters 'very familiar' with Spanish-English only answered two of the 25 items, and four items were answered correctly by all raters 'very familiar' with Arabic-English. The theory that raters 'very familiar' with Dhivehi-English would have the greatest measures of benefits to intelligibility is more than plausible.

The findings from all three accents concerning how raters' differing familiarity levels affected intelligibility were not equal, but they did all confirm that, at the least, accent-familiarity level differences can result in significant differences in intelligibility. Furthermore, the findings from the t-tests confirmed that raters 'very familiar' with any of the three accents were more successful, and often significantly more successful, completing the majority gap-fill items than raters with less familiarity. In the next section, the differences in pronunciation scores and intelligibility between familiarity levels with each accent will be examined further, and attempt to answer the research question concerning whether or not correlations exist between both raters' accent-familiarity levels and pronunciation scores, and raters' accent-familiarity levels and intelligibility.

5.6 Accent-familiarity's correlations with pronunciation scoring and intelligibility

The previous two sections of this chapter determined that raters' differing levels of familiarity affect both pronunciation scores and intelligibility. However, the findings did not reveal whether or not significant positive correlations exist. In this section the two research questions concerning whether or not there are correlations between raters' levels of accent familiarity and pronunciation scores, and with intelligibility are considered. Evidence will be shown suggesting there are significant positive correlations linking increased accent-familiarity levels with higher pronunciation scores among all three accents, and significant correlations between familiarity level and increased intelligibility for Spanish- and Dhivehi-English.

To determine the answers to these questions both the data from the pronunciation scores and intelligibility items were analysed using SPSS. The pronunciation score data was divided according to speaker-accent groups, and mean scores from each rater for each accent group were calculated. Scatter plots were graphed placing the raters' mean pronunciation scores on the y-axis with their accent-familiarity level on the x-axis. The same procedure was conducted for the intelligibility data where the data was divided by speaker-accent groups (x-axis), and each rater's calculated mean intelligibility success rate (y-axis). Regression lines were also calculated to illustrate the correlations. To compliment the scatterplots, Pearson's correlations were conducted to determine if the correlations were significant, and the effect sizes of the correlations were determined by calculating the R Squared for each correlation.

The results for both pronunciation scores and intelligibility concerning Spanish-English revealed significant correlations. Figure 5.4 shows the scatterplots, and Tables 5.29 and 5.30 provide the results of the Pearson's correlations and effect sizes. Of course, significant correlations do not imply causation, and the effect sizes were not large. It is true that the effect sizes can be used to determine that 76% of the variance for pronunciation and 73% of the variance for intelligibility success was shared between the four groups of raters, but the similarities that will be shown across all the accents to these results suggest raters' accent familiarity could be impactful to scores and intelligibility.

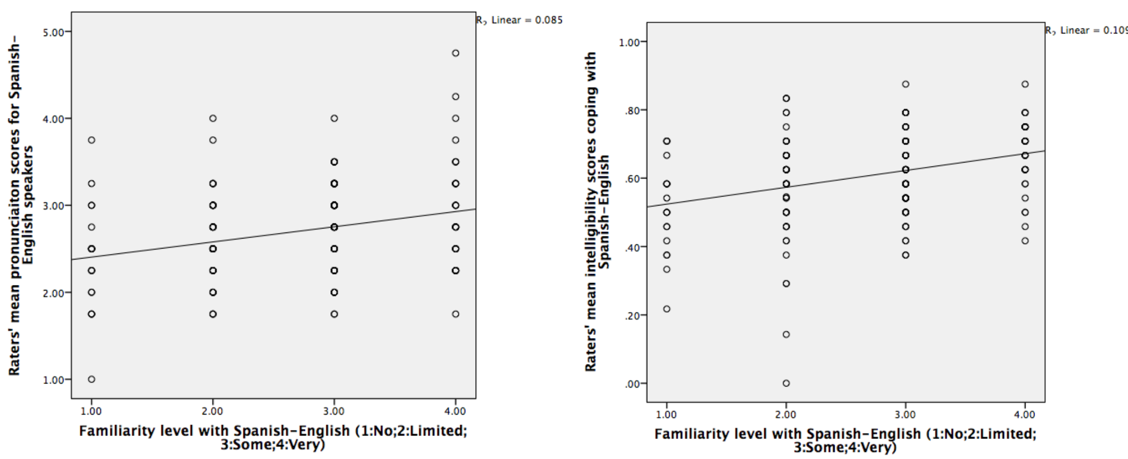


Figure 5.4: Scatterplots with regression lines of the correlations between raters' familiarity levels with Spanish-English and raters' mean pronunciation scores given to the Spanish-English speaker-participants (left) and with the raters' mean intelligibility success rates transcribing the Spanish-English intelligibility items

Table 5.29: Pearson's correlation results measuring familiarity level with Spanish-English and pronunciation scores

		Familiarity level with Spanish-English (1:No;2:Limited; 3:Some;4:Very)	Raters' mean pronunciaiton scores for Spanish-English speakers
Familiarity level with Spanish-English (1:No;2:Limited;3:Some;4:Very)	Pearson Correlation	1	.292**
	Sig. (2-tailed)		0.000
	N	187	187
	R squared		0.850
** Correlation is significant at the 0.01 level (2-tailed).			

Table 5.30: Pearson's correlation results measuring familiarity level with Spanish-English and intelligibility success

		Familiarity level with Spanish-English (1:No;2:Limited; 3:Some;4:Very)	Raters' mean intelligibility scores coping with Spanish-English
Familiarity level with Spanish-English (1:No;2:Limited;3:Some;4:Very)	Pearson Correlation	1	.330**
	Sig. (2-tailed)		0.000
	N	189	189
	R squared		0.109
** Correlation is significant at the 0.01 level (2-tailed).			

As stated previously, the findings concerning Arabic-English determined significant correlations for pronunciation scores but not intelligibility. The scatterplots for both pronunciation scores and intelligibility are shown in Figure 5.5, and Tables 5.31 and 5.32 provide the Pearson's correlation and effect sizes. Again, these correlations and effect sizes may not seem strong, but the findings do indicate that if 'very familiar' raters had more intelligibility success than raters with 'no familiarity'.

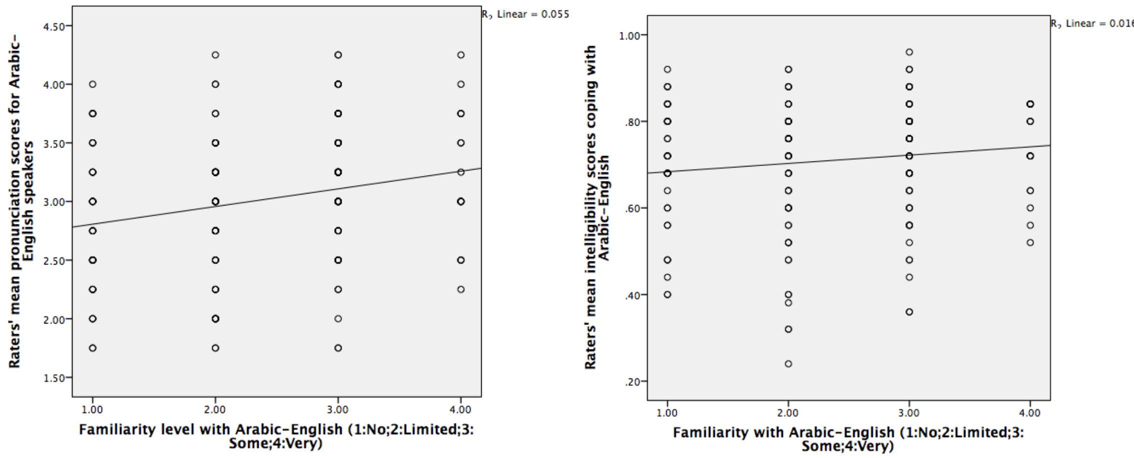


Figure 5.5: Scatterplots with regression lines of the correlations between raters' familiarity levels with Arabic-English and raters' mean pronunciation scores given to the Arabic-English speaker-participants (left) and with the raters' mean intelligibility success rates transcribing the Arabic-English intelligibility items

Table 5.31: Pearson's correlation results measuring familiarity level with Arabic-English and pronunciation scores

		Familiarity level with Arabic-English (1:No;2:Limited; 3:Some;4:Very)	Raters' mean pronunciation scores for Arabic-English speakers
Familiarity level with Arabic-English (1:No;2:Limited;3:Some;4:Very)	Pearson Correlation	1	0.234**
	Sig. (2-tailed)		0.001
	N	187	187
	R squared		0.055
** Correlation is significant at the 0.01 level (1-tailed).			

Table 5.32: Pearson's correlation results measuring familiarity level with Arabic-English and intelligibility success

		Familiarity with Arabic-English (1:No;2:Limited; 3:Some;4:Very)	Raters' mean intelligibility scores coping with Arabic-English
Familiarity with Arabic-English (1:No;2:Limited;3:Some;4:Very)	Pearson Correlation	1	0.127
	Sig. (2-tailed)		0.083
	N	188	188
	R squared		0.016

Significant positive correlations were determined between raters' accent-familiarity levels and both pronunciation scores and intelligibility. The scatterplots are shown in Figure 5.6, and the results of the Pearson's correlations and effect sizes are shown in Tables 5.33 and 5.34. The findings are very similar to those determined for Spanish-English.

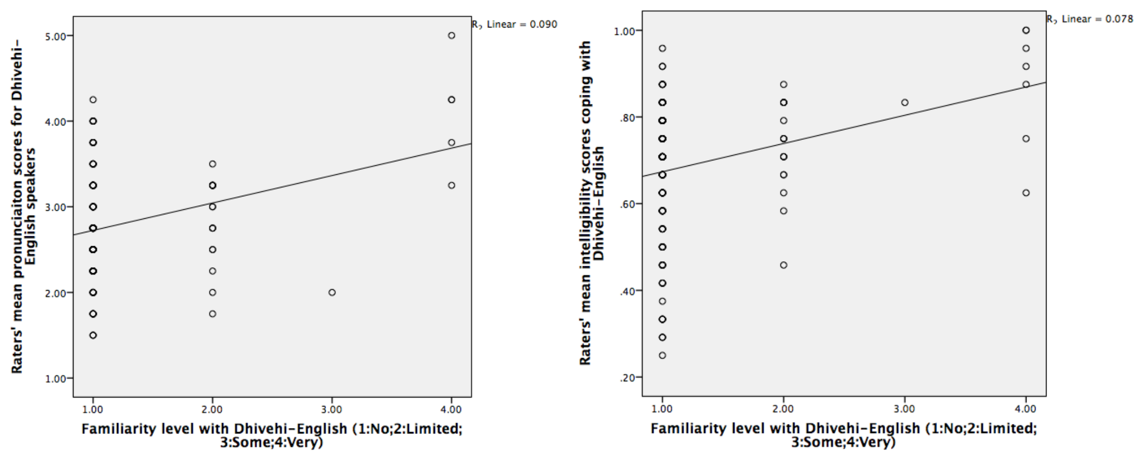


Figure 5.6: Scatterplots with regression lines of the correlations between raters' familiarity levels with Dhivehi-English and raters' mean pronunciation scores given to the Dhivehi-English speaker-participants (left) and with the raters' mean intelligibility success rates transcribing the Dhivehi-English intelligibility items

Table 5.33: Pearson's correlation results measuring familiarity level with Dhivehi-English and pronunciation scores

		Familiarity level with Dhivehi-English (1:No; 2:Limited; 3:Some;4:Very)	Raters' mean pronunciaiton scores for Dhivehi-English speakers
Familiarity level with Dhivehi-English (1:No;2:Limited;3:Some;4:Very)	Pearson Correlation	1	.301**
	Sig. (2-tailed)		0.000
	N	188	188
	R squared		0.090

** Correlation is significant at the 0.01 level (2-tailed).

Table 5.34: Pearson's correlation results measuring familiarity level with Dhivehi-English and intelligibility success

		Familiarity level with Dhivehi-English (1:No;2:Limited; 3:Some;4:Very)	Raters' mean intelligibility scores coping with Dhivehi-English
Familiarity level with Dhivehi-English (1:No;2:Limited;3:Some;4:Very)	Pearson Correlation	1	.279**
	Sig. (2-tailed)		0.000
	N	190	190
	R squared		0.078

** Correlation is significant at the 0.01 level (2-tailed).

The findings from all three accents indicate that significant positive correlations exist between raters' accent-familiarity levels and pronunciation scores, as well as between raters' accent-familiarity levels and intelligibility success rates for two of the three accents. While the effect sizes were not large, and these findings do not guarantee accent-familiarity level is the cause of pronunciation score variance or intelligibility variance between rater-groups, they do determine that these relationships exist. In the context of a high-stakes speaking test such chances for score variance due to individual raters' accent-familiarity levels should be considered a real threat to the reliability and validity of scores on such tests.

Notes concerning the two removed outlier rater-participants

The two rater-participants that were removed from the pronunciation score data set because of their fit statistics deserve a brief discussion here. Both raters demonstrated how strongly raters' accent-familiarity levels are related to both pronunciation scores and intelligibility. The reasoning their pronunciation score fit statistics did not fit the Rasch model is that they both scored all four Dhivehi-English speakers a score of '5 – Speech is generally clear and requires little or no listener effort'. What was problematic about their scores was that they scored all but one of the remaining eight nonnative speaker-participants' pronunciation as either '1- The speaker's pronunciation is unintelligible', or '2 - Consistent pronunciation difficulties cause considerable listener effort throughout the sample. It was necessary to listen several times before attempting to complete the gap fill'. If this study had only examined pronunciation scores and not also included intelligibility tasks, the only conclusions that could have been made would be that the two rater-participants were too biased toward Dhivehi-English test-candidates, and should be removed. But by also investigating intelligibility, the pronunciation scores the two outliers delivered are understandable. When completing the transcriptions of the Dhivehi-English speakers both raters answered 23 of 24 items correctly; they did not miss the same item. To these two raters, it appears, the Dhivehi-English speakers' pronunciation was completely intelligible and easy for them to process, so the scores they determined for them fit the description of their rating scale choices. The intelligibility results for the raters coping with the seven other speaker-participants' they scored so poorly for pronunciation likewise reflected the pronunciation score descriptors; their speech was largely unintelligible to them. Clearly these two individuals are not qualified to act as actual raters of high-stakes tests of spoken English, but they do demonstrate how high-stakes tests' reliance on raters' listener effort to determine pronunciation proficiency can be affected by raters' accent-familiarity levels.

5.7 The unequal effect of accent familiarity

This section addresses the research question concerning whether or not the *rater accent-familiarity effect* on pronunciation scores and intelligibility is more prevalent with some accents than others. This question is included and considered important because if raters' differing accent-familiarity levels affect pronunciation scores on high-stakes tests, it would be considered ideal if the effect were equal among all nonnative accents. If the effect were determined to be equal among accents, test developers could more easily incorporate the raters' accent-familiarity levels into consideration for possible score adjustments. Likewise, if it is determined that accent-familiarity also affects intelligibility, and that effect contributes to pronunciation score differences, it would be beneficial to understand if the effect on intelligibility is equal among all accents. However, as the title of this section implies, the results from this study suggest that the effects of raters' differing accent familiarity levels were found to be different for each language concerning pronunciation scores and intelligibility.

The findings this section presents were determined by examining the results from all three accents' pronunciation and intelligibility data analyses. First discussed will be how the results from the analyses of the pronunciation scores differed, followed by discussions of variance observed in the results of the intelligibility analyses'. It will be shown that though all three accents showed correlations between both increased intelligibility and higher pronunciation scores with accent-familiarity, there were differences in the sizes of the effect observed among the three accents.

5.7.1 Differences in the rater accent-familiarity effect on pronunciation scores

The effects of raters differing levels of accent-familiarity on pronunciation scores were not equal among the three accents investigated. The results of the ANOVA, t-tests and bias interactions revealed for each of the three accents were compared, and it was determined that familiarity with Dhivehi-English showed the largest effect size to pronunciation scores. As Tables 5.11-5.13 show, three of four Dhivehi-English speakers' pronunciation scores varied significantly between the four familiarity related rater-subgroups. All four Dhivehi-English speakers' pronunciation scores showed significant variance between the raters 'very familiar' and all other raters in the t-tests where only two of four Spanish-English speakers' scores differed significantly from the 'very familiar' raters (see Table 5.5), and one Arabic-English speaker's scores were significantly different between 'very familiar' and all other raters (see table 5.8). The bias size (1.85) for the 'very familiar' group of raters with Dhivehi-English was six times larger than that determined for Spanish-English 'very familiar' raters (.29), and more than 30 times larger than the Arabic-English 'very familiar' raters' bias size (.06). Raters' accent-familiarity with Spanish-English showed less effect than Dhivehi-English, but more than Arabic-English. While it is true that the raters 'very familiar' with Dhivehi-English were comprised of only five of the 188 raters, the Spanish-English and Arabic-English rater population sizes did not have the same limitations. Raters 'very familiar' with Spanish-English scored the Spanish-

English speakers on average fifteen percent higher than what the Rasch model predicted. Likewise, raters 'very familiar' with Arabic-English showed a thirteen percent higher than expected average scores scoring the Arabic-English speakers' pronunciation. These findings demonstrate that the differences of the accent-familiarity level effect to pronunciation scores are, in fact, not equal between the three accents included in this study. Therefore, it is difficult to attempt to make a single generalizable explanation of the effect of raters' accent-familiarity level differences on pronunciation scores concerning any or all accents. The effects, it seems, are accent-specific.

5.7.2 Differences in the rater accent-familiarity effect on intelligibility

Determining if the effects of raters' accent-familiarity differences on intelligibility were equal with all three accents was more complicated than what was necessary to determine the differences of pronunciation scores. The difficulty of the items from each accent group had to be considered, as well as the differences in success the rater-subgroups showed coping with each accent group of speaker-participants. It will be shown that differences occurred between the three accents concerning the measures of the impact accent-familiarity level had on intelligibility.

Overall, the greatest differences in intelligibility were determined between the four accent-familiarity rater-subgroups for Spanish-English. The results of the Facets Spanish-English intelligibility item analyses were the only to include a significant Chi-square result between the four rater-subgroups ($\chi^2(3)=17.3$, p=.00). It was surprising to learn that not only were the Spanish-English items determined to be on average more difficult than the items from the other two accents, as determined by the mean logit measure of the items (-.33), but that the overall average success rate all rater-participants showed coping with the Spanish-English was lowest of the three accents (61%) (see again table 5.14). The test was challenging for this particular group of rater participants, but the exact agreement statistic (71.8%) and expected agreement statistic (71.3%) suggest the difficulty level of the test matched the ability level of that particular group of rater-participants. The ANOVA results that examined how the four rater-subgroups coped with each Spanish-English item were also the only ANOVA results that showed significant differences between groups for items from all four speakers. Overall, Spanish-English pronunciation showed the greatest variance in intelligibility when examining accent-familiarity divided into four levels.

Dhivehi-English showed less overall variance of intelligibility success than Spanish-English did, but more than was determined from the Arabic-English items. It was surprising that the Chi-square results included in the Facets analyses did not reveal significant differences between rater-subgroups (p=.06) with intelligibility, and likewise that the rater population demonstrated an overall greater average success rate transcribing the Dhivehi-English speakers' speech (69%) than they did with Spanish-English (61%). It was surprising because of the three accents, Dhivehi-English had the lowest number of raters with any familiarity at all with the accent (25/190). Overall, the results demonstrate that though Dhivehi-English is less-known than Spanish-English, it was not less intelligible. The Facets analyses calculated the mean logit measure of the Dhivehi-English items to be .00 with a standard deviation of 2.14, which was the smallest standard deviation of item difficulty of the three accents. These findings also suggest not only that the rater-population, overall, coped well with Dhivehi-English, but that the Dhivehi-English items – as a group of items, or test – matched the ability level of this particular group of rater-participants.

Though the raters 'very familiar' with Dhivehi-English failed to show significant overall success completing the items, the greatest differences in individual item response success was observed in the t-tests that examined the 'very familiar' level of familiarity with all other raters' performance. Nine items from all four speakers resulted in significant differences in the t-tests – the only accent-group to include significant differences from all four speakers in the t-test results. Raters accent-familiarity differences with Dhivehi-English showed more differences in its effect on intelligibility than the other two accents investigated.

The effects on intelligibility by raters' differing levels of accent-familiarity with Arabic-English were determined to be lowest among the three accents investigated. Similar to the measures of pronunciation score differences due to accent-familiarity with Arabic-English, the results suggest little impact occurred with intelligibility. As mentioned earlier, no significance was determined in the Facets Chi-square results between the four rater-subgroups (p=.21). Overall, the rater-population was most successful transcribing the Arabic-English speakers' speech (71%), and the mean logit measure determined that the Arabic-English items were easiest of the three accents' items (.23) (see Table 5.20 again). The exact agreement results were 72.3%, and the expected were 71.8%, which suggest the Arabic-English intelligibility items were comparatively easier than the other two accents according to the Facets analyses. Perhaps for this reason, less variance in intelligibility success was observed than with the other two accents.

The Arabic-English intelligibility items showed less variance than the other two accents in both the ANOVA and t-test results as well. Only five from the 25 total items resulted in significant intelligibility variance between the four rater-subgroups in the ANOVA, and only three items were found to cause significant intelligibility variance between the 'very familiar' and 'all other raters' from the t-tests. It is possible that this particular group of speaker-participants' pronunciation was largely intelligible to this particular group of rater-participants. It is also possible that the effects of accent-familiarity level with Arabic-English are less than the effects with Spanish- or Dhivehi-English, and it is possible that these results occurred by chance. What does seem plausible is there is reason to suspect that the effects of accent-familiarity levels on intelligibility are not equal among all accents. The effects, very likely, are accent-specific. In the next section, population sizes are considered as a potential means for testing agencies to attempt to predict what levels of accent-familiarity a pool of raters may have.

5.8 The rater accent-familiarity effect and test-takers' L1 population size considerations

In this section the research question concerned with whether or not the population sizes of test candidates' L1s affect the likelihood of raters having familiarity with their accents is considered. As the other findings already presented in this chapter have determined, raters' accent familiarity differences can cause significant differences in both pronunciation scores and intelligibility. Though the effects were not always significant, the correlation analyses suggest that the higher the level of familiarity a rater is with a test-taker's accent, the more likely it is that rater will find that test-taker's speech intelligible, and score their pronunciation more leniently. This research question was asked in order to determine if testing agencies might be able to predict which accents a group of raters will most likely be familiar with and level of familiarity based on the global population sizes of each accent's respective L1.

In order to answer the question, three types of data were analysed. First, the data resulting from the rater-participants' reported levels of accent-familiarity with nine different World English accents were examined for frequencies. Also included in the analyses were data collected from web-based resources providing the global population sizes and percentages for each respective L1. These results are shown in Table 5.29. The accents most raters had 'no familiarity' with appear at the top of the list (Dhivehi-English), and the accent the least number of raters had 'no familiarity' with appearing at the bottom (Spanish-English). They were listed this way because though the 'very familiar' level has been determined to have the greatest amount of benefit to pronunciation scores and intelligibility, it seems more important to consider how many rater-participants had no familiarity at all with different accents. The L1 population sizes and world population percentages provide additional scale and context to the table, as it was theorized that L1 population size might be a useful factor for test administrators to predict accent-familiarity levels from rater populations.

Table 5.35: Rater-participants' reported levels of accent-familiarity with nine World English accents and estimated L1 speaker population sizes

First Language	No familiarity		Limited fam.		Some fam.		Very fam.		Native L1 speakers in millions (2010)	Fraction of world population
	Total	Percent	Total	Percent	Total	Percent	Total	Percent		
Dhivehi	165	86.8	17	8.9	1	0.5	7	3.7	0.4	0.01%
Indonesian	90	47.4	73	38.4	25	13.2	1	0.5	77	1.16%
Portuguese/Brazilian	69	36.3	65	34.2	46	24.2	9	4.7	215	3.08%
Urdu	62	32.6	78	41.1	40	21.1	9	4.7	66	0.99%
Korean	54	28.4	63	33.2	60	31.6	11	5.8	76	1.14%
Arabic	38	20	62	32.6	68	35.8	20	10.5	295	4.23%
German	29	15.3	63	33.2	68	35.8	29	15.3	89	1.39%
Japanese	25	13.2	73	38.4	62	32.6	29	15.3	125	1.92%
Spanish	19	10	57	30	78	41.1	35	18.4	405	5.85%

Note: Native speaker population data extracted from Wikipedia's *List of languages by number of native speakers* (March4, 2016)

The findings determined that it is not practical to presume that L1 population sizes can be utilized to predict raters' accent-familiarity levels with different World English accents. While it may seem understandable that 86.8% of rater-participants had no familiarity with Dhivehi-English because of its population size and remoteness, the percentages of raters with no familiarity with Arabic-(20%) and Portuguese/Brazilian-English (36.3%) were much larger compared to accents with smaller L1 populations like German- (15.3%)- or Japanese-English (13.2%). This study recruited 190 rater-participants born or raised in 35 different countries, so it is likely that this particular group of rater participants might have a more varied exposure to different World Englishes than a group of rater-participants recruited from only one country or consisting of only native English speakers. Based on these results, the answer to the research question concerning whether or not the population sizes of test candidates' L1s affect the likelihood of raters having familiarity with their accents is, 'not necessarily'.

Summary

Overall, the findings of this study suggest that accent-familiarity, as a rater trait or characteristic, affects not only pronunciation scores, but also contributes to differences in intelligibility success. The inclusion of intelligibility success data allowed for the first time more than speculative reasoning why score differences had been observed in a study investigating whether or not test scores are affected by raters' differing accent-familiarities. The findings include compelling evidence to suggest that correlations do exist between both pronunciation scores and intelligibility success with raters' accent-familiarity levels. It was also determined that the effects of raters' differing accent-familiarity levels are not equal between all accents, and that it is not possible to predict which accents more raters will have more familiarity with based on the population sizes of the respective L1s associated with each nonnative English accent. In the next chapter, conclusions and implications of this research will be detailed.

Chapter 6 : Conclusion and Implications

This chapter presents summaries of the main findings as well as some suggested implications of this study. Also included will be broader more generalized conclusions drawn from the results of both the pilot study and main study that serve to both answer the research questions and address the veracity of the hypotheses introduced in the first chapter. Recommendations for future research are also included, so that through continued research a greater, more complete understanding of this threat to valid, reliable high-stakes tests of spoken English can be attained.

Divided into six main sections, the chapter begins with a summary of the research approach and methods included in the study. Section 6.2 focuses on the findings related to pronunciation scoring, and 6.3 addresses the findings concerned with how accent-familiarity levels affected intelligibility. Next, an argument is made in section 6.4 suggesting that 'biasing for the best' by focusing on raters 'very familiar' with any particular accent is presented. Limitations of this study are presented along with recommendations for future research in section 6.5 followed by concluding remarks in section 6.6.

6.1 Review of the research approach and methods

The general aim of this research was to determine what effect raters' differing levels of accent-familiarity with four different nonnative English accents have on pronunciation scores on high-stakes, semi-direct English tests. The problem this research addressed was that raters' differing accent-familiarities with test candidates' speech cause a rater-dependent, construct-irrelevant threat to the reliability and validity of pronunciation scores. Based on a strong theoretical stance outlined in Chapter 1 and detailed in the review of the literature in Chapter 2, research questions and hypotheses emerged linking raters' levels of accent-familiarity with a particular group of L1 accented English speakers to variation in both pronunciation score decisions and intelligibility success with those same speakers. Among these hypotheses included the first attempt to offer potential reasoning for pronunciation score variance occurring between raters with differing levels of accent-familiarity. The hypothesis suggests score variance occurs between raters due to speech-intelligibility benefits raters gain through exposure to different accents; as accent-familiarity increases the listener's ability to find the pronunciation of speakers of that accent more intelligible also increases. Since pronunciation rating scales on high-stakes tests rely on raters to determine their individual 'listener effort' needed to cope with test-takers' pronunciation in order to determine scores, the amount of accent-familiarity a rater has with an accent directly affects the amount of 'listener effort' raters require. It was further hypothesized that raters' accent-familiarity levels have correlations with both pronunciation scores and intelligibility success. To empirically test these hypotheses two tests were designed.

The first test, a pilot study described in Chapter 3, sought to determine how raters' differing accent-familiarity levels with Japanese-English affected the pronunciation scores and intelligibility of five Japanese-English speakers. Eighty-seven rater-participants completed the three-part test. The findings from the MFRM analyses of the pronunciation scores and intelligibility tasks determined that raters' accent-familiarity levels had significant impact on both pronunciation scores and intelligibility success. Positive, significant correlations were also determined between accent familiarity levels and both higher pronunciation scores and increased intelligibility. Experiences gained conducting the pilot study influenced the design decisions of the larger main study. Key changes made to the methodologies included changing how the sentences for the speaker-participants were designed to better feature aspects of phonological difficulty with each accent and to also investigate bias.

The main study described in Chapters 4 and 5 examined the effects raters' differing accent-familiarity levels had on pronunciation scores and intelligibility of speakers of Spanish-English (n=4), Arabic-English (n=4) and Dhivehi-English (n=4). Similar to the first study, the findings revealed significant differences in pronunciation scores and intelligibility success among the four investigated accents between raters with differing accent-familiarity levels, though the measures of variance differed between accents. The key findings from both studies are presented in the next three sections.

6.2 Accent-familiarity levels and their effects on pronunciation scores

<u>Hypothesis 1</u>: Raters with higher levels of accent-familiarity with the L2 accented speech of speakers from a specific L1 background will tend to score those speakers higher than raters less familiar or unfamiliar with the accented speech from the same speakers.

Finding (1): Accent-familiarity caused significant differences of pronunciation scores. Significant score variance occurred between rater groups divided by accent-familiarity level for three of four speaker-participants of Spanish-English and Dhivehi-English, for three of five Japanese-English speaker-participants, and one of four Arabic-English speaker-participants.

Finding (2): The effects on pronunciation scores potentially attributed to raters' accent-familiarity level differences were not equal among the four accents investigated. The largest effect was determined toward Japanese-English, less effect toward Dhivehi-English, and the least effect on Arabic- and Spanish-English pronunciation scores.

<u>Hypothesis 2</u>: Raters' accent-familiarity levels with the L2 accented speech of speakers from a specific L1 and the pronunciation scores raters deliver are correlated.

Finding (3): Significant positive correlations between raters' accent familiarity level and pronunciation scores were determined. As accent-familiarity level increased, so too did pronunciation scores.

The findings that pronunciation scores were affected by raters' differing accent-familiarity levels supports similar findings determined by Carey et al. (2011) and Winke et al. (2011) discussed in Chapter 2. Raters' differing accent-familiarity levels do pose a threat to the reliability and validity of scores on high-stakes tests of spoken English. This was the main research question of this study, though no single answer could be ascertained concerning the degree to which the potential impact to scores may be that are applicable to all accents. Different amounts of score variance also occurred between the accents both Carey et al. (2011) and Winke et al. (2011) investigated, though this study was the first to include effect sizes that also varied between accents.

The significant positive correlations between familiarity levels and pronunciation scores were hypothesized, though the effect sizes were small. The small effect sizes could have been due to the structure of the familiarity level scale (no familiarity; limited familiarity; some familiarity; very familiar). If the accent-familiarity scale had allowed raters to self-score between levels (e.g. between limited familiarity and some familiarity with Arabic-English), it may have strengthened the effect sizes. Whatever the reason(s) for the small effect sizes, the amount of impact on individual scores due to rater's accent-familiarity levels should not be considered as either an *acceptable* source of score variance or as *insignificant* variance.

The main implication of these findings is that the threat to the validity and reliability of scores on high-stakes tests of spoken English can no longer be ignored or considered as a '*possible threat*'. It has been empirically researched now in multiple studies, and should henceforth be considered a '*valid threat*'. Raters' accent-familiarity levels do affect pronunciation scores, and that effect is often significant.

The finding that carries the most potentially troublesome implications is that the effects raters' accent-familiarity levels have on pronunciation scores are not equal among all accents. The results determined the effects on pronunciation scores were weakest with Arabic-English, and strongest toward Japanese-English. There was no evidence to indicate the impact between the four accents were equal or arguably similar. These findings suggest that pronunciation scores will be more or less affected by raters' differing familiarity levels with some test-takers' accents than with others. Winke et al. (2011; also Kim, 2009) had similar findings to those in this study for Arabic-English with no significant score variance occurring for Korean L1 test-takers from raters with Korean as their L2. While Winke (et al.) suspected the lack of significant variance may have been due to a small sample size (n=11) of Korean L2 speaking raters, the present study did not have sample size problems concerning the rater-groups for Arabic-English. The only explanation I can make is that the effects of accent-familiarity on pronunciation scores (and intelligibility) appear to be weaker with some accents than with others. It appears that the effects of raters' accent-familiarity differences are 'accent-specific', and therefore, are not easily generalizable. As a result, another implication is that the threat cannot be managed or adjusted by a single solution. It is not likely that one algorithm could be devised to adjust scores for use with all accents in all tests, nor is it likely test administrators could reliably predict how much scores will be affected by raters' accent-familiarity differences with any particular accent without prior research.

Possible solutions could include testing agencies maintaining accent-familiarity level data from all raters, and tracking rater-to-accent scoring data as an additional means of measuring raters' exposure and experience with different accents. Raters could also be tested periodically to determine any changes in intra-rater reliability with different accents in order to better understand how increased familiarity impacts rater judgments. It may also be necessary for testing agencies to reconsider the practice of randomly assigning test samples to raters for scoring, and instead include raters' accent-familiarity data as a facet of the scoring process.

Potential implications of this study concerning pronunciation scores include addressing the current hybridized holistic scales currently in use in the TOEFL iBT. This study focused on how pronunciation scores are affected by raters' differing accent-familiarities and did not attempt to determine how much pronunciation scores impact holistic scores. Nevertheless, the matter remains that pronunciation scores do impact holistic scores on tests like the TOEFL iBT. There is no research that I am aware of that has examined how TOEFL iBT raters compartmentalize their different ratings of the four sub-section scoring categories ('general description', 'delivery', 'language use' and 'topic development') when making final score decisions for the integrated and independent speaking tasks. Such research is recommended, and would be invaluable to possibly determining the precise impact raters' accent familiarity differences have on holistic scores; however, the simplest solution ETS could make would be reconsidering how the current hybridized scales are utilized for integrated and independent speaking tasks on the TOEFL iBT. The current hybridized holistic scales could be better utilized as multiple trait scales. They already include score-specific descriptions in the rubrics, and it seems certain that raters are at least considering these when making their final holistic scoring decisions. By fully implementing and disclosing test-takers' scores for each trait would benefit all stake-holders. Test-takers would be greater informed of their skill-specific performance, and the question of how much pronunciation scores impact holistic scores would be redundant. In section 6.3 how accent-familiarity affected intelligibility in the study is examined.

6.3 Accent-familiarity levels and their effects on intelligibility

<u>Hypothesis 3</u>: The higher the level of accent-familiarity a rater has with the L2 accented speech of speakers from a specific L1, the more intelligible the utterances produced by those speakers will be.

Finding (4): Level of accent-familiarity was determined to affect raters' abilities to accurately transcribe the utterances of the speaker-participants. Significant differences of overall intelligibility were determined between rater-groups transcribing Spanish-English and Japanese-English, and instances of significant intelligibility variance were determined between familiarity levels on individual test items from all four accents.

Finding (5): Significant, positive correlations occurred between raters' accent familiarity levels and intelligibility success transcribing Japanese-, Spanish- and Dhivehi-English. The strengths and effect sizes of each correlation, however, were not equal between the four accents investigated.

These findings support those of Gass and Varonis (1984) and Bradlow and Bent (2003) that the amount of familiarity with a particular accent a listener has positively affects their ability to find speech from speakers of that particular accent more intelligible. The significant positive correlations between accent-familiarity levels and intelligibility with three of the four investigated accents provides additional support to the claim that *secondary interlanguage speech intelligibility benefits* can occur as a result of increased accent-familiarity. The findings also support the veracity of both the Perceptual Magnet Effect and Exemplar theory; exposure to particular accents positively impacts raters' speech perception of those accents. The evidence presented suggests that the hypothesis was not rejected. Accent-familiarity level did affect raters' abilities to accurately transcribe the speech of the speaker-participants, and the effect was significant at times with each accent.

This helps to explain how intelligibility and 'listener effort' are connected. Prior to this study, it could only be speculated that the significant score variance observed in studies like Carey et al. (2011) and Winke et al. (2011) occurred because either raters' speech processing was affected by accent-familiarity or because of rater bias. Pronunciation rating scales that rely on raters' 'listener effort' are arguably flawed, as they rely only on raters' subjective reasoning between feelings of 'some', 'significant' or 'considerable' listener effort to determine scores. Intelligibility measures included in this study, however, provided evidence revealing that accent-familiarity reduces the necessary listener effort needed to process speech.

Intelligibility and pronunciation scores are connected. Intelligibility differences, at least in part, account for the score variance observed in raters' applications of the pronunciation score rating scales. The effect sizes of the correlations between intelligibility and familiarity levels were not large, and little significant evidence was provided concerning Arabic-English, but significant pronunciation score variance can occur due to insignificant variance in intelligibility. Part of the problem concerning 'listener effort' estimations used to determine scores in semi-direct tests and how intelligibility contributes to rater-decision-making is that from the rater's perspective, intelligibility can only be inferred but never confirmed. When pronunciation scores are based on spontaneous speech samples, it is not possible for raters to refer to a transcript of the test-takers' utterance. There is no means for raters to determine if what they believed they perceived in fact matched the actual content of the test-candidate's utterance. Test designers could address this issue by including 'reading-aloud' tasks that would allow for raters to better determine the intelligibility of test-candidates' pronunciation.

The main implication of these findings is that it is not possible for testing agencies to claim that all raters can, or will, process all test-takers' speech equally. Some raters will process certain test-takers' speech more effectively than others, and raters' accent-familiarity levels will, at least partly, contribute to the differences in speech processing effectiveness. Since accent-familiarity level impacts not only pronunciation scoring but intelligibility, it is not likely that a single rater-training session could *normalize* a pool of raters' scoring and speech processing capabilities with all possible accents of test-takers. Rater training should instead require accent-specific sessions designed to train raters for scoring certain groups of test-takers, not unlike the specialized training the treatment group received in Xi and Mollaun's (2009) study where some raters were provided with a specialized set of benchmark responses of only Indian-English. Such accent-specific training sessions could not only scrutinize raters' pronunciation scoring, but also examine raters' intelligibility success similarities. In the next section the findings concerning the 'very familiar' level of familiarity are discussed, and an argument is presented suggesting that perhaps the best potential solution to the problem of raters' differing familiarity levels impacting scores may be by utilizing primarily raters who are 'very familiar' with each accent to score those test-takers' pronunciation.

6.4 The 'very familiar' familiarity level and 'bias for best' (Fox, 2004)

Hypothesis 4: Raters 'very familiar' with a particular accent will demonstrate the most leniency scoring the pronunciation of speakers of that accent.

Hypothesis 5: Raters 'very familiar' with a particular accent will be more successful transcribing the pronunciation of speakers of that accent than raters with less familiarity.

Finding (6): Raters that reported being 'very familiar' with all included accents except

Arabic-English tended to show the greatest leniency when scoring pronunciation. The findings determined that when the raters were divided into either ‘very familiar’ or ‘not very familiar’, the ‘very familiar’ rater-subgroups’ mean pronunciation scores were higher than the mean scores from all other raters. Significant score variance between the two groups were determined for ten of the seventeen speaker-participants’ ($p \leq .001$, d=.6~2.1, n=6; $p \leq .05$, d=.4~1.69, n=4).

Finding (7): Evidence suggesting accent-familiarity level was linked to positive bias in pronunciation scores was determined. The findings from the Spanish-English and Dhivehi-English revealed that as raters’ accent-familiarity increased so, too, did the measures of positive bias. However, only two instances of significant bias toward individual speaker-participants by ‘very familiar’ raters were observed (Spanish-English n=1; Dhivehi-English n=1). The Arabic-English results revealed no instances of significant bias, but the findings did indicate an increase in positive bias between the ‘limited familiarity’ and ‘some familiarity’ levels with no increase at the ‘very familiar’ level.

Gass and Varonis (1984) concluded that the more familiarity a listener has with different aspects of either the speaker or the content of a speaker’s utterance will positively affect the listener’s ability to transcribe the contents of a speaker’s utterances. This study’s findings support those of Gass and Varonis, and provide the added aspect of levels to accent-familiarity. The higher the level, the greater the benefit to speech perception is possible or likely. Pronunciation scales in tests like the TOEFL iBT that rely on raters’ estimations of listener effort to make scoring decisions require a kind of rater-homogeneity of speech-processing ability that this study has established does not exist. The findings show that raters ‘very familiar’ with a particular accent demonstrate the greatest overall intelligibility success coping with that accent than other familiarity levels. Very familiar raters, this study has shown, also exhibit greater pronunciation scoring leniency than all other raters, likely due to the intelligibility benefits from being very familiar with the accent. From these findings the question emerges concerning what accent-familiarity level of raters are most appropriate for scoring pronunciation.

In order to achieve the highest measures of both inter-rater and intra-rater reliability, I argue that raters 'very familiar' with any accent would be the most appropriate for scoring those test-candidates' pronunciation. It is true that the 'very familiar' raters demonstrated the highest measures of scoring bias in this study, and as a result, an argument could be made that 'very familiar' raters with a particular accent should not be assigned to score the pronunciation of test-takers of that accent; however, such a decision depends on how bias should be interpreted. While the findings did reveal some positive bias occurred from the 'very familiar' raters, it is also arguable that the findings equally revealed negative bias from all other raters. I argue in favor of an approach of "biasing for the best" (see Fox, 2004, p.235) concerning pronunciation scoring based on two lines of reasoning:

1. Only the 'very familiar' level of accent-familiarity can be considered permanent.
2. Test scores should reflect the best possible outcome of a test-taker's performance.

It would be beneficial if raters' speech processing capabilities with a particular accent remain constant. All levels of accent-familiarity other than 'very familiar' can be considered as 'developing stages'. They are periods of accent-familiarity growth. A rater can no longer be considered to have 'no familiarity' with any given accent once they have encountered it for the first time, and the levels 'limited' and 'some' familiarity cannot remain constant, as raters' familiarity at these stages may gradually increase with continued exposure. Consistent speech processing capabilities seems the best means to maintain raters' intra-rater reliability, though this particular point was not included in this study. Research is therefore recommended to investigate whether or not intra-rater reliability of pronunciation scoring is affected throughout raters' accent-familiarity level development.

Testing agencies should adopt a notion of 'biasing for the best' when attempting to address raters' accent-familiarity levels as a potential source of pronunciation score bias. In an interview with Merrill Swain (Fox, 2004) she discussed testing that allowed for eliciting the best performance from test-takers and called it "biasing for the best". It is an ethical argument that test-takers should be able to receive the best possible score their performance can achieve. Significant score variance has been determined due to raters' differing levels of accent-familiarity, and it has also been determined that raters 'very familiar' are most likely to deliver the highest scores for pronunciation. The best way that testing agencies can guarantee that test-takers will receive the highest possible score based on their performance is by assigning raters that are very familiar with their particular accent to score their pronunciation. Of course, this suggestion is limited to semi-direct tests like the TOEFLiBT that are used to determine whether or not a test candidate may gain entrance into a university, but would not be advisable, for example, for rater to test-candidate selection for the test of English Language Proficiency for Aeronautical Communication (ELPAC). In such a case where test-candidates' English proficiency is needed to ensure the safety of lives, it would be advisable that raters be assigned to score test-takers with as little familiarity as possible with the test-takers' accents.

The implications of biasing for the best would require changes be made to how test rating is administered. Again, assigning raters randomly to score test-takers' speech samples would not be possible, and instead would include deliberately managed assignments. This approach could reduce or eliminate significant score variance. It would also be an ethical response to the problem because it would be a policy of assigning raters with the most personal experience with each test-taker's accent to determine pronunciation performance, and could be considered as an additional measure of quality control rating. In the next section some limitations of this study are discussed, as well as suggestions for future research.

6.5 Limitations and recommendations for future research

Here some limitations of the study not previously addressed are presented, and suggestions for future research are included. Again, the findings and implications of this thesis are limited to semi-direct tests, and differences may occur in traditional face-to-face speaking tests. As the findings showed, the impact raters' differing levels of accent-familiarity have on pronunciation scores vary from accent to accent. One limitation of the study is its lack of ability to determine a single, generalized description of precisely how much raters' accent-familiarity differences affect pronunciation scores for all accents. One can only speculate what the measure of impact raters' accent-familiarity differences may have on scores for any other accent not investigated in this study; however, based on the findings of this study the likelihood that the impact would be significant is high. It is therefore recommended that continued research be conducted both with the accents included in this study and with all other accents. Though the matter concerning the lack of training the raters had was previously discussed, it is worth mentioning that the results could have been different if the participants were all trained, experienced raters. Such experience, if researched, may reveal raters have developed compensatory strategies when scoring speakers with accents they are very familiar with. Concerning the intelligibility tasks and resulting data, since the rater-participants were able to listen to each speaker's utterances as many times as they wished, it is likely the resulting data would have been different had the raters been only allowed one listening. It is possible that a greater impact to intelligibility may have been revealed due to accent-familiarity level. Finally, though it was mentioned earlier that the rater-population's diversity concerning their familiarity with Dhivehi-English was less than hoped for, the decision to include an accent that can fairly be judged as *not well known* in the global English speaking community should not be considered a limitation of the study. It is of particular importance that *lesser-known accents* like Dhivehi-English be researched because it is these test-takers' scores that are most likely at risk of receiving lower scores than test-candidates with accents more globally well-known. Test-takers with these accents are least likely to be scored by a rater very familiar with their accent, and it is probable that a randomly assigned rater will have no prior familiarity at all with their accent. Research of these accents could include examining pronunciation scores between raters with different levels of familiarity for variance as in this study. Research could also scrutinize pronunciation scores from test-takers from both well-known and lesser-known accents that received the same scores for reading and listening. It is possible that significant pronunciation score variance is occurring between these accent groups that is due to raters' accent-familiarity differences. Overall, it is believed that the limitations of this study were outweighed by the potential benefits to the field of language testing.

6.6 Concluding remarks

Raters' accent-familiarity level differences are impacting pronunciation scores, and this research has shown that accent-familiarity differences also affect intelligibility. Thus, there are four clear implications from this research for the providers of large-scale international tests:

1. Random assignment of raters to test-takers should be replaced with appointing raters 'very familiar' with test-takers' accents, or at least limiting rater assignment to those that have received specific rater training and evaluation of scoring test-takers with that particular accent.
2. Testing organisations such as ETS should consider retiring the current hybridized holistic scales used for scoring the independent and integrated speaking tasks that include scoring rubrics for 'general description', 'delivery', 'language use' and 'topic development'. It would probably be better to consider utilizing the existing scales as a multiple trait scoring system rather than using the current *not-so-holistic* scale. There is no logical reason I can think of to employ rating scales like the ones in the TOEFL iBT and not have raters reveal the different scores they determined for each trait. Clearly since they are so defined as they are, raters are referring to them; why not collect those assessments? Test-takers would benefit more from understanding precisely how their performance was scored for each category, and this would also offer ETS greater opportunities to address and potentially control the impact raters' accent-familiarity levels have on pronunciation scores.
3. I would recommend testing agencies treat this problem not as a "potential threat" but as a verified problem. Additional research is needed, of course, but there is enough evidence both from this study and others to reliably confirm it is not

merely a theory. Raters' differing levels of accent-familiarity with test-takers' accents are affecting scores.

4. In the matter of how high-stakes tests of spoken English can best include pronunciation as a test construct, the implications of this research are that test administrators must include raters' levels of familiarity with the accents of test-takers as an impactful facet of the construct. This study supports Carey et al.'s (2011) "interlanguage phonology familiarity" (p.204) concept that raters' accent familiarity enables enhanced speech perception, and expands on Bent and Bradlow's (2003) "interlanguage speech intelligibility benefit" (p.1600) to determine that raters' accent-familiarities cause what I refer to as a *secondary interlanguage speech intelligibility benefit*. Gass and Varonis (1984) determined more than thirty years ago that familiarity with a particular accent significantly affects intelligibility of speech in that accent, yet language testing research has either overlooked or dismissed how raters' accent-familiarity differences might affect inter-rater score reliability until recently.

The findings from this study provide both supportive evidence and important new information concerning our understanding of this threat to pronunciation scores. Implications and potential means to address the threat have also been provided. The intelligibility of speech in a particular accent has been shown to be affected by the amount of accent-familiarity raters have with that particular accent. As long as English pronunciation scores on high-stakes tests are based on raters' personal degree of 'listener effort', test raters' accent-familiarity levels will impact scores. Not until active measures are taken by testing agencies to attend to this problem can the validity and reliability of pronunciation scores on such tests be ensured.

Appendix A

Included in this appendix:

- A copy of the pilot study from Survey Monkey
- Examples from the original BKB-R Sentence Lists
- Logit-to-probability conversion table
- The Facets analyses output of the pronunciation score data
- The Facets analyses output of the intelligibility data

Copy of the pilot study from Survey Monkey

LT Project Pilot 1.1

Overview and instructions

Thank you for participating in this pilot study. Your contributions and comments will aid in the development of a future instrument to complete my doctoral studies.

There are 3 sections to this survey:

1. Reporting biographical, professional and linguistic experiences

2.Completing intelligibility gap-fill exercises and scoring pronunciation of different speakers' audio recordings

3. Comments and suggestions to the researcher

Overview - This pilot and future main study are the focus of my doctoral studies in Language Testing through the University of Leicester, and is partially funded by ETS through a TOEFL grant for doctoral research in second and foreign language assessment. This research aims to examine the effects of raters' familiarity of various non-native English accents on pronunciation scores on high stakes tests of English, such as the TOEFLiBT. Your contribution, comments and suggestions concerning the items, instrument, instructions and experiences while completing the survey will be vital to the success of this study and the final study to follow.

Note concerning the content of the recorded samples

- The recordings you will listen to and score in the test are comprised of prepared sentences read aloud by various non-native speakers. The sentences are influenced by BKB-R materials, (Bench, Kowald & Bamford, 1979) though the sentences include non-syntactic anomalies. In other words, the grammar of each sentence is orthodox: i.e. subject-verb-complement/object;however, key words of the same class (noun, verb, adverb, etc.) are replaced with word selections that do not normally occur in usual contexts. The reasoning for this choice of stimuli is to attempt to reduce the effect knowledge of context has on comprehension (Gass & Varonis, 1984). When scoring the samples for comprehension, it is not necessary that you understand the meaning of the sentence but only the words the speaker said.

Instructions: This test should take about 10 minutes to complete. Incomplete surveys will not be included, so make sure you have enough time available before you begin.

Please feel free to make comments about the survey at any time using the comments boxes on each page.

Thank you in advance for your participation and cooperation.

Sincerely,

Kevin C. Browne

1

Exit this survey

LT Project Pilot 1.1

Overview and instructions

1 / 9

Thank you for participating in this pilot study. Your contributions and comments will aid in the development of a future instrument to complete my doctoral studies.

There are 3 sections to this survey:

1. Reporting biographical, professional and linguistic experiences

2.Completing intelligibility gap-fill exercises and scoring pronunciation of different speakers' audio recordings

3. Comments and suggestions to the researcher

Overview - This pilot and future main study are the focus of my doctoral studies in Language Testing through the University of Leicester, and is partially funded by ETS through a TOEFL grant for doctoral research in second and foreign language assessment. This research aims to examine the effects of raters' familiarity of various non-native English accents on pronunciation scores on high stakes tests of English, such as the TOEFLiBT. Your contribution, comments and suggestions concerning the items, instrument, instructions and experiences while completing the survey will be vital to the success of this study and the final study to follow.

Note concerning the content of the recorded samples

- The recordings you will listen to and score in the test are comprised of prepared sentences read aloud by various non-native speakers. The sentences are influenced by BKB-R materials, (Bench, Kowald & Bamford, 1979) though **the sentences include non-syntactic anomalies**. In other words, the grammar of each sentence is orthodox: i.e. subject-verb-

PREVIEW & TEST

Exit this survey

LT Project Pilot 1.1

Part One - Reporting biographical, professional and linguistic experiences

2 / 9

Please answer a few short questions about your background and experiences.

*** 2. What is your home country?**

*** 3. What is your first, or native language?**

*** 4. In what country do you currently reside?**

*** 5. Are you currently an ESL/EFL teacher?**

- Yes
- No

6. Are you currently enrolled in a graduate program for education, TESOL or Applied Linguistics?

- Yes
- No

*** 7. If teaching, what level of education do you currently teach?**

- Higher Education - University, College, Junior or Technical College

Exit this survey

LT Project Pilot 1.1

Part Two - Scoring audio samples of different speakers for pronunciation

3 / 9

Instructions - In Part 1 of the test you will listen to 4 speakers.

There are 2 questions for each speaker.

You will be directed to first listen to a recorded sample via a web-link, and then complete the gap-fill exercise.

The following question will ask you to score the speaker's pronunciation using a provided scale.

Pronunciation Scale:

- **Speech is generally clear and requires little or no listener effort. Only one listening required.**
- **Speech is generally clear, with some fluidity of expression, but it exhibits minor difficulties with pronunciation and may require some listener effort at times. Only one listening required.**
- **Speech is clear at times, though it exhibits problems with pronunciation and so may require more listener effort. It was necessary to listen more than once before attempting to complete the gap fill.**
- **Consistent pronunciation difficulties cause considerable listener effort throughout the sample. It was necessary to listen more than once before attempting to complete the gap fill.**
- **Cannot comprehend at all.**

Since this test involves listening exercises, please work in a quiet room; headphones are suggested.

Exit this survey

LT Project Pilot 1.1

4 / 9

*** 12. Speaker B - Fill in the blanks. Please click on the following link and listen to the recording, and fill in the blanks with the missing words. Each blank has only one missing word.**

LT Voice 1.m4v

They are _(a)__ some _(b)__.

The __(c)__ had nine __(d)__.

a.

b.

c.

d.

Exit this survey

LT Project Pilot 1.1

5 / 9

14. Speaker C - Fill in the blanks. Please click on the following link and listen to the recording, and fill in the blanks with the missing words. Each blank has only one missing word.

LT Voice 2

The institution __(a)__ was __(b)__.

The _(c)__ made an _(d)__ _(e)__.

a.

b.

c.

d.

Exit this survey

LT Project Pilot 1.1

6 / 9

16. Speaker D - Fill in the blanks. Please click on the following link and listen to the recording, and fill in the blanks with the missing words. Each blank has only one missing word.

LT Voice 3

The _(a)__ is _(b)__ the _(c)__.

He __(d)__ his __(e)__.

a.

b.

c.

d.

Exit this survey

LT Project Pilot 1.1

7 / 9

*** 18. Speaker E - Fill in the blanks. Please click on the following link and listen to the recording, and fill in the blanks with the missing words. Each blank has only one missing word.**

LT Voice 4.m4v

The __(a)__ cut some __(b)__.

She _(c)__ _(d)__ her __(e).

a.

b.

c.

d.

Exit this survey

LT Project Pilot 1.1

8 / 9

*** 20. Speaker F - Fill in the blanks. Please click on the following link and listen to the recording, and fill in the blanks with the missing words. Each blank has only one missing word.**

LT Voice 5.m4v

__(a)__ __(b)__ the money.

The _(c)__ _(d)__ on the __(e).

a.

b.

c.

d.

Exit this survey

LT Project Pilot 1.1

Part Three - Comments and suggestions to the Researcher

9 / 9

These questions are not required for completion of the survey, but your comments and suggestions are encouraged.

22. Do you have any comments or suggestions concerning the questions?

23. Do you have any comments or suggestions concerning the audio/video links?

24. Do you have any comments or suggestions concerning the scales for measuring comprehension?

25. Do you have any comments or suggestions concerning the information collected from respondents?

26. Are there any other comments or suggestions you would like to offer the

Examples from the BKB-R Sentence Lists

Example list 1

1. The children dropped the bag.
2. The dog came back.
3. The floor looked clean.
4. She found her purse.
5. The fruit is on the ground.
6. Mother got a saucepan.
7. They washed in cold water.
8. The young people are dancing.
9. The bus left early.
10. They had two empty bottles.
11. The ball is bouncing very high.
12. Father forgot the bread.
13. The girl has a picture book.
14. The orange was very sweet.
15. He is holding his nose.
16. The new road is on the map.

Example list 2

1. The boy forgot his book.
2. A friend came for lunch.
3. The match boxes are empty.
4. He climbed his ladder.
5. The family bought a house.
6. The jug is on the shelf.
7. The ball broke the window.
8. They are shopping for cheese.
9. The pond water is dirty.
10. They heard a funny noise.
11. The police are clearing the road.
12. The bus stopped suddenly.
13. She writes to her brother.
14. The football player lost a shoe.
15. The three girls are listening.
16. The coat is on a chair.

Example list 3

1. The book tells a story.
2. The young boy left home.
3. They are climbing the tree.
4. She stood near her window.
5. The table has three legs.
6. A letter fell on the floor.
7. The five men are working.

8. He listened to his father.
9. The shoes were very dirty.
10. They went on a vacation.
11. The baby broke his cup.
12. The lady packed her bag.
13. The dinner plate is hot.
14. The train is moving fast.
15. The child drank some milk.
16. The car hit a wall.

Example list 4

1. A dish towel is by the sink.
2. The janitor used a broom.
3. She looked in her mirror.
4. The good boy is helping.
5. They followed the path.
6. The kitchen clock was wrong.
7. The dog jumped on the chair.
8. Someone is crossing the road.
9. The mailman brought a letter.
10. They are riding their bicycles.
11. He broke his leg.
12. The milk was by the front door.
13. The shirts are hanging in the closet.
14. The ground was very hard.
15. The buckets hold water.
16. The chicken laid some eggs.

Example list 5

1. The angry man shouted.
2. The dog sleeps in a basket.
3. They're drinking tea.
4. Mother opens the drawer.
5. An old woman was at home.
6. He dropped his money.
7. They broke all the eggs.
8. The kitchen window was clean.
9. The girl plays with the baby.
10. The big fish got away.
11. She's helping her friend.
12. The children washed the plates.
13. The postman comes early.
14. The sign showed the way.
15. The grass is getting long.
16. The match fell on the floor.

Logit-to-probability conversion table

Logit Difference	Probability of Success	Logit Difference	Probability of Success
5.0	99%	-5.0	1%
4.6	99%	-4.6	1%
4.0	98%	-4.0	2%
3.0	95%	-3.0	5%
2.2	90%	-2.2	10%
2.0	88%	-2.0	12%
1.4	80%	-1.4	20%
1.1	75%	-1.1	25%
1.0	73%	-1.0	27%
0.8	70%	-0.8	30%
0.5	62%	-0.5	38%
0.4	60%	-0.4	40%
0.2	55%	-0.2	45%
0.1	52%	-0.1	48%
0.0	50%	0.0	50%

Copied from *Logit and probit: what are they?* Linacre,M. 2016.
http://www.winsteps.com/winman/whatisalogit.htm

The Facets analyses output of the pronunciation score data

```
Facets (Many-Facet Rasch Measurement) Version No. 3.71.1  Copyright �(c) 1987-2013, John M. Linacre. All rights reserved.
24/02/2014 13:18:33

Rasch Analysis of Expert Ratings 24/02/2014 13:18:33
Table 1. Specifications from file "C:\Users\kevin\Dropbox\C-Leicester Stuff\Pilot\FINAL Facets Output files\Final Pronunciation FACETS Operating
File Feb6-2014.txt".

Title = Rasch Analysis of Expert Ratings 24/02/2014 13:18:33
Data file = "C:\Users\kevin\Desktop\Facets Pilot Data\Pronunciation Expanded 1 is poor 5 is clear Data.xlsx"
Output file = C:\Users\kevin\Dropbox\C-Leicester Stuff\Pilot\FINAL Facets Output files\Final Pronunciation FACETS Operating File Feb24-
2014.out.txt

; Data specification
Facets = 3
Delements = N
Non-centered = 1
Positive =  1, 2, 3
Labels =
 1,rater ; (elements = 87)
 2,Familiarity Level, G  ; (elements = 4)
 3,Speaker ; (elements = 5)
Model = ?,?,?,R5,1

; Output description
Arrange tables in order = MN
Bias/Interaction direction = plus ; ability, easiness, leniency: higher score = positive logit
Fair score = Mean
Pt-biserial = Measure
Heading lines in output data files = Y
Inter-rater coefficients reported for facet = 1
Barchart = Yes
Total score for elements = Yes
T3onscreen show only one line on screen iteration report = Y
T4MAX maximum number of unexpected observations reported in Table 4 = 100
T8NBC show table 8 numbers-barcharts-curves = NBC
Unexpected observations reported if standardized residual >= 3
Usort unexpected observations sort order = u
WHexact - Wilson-Hilferty standardization = Y

; Convergence control
Convergence = .1, .01
Iterations (maximum) = 0 ; unlimited
Xtreme scores adjusted by = .3, .5  ;(estimation, bias)

Rasch Analysis of Expert Ratings 24/02/2014 13:18:33
Table 2. Data Summary Report.

Assigning models to Data= "C:\Users\kevin\Desktop\Facets Pilot Data\Pronunciation Expanded 1 is poor 5 is clear Data.xlsx"
Total lines in data file = 91
Total data lines = 87
Responses matched to model: ?,?,?,R5,1 = 432
    Total non-blank responses found = 432
              Number of blank lines = 3
Number of missing-null observations = 3
Valid responses used for estimation = 432

Rasch Analysis of Expert Ratings 24/02/2014 13:18:33
Table 3. Iteration Report.
```

Iteration	Max. Score Residual Elements	%	Categories	Max. Logit Change Elements	Steps
PROX 1				-2.9444	
PROX 2				.3639	
JMLE 3	-232.3992	-34.0	120.8063	.9189	-2.9813
JMLE 4	-33.7712	-18.0	-26.9918	-.7561	.3667
JMLE 5	-20.4683	-8.5	21.2041	-.4025	-.2331
JMLE 6	-18.0159	-7.4	18.0115	-.3351	-.1519
JMLE 7	-14.9140	-6.1	15.0367	-.2658	-.1297
JMLE 8	-12.7264	-5.1	12.7455	-.2154	-.1110
JMLE 9	-11.1476	-4.4	10.9637	-.1779	-.0969
JMLE 10	-9.8512	-3.7	9.5571	-.1497	-.0856
JMLE 11	-8.7725	-3.3	8.4242	-.1283	-.0763
JMLE 12	-7.8629	-2.9	7.4939	-.1116	-.0685
JMLE 13	-7.0866	-2.6	6.7169	-.0984	-.0619
JMLE 14	-6.4168	-2.3	6.0580	-.0885	-.0567
JMLE 15	-5.8278	-2.1	5.4865	-.0801	-.0521
JMLE 16	-5.3063	-1.9	4.9860	-.0730	-.0480
JMLE 17	-4.8418	-1.8	4.5440	-.0668	-.0443
JMLE 18	-4.4261	-1.6	4.1510	-.0614	-.0409
JMLE 19	-4.0523	-1.5	3.7993	-.0566	-.0379
JMLE 20	-3.7150	-1.3	3.4830	-.0522	-.0351
JMLE 21	-3.4095	-1.2	3.1973	-.0483	-.0326
JMLE 22	-3.1320	-1.1	2.9382	-.0448	-.0303
JMLE 23	-2.8793	-1.0	2.7025	-.0416	-.0282
JMLE 24	-2.6486	-.9	2.4875	-.0386	-.0262
JMLE 25	-2.4377	-.9	2.2909	-.0359	-.0244
JMLE 26	-2.2445	-.8	2.1107	-.0333	-.0227
JMLE 27	-2.0673	-.7	1.9453	-.0310	-.0211
JMLE 28	-1.9044	-.7	1.7932	-.0289	-.0197
JMLE 29	-1.7546	-.6	1.6533	-.0269	-.0183
JMLE 30	-1.6167	-.6	1.5243	-.0250	-.0170
JMLE 31	-1.4897	-.5	1.4054	-.0233	-.0159
JMLE 32	-1.3725	-.5	1.2957	-.0217	-.0148
JMLE 33	-1.2645	-.4	1.1943	-.0202	-.0138
JMLE 34	-1.1647	-.4	1.1006	-.0188	-.0128
JMLE 35	-1.0725	-.4	1.0141	-.0175	-.0119
JMLE 36	-.9874	-.3	.9340	-.0162	-.0111
JMLE 37	-.9088	-.3	.8600	-.0151	-.0103
JMLE 38	-.8361	-.3	.7916	-.0140	-.0096
JMLE 39	-.7689	-.3	.7283	-.0130	-.0089
JMLE 40	-.7069	-.2	.6697	-.0121	-.0083
JMLE 41	-.6495	-.2	.6156	-.0112	-.0077
JMLE 42	-.5965	-.2	.5655	-.0104	-.0071
JMLE 43	-.5476	-.2	.5193	-.0097	-.0066
JMLE 44	-.5023	-.2	.4765	-.0090	-.0061
JMLE 45	-.4606	-.2	.4370	-.0083	-.0057
JMLE 46	-.4220	-.1	.4005	-.0077	-.0053
JMLE 47	-.3865	-.1	.3669	-.0071	-.0049
JMLE 48	-.3537	-.1	.3358	-.0066	-.0045
JMLE 49	-.3234	-.1	.3071	-.0061	-.0041
JMLE 50	-.2956	-.1	.2807	-.0056	-.0038

```
| JMLE  51      -.2699   -.1      .2564     -.0052   -.0035 |
| JMLE  52      -.2463   -.1      .2340     -.0048   -.0033 |
| JMLE  53      -.2246   -.1      .2134     -.0044   -.0030 |
| JMLE  54      -.2046   -.1      .1944     -.0040   -.0028 |
| JMLE  55      -.1862   -.1      .1770     -.0037   -.0025 |
| JMLE  56      -.1694   -.1      .1610     -.0034   -.0023 |
| JMLE  57      -.1539   -.1      .1463     -.0031   -.0021 |
| JMLE  58      -.1397    .0      .1328     -.0029   -.0020 |
| JMLE  59      -.1268    .0      .1205     -.0026   -.0018 |
| JMLE  60      -.1149    .0      .1092     -.0024   -.0016 |
| JMLE  61      -.1040    .0      .0989     -.0022   -.0015 |
| JMLE  62      -.0941    .0      .0895     -.0020   -.0014 |
+-----------------------------------------------------------+
Warning (6)! There may be 4 disjoint subsets


Rasch Analysis of Expert Ratings 24/02/2014 13:18:33
Table 4. Unexpected Responses - appears after Table 8.


Rasch Analysis of Expert Ratings 24/02/2014 13:18:33
Table 5. Measurable Data Summary.

+-------------------------------------------------+
| Cat  Score  Exp.  Resd StRes|                   |
|-----------------------------+-------------------|
| 2.56  2.56  2.56   .00  .00 | Mean (Count: 432) |
|  .86   .86   .63   .59 1.01 | S.D. (Population) |
|  .86   .86   .63   .59 1.01 | S.D. (Sample)     |
+-------------------------------------------------+
Data log-likelihood chi-square = 760.2443
Approximate degrees of freedom = 335
Chi-square significance prob.  = .0000
                                      Count   Mean   S.D.   Params
Responses used for estimation      =       432   2.56   0.86       97
Count of measurable responses      =       432
Raw-score variance of observations   =   0.74 100.00%
Variance explained by Rasch measures =   0.40  53.35%
Variance of residuals                =   0.35  46.65%

Rasch Analysis of Expert Ratings 24/02/2014 13:18:33
Table 6.0  All Facet Vertical "Rulers".

Vertical = (1*,2A,3A,S) Yardstick (columns lines low high extreme)= 0,5,-7,2,End
+-----------------------------------------------------------------------------+
|Measr|+rater      |+Familiarity Level                         |+Speaker   |Scale|
|-----+------------+-------------------------------------------+-----------+-----|
|   2 +            +                                           +           + (5) |
|     |            |                                           |           |     |
|     |            |                                           |           |     |
|     |            |                                           | Speaker E | --- |
|     | **         |                                           |           |     |
|   1 + ***        +                                           + Speaker C +     |
|     |            |                                           |           |     |
|     | **         |                                           |           |     |
|     | **         | Very Familiar                             |           |     |
|     | *******    |                                           |           |  3  |
*   0 * **         * Limited Familiarity  Some Familiarity     *           *     *
|     |            |                                           |           |     |
|     | ********   | No Familiarity                            | Speaker B |     |
|     | **         |                                           |           |     |
|     | *********  |                                           | Speaker D |     |
|  -1 + *****      +                                           +           +     |
|     | *          |                                           | Speaker F | --- |
|     | *******    |                                           |           |     |
|     | ******     |                                           |           |     |
|     | ***        |                                           |           |     |
|  -2 + *******    +                                           +           +     |
|     | ***        |                                           |           |     |
|     | *****      |                                           |           |     |
|     | *          |                                           |           |     |
|     | ****       |                                           |           |  2  |
|  -3 +            +                                           +           +     |
|     | ***        |                                           |           |     |
|     | *          |                                           |           |     |
|     |            |                                           |           |     |
|     | *          |                                           |           |     |
|  -4 +            +                                           +           +     |
|     |            |                                           |           |     |
|     |            |                                           |           | --- |
|     | *          |                                           |           |     |
|     |            |                                           |           |     |
|  -5 + *          +                                           +           +     |
|     |            |                                           |           |     |
|     |            |                                           |           |     |
|     |            |                                           |           |     |
|     |            |                                           |           |     |
|  -6 +            +                                           +           +     |
|     | *          |                                           |           |     |
|     |            |                                           |           |     |
|     |            |                                           |           |     |
|     |            |                                           |           |     |
|  -7 +            +                                           +           + (1) |
|-----+------------+-------------------------------------------+-----------+-----|
|Measr| * = 1      |+Familiarity Level                         |+Speaker   |Scale|
+-----------------------------------------------------------------------------+

S.1: Model = ?,?,?,R5
There are 4 disconnected subsets identified in Table 7.

Rasch Analysis of Expert Ratings 24/02/2014 13:18:33
Table 6.1  rater Facet Summary.

Logit:
     1     1 1   1 13 4153736715928 2722 32
 +----+----+----Q----+-S--+---M+----+S---+-Q--+
-7   -6   -5   -4   -3   -2   -1    0    1    2

Infit MnSq:
    1 1
  8114134624354331311   3 11     1  1  1     1
 +S--------M--------S+------Q--+---------+---------+
 0         1         2         3         4         5

Outfit MnSq:
    1
  82078427513741212 12  21111   1    11     1
```

```
 +S--------M--------S+-------Q-+---------+---------+
 0         1         2         3         4         5

Infit ZStd:
    3  5      1 26213 25533 133 51 423441131 1  2 12     1 11   1
 +-Q-------+----S----+-------M-+---------+-S-------+----Q----+---------+
-3         -2        -1        0         1         2         3         4

Outfit ZStd:
    3   5     1 172115252132 4 441 3632112 22111 112    1  11  1
 +-Q-------+----S----+-------M-+---------+-S-------+----Q----+---------+
-3         -2        -1        0         1         2         3         4
```

Rasch Analysis of Expert Ratings 24/02/2014 13:18:33
Table 6.2 Familiarity Level Facet Summary.

```
Logit:
                                  1 2 1
 +----+----+----+----+----+----+-Q-SMS-Q-+----+
-7   -6   -5   -4   -3   -2   -1    0    1    2

Infit MnSq:
        1 2 1
 +-----Q-SMS-Q-------+---------+---------+---------+
 0         1         2         3         4         5

Outfit MnSq:
        1 111
 +-----Q-SM+SQ-------+---------+---------+---------+
 0         1         2         3         4         5

Infit ZStd:
                      1 1 1             1
 +---------+--Q------S------M--+---S-----+Q--------+---------+---------+
-3         -2        -1        0         1         2         3         4

Outfit ZStd:
                     1    1 1              1
 +---------+-Q-------S--------M+------S--+----Q----+---------+---------+
-3         -2        -1        0         1         2         3         4
```

Rasch Analysis of Expert Ratings 24/02/2014 13:18:33
Table 6.3 Speaker Facet Summary.

```
Logit:
                              1 1 1      1 1
 +----+----+----+----+----Q----S----M----S----Q
-7   -6   -5   -4   -3   -2   -1    0    1    2

Infit MnSq:
          122
 +-------QSMS--------+---------+---------+---------+
 0         1         2         3         4         5

Outfit MnSq:
          2 3
 +-------QSMSQ-------+---------+---------+---------+
 0         1         2         3         4         5

Infit ZStd:
                     1      1 1    11
 +---------+-------Q-+---S----M+----S----Q---------+---------+---------+
-3         -2        -1        0         1         2         3         4

Outfit ZStd:
                       1    1      1 1 1
 +---------+--------Q+----S----+M-----S--+--Q------+---------+---------+
-3         -2        -1        0         1         2         3         4
```

Rasch Analysis of Expert Ratings 24/02/2014 13:18:33
Table 7.1.1 rater Measurement Report (arranged by MN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	Exact Agree. Obs %	Exact Agree. Exp %	Nu rater	
6	5	1.20	1.15	-6.17	1.19	1.66	1.0	2.71	1.3	.11	-.36	.47	17.1	14.4	73 Limited Familiarity	in subset: 2
7	5	1.40	1.35	-5.04	.97	1.02	.2	.96	.1	.98	.39	.58	21.1	20.6	69 Limited Familiarity	in subset: 2
8	5	1.60	1.47	-4.58	.88	.53	-.7	.54	-.7	1.56	.66	.59	15.7	18.8	83 Very Familiar	in subset: 1
8	5	1.60	1.70	-3.77	.88	.71	-.3	.81	-.1	1.29	.47	.59	38.3	35.2	59 No Familiarity	in subset: 3
9	5	1.80	1.79	-3.45	.83	3.49	2.8	3.58	2.9	-2.17	.07	.58	14.5	32.8	70 Limited Familiarity	in subset: 2
10	5	2.00	1.87	-3.17	.81	1.94	1.4	2.05	1.5	-.10	.55	.58	19.5	31.2	33 Very Familiar	in subset: 1
10	5	2.00	1.87	-3.17	.81	2.54	1.9	2.55	1.9	-.86	.19	.58	31.4	31.2	84 Very Familiar	in subset: 1
10	5	2.00	1.87	-3.17	.81	4.40	3.3	4.26	3.2	-2.52	.42	.58	24.3	31.2	87 Very Familiar	in subset: 1
10	5	2.00	1.98	-2.78	.81	.31	-1.3	.30	-1.4	1.78	.00	.58	45.4	37.5	47 Limited Familiarity	in subset: 2
10	5	2.00	1.98	-2.78	.81	.31	-1.3	.30	-1.4	1.78	.00	.58	45.4	37.5	53 Limited Familiarity	in subset: 2
10	5	2.00	1.98	-2.78	.81	.31	-1.3	.30	-1.4	1.78	.00	.58	45.4	37.5	57 Limited Familiarity	in subset: 2
10	5	2.00	1.98	-2.78	.81	.94	.1	.94	.1	1.06	.54	.58	42.8	37.5	64 Limited Familiarity	in subset: 2
10	5	2.00	2.01	-2.66	.81	.75	-.2	.76	-.2	1.25	.67	.58	60.0	39.7	67 Some Familiarity	in subset: 4
10	5	2.00	2.09	-2.37	.81	1.31	.6	1.34	.6	.60	.89	.58	33.3	45.2	10 No Familiarity	in subset: 3
10	5	2.00	2.09	-2.37	.81	.31	-1.3	.30	-1.4	1.78	.00	.58	55.0	45.2	41 No Familiarity	in subset: 3
10	5	2.00	2.09	-2.37	.81	.31	-1.3	.30	-1.4	1.78	.00	.58	55.0	45.2	45 No Familiarity	in subset: 3
10	5	2.00	2.09	-2.37	.81	.31	-1.3	.30	-1.4	1.78	.00	.58	55.0	45.2	46 No Familiarity	in subset: 3
10	5	2.00	2.09	-2.37	.81	1.50	.9	1.54	.9	.38	.79	.58	36.7	45.2	86 No Familiarity	in subset: 3
11	5	2.20	2.16	-2.14	.79	.45	-.8	.38	-.9	1.56	.49	.61	53.3	40.5	7 Limited Familiarity	in subset: 2
11	5	2.20	2.16	-2.14	.79	.45	-.8	.38	-.9	1.56	.49	.61	53.3	40.5	42 Limited Familiarity	in subset: 2
11	5	2.20	2.16	-2.14	.79	2.32	1.6	2.25	1.6	-.16	.92	.61	29.6	40.5	79 Limited Familiarity	in subset: 2
11	5	2.20	2.20	-2.02	.79	.25	-1.4	.22	-1.5	1.73	.69	.61	66.7	43.2	75 Some Familiarity	in subset: 4
12	5	2.40	2.22	-1.93	.77	1.09	.3	1.29	.6	.73	.71	.63	37.8	39.1	11 Very Familiar	in subset: 1
12	5	2.40	2.22	-1.93	.77	.08	-2.3	.07	-2.3	1.82	.96	.63	38.9	39.1	16 Very Familiar	in subset: 1
12	5	2.40	2.22	-1.93	.77	.82	.0	.84	.0	1.15	.26	.63	42.7	39.1	27 Very Familiar	in subset: 1
12	5	2.40	2.22	-1.93	.77	.08	-2.3	.07	-2.3	1.82	.96	.63	38.9	39.1	28 Very Familiar	in subset: 1
12	5	2.40	2.22	-1.93	.77	1.32	.6	1.33	.6	.73	.97	.63	30.8	39.1	77 Very Familiar	in subset: 1
12	5	2.40	2.22	-1.93	.77	1.72	1.1	1.91	1.2	.18	.35	.63	36.8	39.1	82 Very Familiar	in subset: 1
11	5	2.20	2.29	-1.73	.79	1.77	1.1	1.97	1.3	.09	.16	.61	35.0	46.3	20 No Familiarity	in subset: 3
11	5	2.20	2.29	-1.73	.79	.45	-.8	.38	-.9	1.56	.49	.61	56.7	46.3	49 No Familiarity	in subset: 3
11	5	2.20	2.29	-1.73	.79	.80	-.1	.86	.0	1.10	.82	.61	43.3	46.3	76 No Familiarity	in subset: 3
12	5	2.40	2.35	-1.54	.77	.08	-2.3	.07	-2.3	1.82	.96	.63	53.9	41.5	12 Limited Familiarity	in subset: 2
12	5	2.40	2.35	-1.54	.77	.08	-2.3	.07	-2.3	1.82	.96	.63	53.9	41.5	39 Limited Familiarity	in subset: 2
12	5	2.40	2.35	-1.54	.77	1.36	.7	1.52	.8	.43	.51	.63	36.8	41.5	43 Limited Familiarity	in subset: 2
12	5	2.40	2.35	-1.54	.77	.91	.0	.85	.0	1.11	.22	.63	48.7	41.5	60 Limited Familiarity	in subset: 2
12	5	2.40	2.35	-1.54	.77	1.45	.8	1.21	.5	.84	.49	.63	48.0	41.5	78 Limited Familiarity	in subset: 2
12	5	2.40	2.35	-1.54	.77	.08	-2.3	.07	-2.3	1.82	.96	.63	53.9	41.5	81 Limited Familiarity	in subset: 2
12	5	2.40	2.39	-1.42	.77	.82	.0	.84	.0	1.15	.26	.63	60.0	43.6	71 Some Familiarity	in subset: 4

13	5	2.60	2.41	-1.36	.74	3.82	2.8	3.69	2.7	-1.91	.54	.65	17.3	40.4	4 Very Familiar	in subset: 1
13	5	2.60	2.41	-1.36	.74	.39	-1.0	.42	-.9	1.55	.66	.65	44.9	40.4	5 Very Familiar	in subset: 1
13	5	2.60	2.41	-1.36	.74	.54	-.6	.49	-.7	1.55	.93	.65	37.8	40.4	30 Very Familiar	in subset: 1
13	5	2.60	2.41	-1.36	.74	1.20	.5	1.23	.5	.76	-.10	.65	33.0	40.4	32 Very Familiar	in subset: 1
13	5	2.60	2.41	-1.36	.74	2.25	1.6	2.35	1.7	-.06	-.02	.65	30.8	40.4	51 Very Familiar	in subset: 1
13	5	2.60	2.41	-1.36	.74	.56	-.6	.66	-.3	1.35	.47	.65	34.1	40.4	72 Very Familiar	in subset: 1
12	5	2.40	2.50	-1.13	.77	1.45	.8	1.21	.5	.84	.49	.63	50.0	44.6	58 No Familiarity	in subset: 3
13	5	2.60	2.56	-.97	.74	.27	-1.4	.28	-1.3	1.67	.79	.65	49.3	40.5	35 Limited Familiarity	in subset: 2
13	5	2.60	2.56	-.97	.74	.54	-.6	.49	-.7	1.55	.93	.65	51.3	40.5	36 Limited Familiarity	in subset: 2
13	5	2.60	2.56	-.97	.74	.54	-.6	.49	-.7	1.55	.93	.65	51.3	40.5	56 Limited Familiarity	in subset: 2
13	5	2.60	2.56	-.97	.74	.39	-1.0	.42	-.9	1.55	.66	.65	45.4	40.5	68 Limited Familiarity	in subset: 2
13	5	2.60	2.56	-.97	.74	1.71	1.1	1.74	1.1	.14	.71	.65	33.6	40.5	80 Limited Familiarity	in subset: 2
10	4	2.50	2.60	-.85	.84	.32	-1.1	.32	-1.0	1.64	.76	.61	52.8	41.1	8 Limited Familiarity	in subset: 2
14	5	2.80	2.61	-.83	.72	.60	-.5	.60	-.5	1.47	.80	.66	43.8	40.2	1 Very Familiar	in subset: 1
14	5	2.80	2.61	-.83	.72	.65	-.4	.63	-.5	1.43	.78	.66	38.4	40.2	15 Very Familiar	in subset: 1
14	5	2.80	2.61	-.83	.72	1.42	.8	1.38	.7	.59	.30	.66	31.9	40.2	21 Very Familiar	in subset: 1
14	5	2.80	2.61	-.83	.72	1.31	.6	1.29	.6	.59	-.49	.66	38.9	40.2	22 Very Familiar	in subset: 1
14	5	2.80	2.61	-.83	.72	1.24	.5	1.17	.4	.81	.43	.66	42.7	40.2	29 Very Familiar	in subset: 1
14	5	2.80	2.61	-.83	.72	.78	-.2	.76	-.2	1.29	.70	.66	43.2	40.2	31 Very Familiar	in subset: 1
14	5	2.80	2.61	-.83	.72	1.24	.5	1.17	.4	.81	.43	.66	42.7	40.2	44 Very Familiar	in subset: 1
14	5	2.80	2.61	-.83	.72	.65	-.4	.63	-.5	1.43	.78	.66	38.4	40.2	52 Very Familiar	in subset: 1
13	5	2.60	2.72	-.56	.74	.54	-.6	.49	-.7	1.55	.93	.65	46.7	40.7	38 No Familiarity	in subset: 3
13	5	2.60	2.72	-.56	.74	.56	-.6	.66	-.3	1.35	.47	.65	43.3	40.7	54 No Familiarity	in subset: 3
14	5	2.80	2.77	-.43	.72	3.17	2.5	3.07	2.4	-1.25	-.74	.66	14.5	37.8	9 Limited Familiarity	in subset: 2
14	5	2.80	2.77	-.43	.72	.65	-.4	.63	-.5	1.43	.78	.66	44.1	37.8	40 Limited Familiarity	in subset: 2
14	5	2.80	2.77	-.43	.72	.47	-.9	.47	-.9	1.62	.88	.66	46.7	37.8	66 Limited Familiarity	in subset: 2
14	5	2.80	2.77	-.43	.72	.78	-.2	.76	-.2	1.29	.70	.66	40.1	37.8	74 Limited Familiarity	in subset: 2
15	5	3.00	2.82	-.32	.71	.37	-1.2	.37	-1.2	1.71	.81	.66	44.9	38.6	6 Very Familiar	in subset: 1
15	5	3.00	2.82	-.32	.71	2.34	1.8	2.34	1.8	-.39	.08	.66	32.4	38.6	13 Very Familiar	in subset: 1
15	5	3.00	2.82	-.32	.71	1.25	.5	1.23	.5	.83	.64	.66	35.1	38.6	18 Very Familiar	in subset: 1
15	5	3.00	2.82	-.32	.71	.47	-.9	.48	-.9	1.48	.00	.66	41.1	38.6	34 Very Familiar	in subset: 1
15	5	3.00	2.98	.08	.71	.54	-.7	.54	-.7	1.51	.69	.66	35.5	34.0	17 Limited Familiarity	in subset: 2
15	5	3.00	2.98	.08	.71	.95	.1	.94	.0	1.14	.79	.66	34.2	34.0	55 Limited Familiarity	in subset: 2
16	5	3.20	3.03	.19	.71	1.16	.4	1.12	.3	.88	.45	.66	30.3	35.9	2 Very Familiar	in subset: 1
16	5	3.20	3.03	.19	.71	.34	-1.2	.34	-1.2	1.75	.95	.66	43.2	35.9	19 Very Familiar	in subset: 1
16	5	3.20	3.03	.19	.71	.34	-1.2	.34	-1.2	1.75	.95	.66	43.2	35.9	24 Very Familiar	in subset: 1
16	5	3.20	3.03	.19	.71	.95	.1	.93	.0	1.10	.57	.66	41.1	35.9	26 Very Familiar	in subset: 1
16	5	3.20	3.03	.19	.71	1.11	.3	1.11	.3	.91	.47	.66	40.5	35.9	37 Very Familiar	in subset: 1
16	5	3.20	3.03	.19	.71	1.11	.3	1.11	.3	.91	.47	.66	40.5	35.9	85 Very Familiar	in subset: 1
15	5	3.00	3.03	.20	.71	.95	.1	.94	.0	1.14	.79	.66	26.7	30.7	62 Some Familiarity	in subset: 4
9	3	3.00	3.08	.31	.91	.38	-.8	.39	-.8	1.60	.00	.73	30.1	32.6	23 Limited Familiarity	in subset: 2
15	5	3.00	3.15	.49	.71	1.65	1.1	1.68	1.1	.26	-.13	.66	15.0	28.8	61 No Familiarity	in subset: 3
17	5	3.40	3.24	.70	.72	.06	-2.7	.06	-2.7	1.95	.96	.65	38.4	32.3	3 Very Familiar	in subset: 1
17	5	3.40	3.24	.70	.72	.06	-2.7	.06	-2.7	1.95	.96	.65	38.4	32.3	25 Very Familiar	in subset: 1
17	5	3.40	3.41	1.09	.72	1.17	.4	1.22	.5	.75	.89	.65	28.3	24.8	50 Limited Familiarity	in subset: 2
17	5	3.40	3.41	1.09	.72	.06	-2.7	.06	-2.7	1.95	.96	.65	21.1	24.8	63 Limited Familiarity	in subset: 2
17	5	3.40	3.41	1.09	.72	2.62	2.0	2.56	1.9	-.55	.62	.65	27.6	24.8	65 Limited Familiarity	in subset: 2
18	5	3.60	3.46	1.22	.73	.23	-1.6	.22	-1.7	1.82	.79	.63	34.6	27.9	14 Very Familiar	in subset: 1
18	5	3.60	3.46	1.22	.73	1.43	.8	1.60	1.0	.38	.38	.63	30.3	27.9	48 Very Familiar	in subset: 1
12.7	5.0	2.56	2.49	-1.27	.77	1.00	-.2	1.01	-.2		.52				Mean (Count: 87)	
2.5	.2	.50	.47	1.38	.07	.86	1.3	.87	1.3		.37				S.D. (Population)	
2.5	.2	.50	.48	1.39	.07	.86	1.3	.88	1.3		.38				S.D. (Sample)	

Model, Populn: RMSE .77 Adj (True) S.D. 1.14 Separation 1.49 Strata 2.32 Reliability (not inter-rater) .69
Model, Sample: RMSE .77 Adj (True) S.D. 1.15 Separation 1.50 Strata 2.34 Reliability (not inter-rater) .69
Model, Fixed (all same) chi-square: 246.1 d.f.: 86 significance (probability): .00
Model, Random (normal) chi-square: 72.2 d.f.: 85 significance (probability): .84
Inter-Rater agreement opportunities: 6323 Exact agreements: 2399 = 37.9% Expected: 2338.7 = 37.0%

Rasch Analysis of Expert Ratings 24/02/2014 13:18:33
Table 7.2.1 Familiarity Level Measurement Report (arranged by MN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	N Familiarity Level	
144	65	2.22	2.32	-.38	.22	.91	-.4	.91	-.4	1.10	.64	.65	1 No Familiarity	in subset: 3
48	20	2.40	2.42	-.09	.38	.71	-.9	.69	-.9	1.32	.73	.68	3 Some Familiarity	in subset: 4
388	157	2.47	2.46	.03	.14	.92	-.6	.96	-.3	1.08	.74	.73	2 Limited Familiarity	in subset: 2
526	190	2.77	2.61	.43	.12	1.09	.9	1.12	1.2	.87	.69	.70	4 Very Familiar	in subset: 1
276.5	108.0	2.46	2.45	.00	.21	.91	-.3	.92	-.1		.70		Mean (Count: 4)	
190.0	68.4	.20	.11	.29	.10	.14	.7	.15	.8		.04		S.D. (Population)	
219.4	79.0	.23	.12	.33	.12	.16	.8	.18	.9		.04		S.D. (Sample)	

Model, Populn: RMSE .24 Adj (True) S.D. .16 Separation .68 Strata 1.25 Reliability .32
Model, Sample: RMSE .24 Adj (True) S.D. .23 Separation .98 Strata 1.64 Reliability .49
Model, Fixed (all same) chi-square: 12.3 d.f.: 3 significance (probability): .01
Model, Random (normal) chi-square: 2.4 d.f.: 2 significance (probability): .30

Rasch Analysis of Expert Ratings 24/02/2014 13:18:33
Table 7.3.1 Speaker Measurement Report (arranged by MN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	N Speaker
185	87	2.13	2.05	-1.26	.19	.85	-.9	.86	-.8	1.14	.57	.61	5 Speaker F
196	86	2.28	2.19	-.77	.19	.98	.0	1.12	.7	.95	.55	.63	3 Speaker D
208	87	2.39	2.29	-.45	.18	.95	-.2	.95	-.2	1.06	.72	.64	1 Speaker B
255	87	2.93	2.85	1.03	.17	1.07	.5	1.08	.5	.91	.72	.67	2 Speaker C
262	85	3.08	3.02	1.44	.17	1.06	.4	1.05	.4	.95	.63	.67	4 Speaker E
221.2	86.4	2.56	2.48	.00	.18	.98	-.1	1.01	.1		.64		Mean (Count: 5)
31.4	.8	.38	.38	1.05	.01	.08	.5	.09	.6		.07		S.D. (Population)
35.1	.9	.42	.43	1.17	.01	.09	.6	.11	.7		.08		S.D. (Sample)

Model, Populn: RMSE .18 Adj (True) S.D. 1.03 Separation 5.67 Strata 7.89 Reliability .97
Model, Sample: RMSE .18 Adj (True) S.D. 1.16 Separation 6.36 Strata 8.81 Reliability .98
Model, Fixed (all same) chi-square: 167.9 d.f.: 4 significance (probability): .00
Model, Random (normal) chi-square: 3.9 d.f.: 3 significance (probability): .27

Rasch Analysis of Expert Ratings 24/02/2014 13:18:33
Table 8.1 Category Statistics.

Model = ?,?,?,R5

DATA					QUALITY CONTROL			RASCH-ANDRICH		EXPECTATION		MOST	RASCH-	Cat
	Category Counts			Cum.	Avge	Exp.	OUTFIT	Thresholds		Measure at		PROBABLE	THURSTONE	PEAK
Score	Total	Used	%	%	Meas	Meas	MnSq	Measure	S.E.	Category	-0.5	from	Thresholds	Prob
1	37	37	9%	9%	-3.72	-3.77	1.0			(-5.47)		low	low	100%
2	182	182	42%	51%	-1.93	-1.93	1.1	-4.39	.21	-2.68	-4.45	-4.39	-4.41	73%

```
|  3     151        151   35%  86%|  -.47   -.42  1.0 |  -.98    .13|    .12   -1.11|   -.98 |  -1.04    | 60%|
|  4      58         58   13%  99%|  1.01    .93   .9 |  1.23    .17|   2.70    1.32|   1.23 |   1.26    | 68%|
|  5       4          4    1% 100%|  1.98   1.92  1.0 |  4.14    .52|(  5.25)   4.26|   4.14 |   4.18    |100%|
+-----------------------------------------------------------------------(Mean)---------(Modal)--(Median)------+
```

```
Scale structure

Measr:-6.0       -4.0       -2.0        0.0        2.0        4.0        6.0
        +          +          +          +          +          +          +
  Mode:<1-(^)---12--------^--------23-----^----34-------^------45----(^)--5>

Median:<1-(^)---12--------^-------23------^----34-------^------45----(^)--5>

  Mean:<1-(^)---12--------^-------23------^----34-------^------45----(^)--5>
        +          +          +          +          +          +          +
Measr:-6.0       -4.0       -2.0        0.0        2.0        4.0        6.0
```

```
Probability Curves
      -6.0       -4.0       -2.0        0.0        2.0        4.0        6.0
       ++----------+----------+----------+----------+----------+----------++
     1 |                                                                   |
       |                                                                   |
       |                                                                   |
       |1                                                                55|
       | 11                                                            55  |
     P |   1             222                                          5    |
     r |    1         222   222                        44           55     |
     o |     1       2         2                     44  444       5       |
     b |      11   22           22      33333      44       44    5        |
     a |        1 2               2   33     3    4           4  5         |
     b |         *                 233        33 4             4*          |
     i |        2 1                32          4*              5 4         |
     l |       2   1              3  2        4  3            5   4        |
     i |     22     1            3    2      4    33         5     4       |
     t |    2        1         33      2    4       3       5       44     |
     y |   2          11      3         2  4         3    55          4    |
       | 22             1   33           2*           33 5             44  |
       |2                1*3            44 22          5*3               44|
       |                33 111        44     22      55   33               |
       |           33333      11114444         2***55       3333           |
     0 |****************************************111************************|
       ++----------+----------+----------+----------+----------+----------++
      -6.0       -4.0       -2.0        0.0        2.0        4.0        6.0
```

```
Expected Score Ogive (Model ICC)
      -6.0       -4.0       -2.0        0.0        2.0        4.0        6.0
       ++----------+----------+----------+----------+----------+----------++
     5 |                                                                   |
       |                                                           55555555|
       |                                                       4455        |
       |                                                  44444            |
     4 |                                              4444                 |
       |                                           444                     |
       |                                       3344                        |
       |                                    333                            |
     3 |                                 333                               |
       |                             3333                                  |
       |                         2233                                      |
       |                     2222                                          |
     2 |                22222                                              |
       |           22222                                                   |
       |      11122                                                        |
       |111111                                                             |
     1 |                                                                   |
       ++----------+----------+----------+----------+----------+----------++
      -6.0       -4.0       -2.0        0.0        2.0        4.0        6.0
```

```
Rasch Analysis of Expert Ratings 24/02/2014 13:18:33
Table 4.1 Unexpected Responses (1 residuals sorted by u).

+------------------------------------------------------------------------------------+
| Cat  Score  Exp.  Resd StRes| Nu rater              N Familiarity Level   N Speaker  |
|-----------------------------+------------------------------------------------------|
|  2     2     1.1    .9  3.5 | 73 Limited Familiarity 2 Limited Familiarity 3 Speaker D |
|-----------------------------+------------------------------------------------------|
| Cat  Score  Exp.  Resd StRes| Nu rater              N Familiarity Level   N Speaker  |
+------------------------------------------------------------------------------------+
```

The Facets analyses output of the intelligibility data

```
Facets (Many-Facet Rasch Measurement) Version No. 3.71.1  Copyright �(c) 1987-2013, John M. Linacre. All rights reserved.
24/02/2014 13:04:14

Rasch Analysis of Expert Ratings 24/02/2014 13:04:14
Table 1. Specifications from file "C:\Users\kevin\Dropbox\C-Leicester Stuff\Pilot\FINAL Facets Output files\Intelligibility
Results\intelligibility Scores Expanded File for Analysis NEWNEW Feb 24.txt".

Title = Rasch Analysis of Expert Ratings 24/02/2014 13:04:14
Data file = "C:\Users\kevin\Dropbox\Yahoo! Mail\Inteligibility Rater Expanded.xls"
Output file = C:\Users\kevin\Dropbox\C-Leicester Stuff\Pilot\FINAL Facets Output files\Intelligibility Results\intelligibility Scores Expanded
File for Analysis NEWNEW Feb 24.out.txt

; Data specification
Facets = 3
Delements = N
Non-centered = 1
Positive =  1, 2, 3
Labels =
 1,rater ; (elements = 87)
 2,Familiarity Level, G  ; (elements = 4)
 3,Item ; (elements = 24)
Model = ?,?,?,D,1

; Output description
Arrange tables in order = MN
Bias/Interaction direction = plus ; ability, easiness, leniency: higher score = positive logit
Fair score = Mean
Pt-biserial = Measure
Heading lines in output data files = Y
Inter-rater coefficients reported for facet = 1
Barchart = Yes
Total score for elements = Yes
T3onscreen show only one line on screen iteration report = Y
T4MAX maximum number of unexpected observations reported in Table 4 = 100
T8NBC show table 8 numbers-barcharts-curves = NBC
Unexpected observations reported if standardized residual >= 3
Usort unexpected observations sort order = u
WHexact - Wilson-Hilferty standardization = Y

; Convergence control
Convergence = .1, .01
Iterations (maximum) = 0 ; unlimited
Xtreme scores adjusted by = .3, .5  ;(estimation, bias)

Rasch Analysis of Expert Ratings 24/02/2014 13:04:14
Table 2. Data Summary Report.

Assigning models to Data= "C:\Users\kevin\Dropbox\Yahoo! Mail\Inteligibility Rater Expanded.xls"
Total lines in data file = 91
Total data lines = 87
Responses matched to model: ?,?,?,D,1 = 2088
    Total non-blank responses found = 2088
              Number of blank lines = 3
Valid responses used for estimation = 2088

Rasch Analysis of Expert Ratings 24/02/2014 13:04:14
Table 3. Iteration Report.

+-----------------------------------------------------------+
| Iteration      Max. Score Residual      Max. Logit Change |
|            Elements    %  Categories   Elements    Steps  |
|-----------------------------------------------------------|
| PROX   1     Recount required            4.7184           |
| PROX   2                                 1.4950           |
| JMLE   3   -131.9003 -25.4      .0000      .5522    .0000 |
| JMLE   4    -55.4440 -12.5      .0000    -.2522    .0000 |
| JMLE   5    -28.3212  -7.0      .0000    -.1423    .0000 |
| JMLE   6    -15.5668  -4.0      .0000    -.0827    .0000 |
| JMLE   7     -8.9160  -2.4      .0000    -.0483    .0000 |
| JMLE   8     -5.2574  -1.4      .0000    -.0283    .0000 |
| JMLE   9     -3.1737   -.8      .0000      .0167    .0000 |
| JMLE  10     -1.9550   -.5      .0000      .0106    .0000 |
| JMLE  11     -1.2264   -.3      .0000      .0069    .0000 |
| JMLE  12      -.7823    .2      .0000      .0046    .0000 |
| JMLE  13      -.5068    .1      .0000      .0032    .0000 |
| JMLE  14      -.3331    .1      .0000      .0023    .0000 |
| JMLE  15      -.2219    .1      .0000      .0017    .0000 |
| JMLE  16      -.1497    .0      .0000      .0013    .0000 |
| JMLE  17      -.1021    .0      .0000      .0010    .0000 |
| JMLE  18      -.0704    .0      .0000      .0007    .0000 |
+-----------------------------------------------------------+
Warning (6)! There may be 4 disjoint subsets


Rasch Analysis of Expert Ratings 24/02/2014 13:04:14
Table 4. Unexpected Responses - appears after Table 8.


Rasch Analysis of Expert Ratings 24/02/2014 13:04:14
Table 5. Measurable Data Summary.

+-------------------------------------------------+
| Cat  Score  Exp.  Resd StRes|                   |
|-----------------------------+-------------------|
|  .58   .58   .58   .00  .00 | Mean (Count: 1914) |
|  .49   .49   .35   .35 1.01 | S.D. (Population)  |
|  .49   .49   .35   .35 1.01 | S.D. (Sample)      |
+-------------------------------------------------+
Data log-likelihood chi-square = 1467.5243
Approximate degrees of freedom = 1803
Chi-square significance prob.  = 1.0000
                                      Count   Mean   S.D.   Params
Responses used for estimation      =    1914   0.58   0.49      111
Responses in one extreme score     =     174   1.00   0.00        2
All Responses                      =    2088   0.61   0.49      113
Count of measurable responses      =    2088
Raw-score variance of observations   =   0.24 100.00%
Variance explained by Rasch measures =   0.12  50.12%
Variance of residuals                =   0.12  49.88%

Rasch Analysis of Expert Ratings 24/02/2014 13:04:14
Table 6.0  All Facet Vertical "Rulers".
```

```
Vertical = (1*,2A,3*,S) Yardstick (columns lines low high extreme)= 0,5,-6,4,End
+----------------------------------------------+
|Measr|+rater      |+Familiarity Level   |+Item|
|-----+------------+---------------------+-----|
|   4 +            +                     + **  |
|     |            |                     | *   |
|     | *          |                     |     |
|     |            |                     |     |
|     |            |                     |     |
|   3 +            +                     + **  |
|     | *          |                     | *   |
|     |            |                     |     |
|     |            |                     |     |
|     | *****      |                     | *   |
|   2 + *          +                     +     |
|     | *          |                     |     |
|     | ****       |                     |     |
|     | ***        |                     | *   |
|     | *********  |                     | **  |
|   1 + *****      +                     + *   |
|     | *****      |                     |     |
|     | ***        |                     |     |
|     | *********  | Very Familiar       |     |
|     | **         |                     | *   |
*   0 * ********** * Some Familiarity    * *   *
|     | ***        | Limited Familiarity | **  |
|     | *********  | No Familiarity      | *   |
|     | ******     |                     | *** |
|     | **         |                     |     |
|  -1 + *          +                     +     |
|     | ***        |                     |     |
|     | ***        |                     |     |
|     |            |                     |     |
|     | *          |                     | *   |
|  -2 +            +                     +     |
|     |            |                     |     |
|     |            |                     |     |
|     |            |                     |     |
|     |            |                     |     |
|  -3 +            +                     + *   |
|     |            |                     |     |
|     |            |                     | *   |
|     |            |                     |     |
|     |            |                     | *   |
|  -4 +            +                     +     |
|     |            |                     |     |
|     |            |                     |     |
|     |            |                     |     |
|     |            |                     |     |
|  -5 +            +                     + *   |
|     |            |                     |     |
|     |            |                     |     |
|     |            |                     |     |
|     |            |                     |     |
|  -6 +            +                     +     |
|-----+------------+---------------------+-----|
|Measr| * = 1      |+Familiarity Level   | * = |
+----------------------------------------------+

S.1: Model = ?,?,?,D
There are 4 disconnected subsets identified in Table 7.

Rasch Analysis of Expert Ratings 24/02/2014 13:04:14
Table 6.1  rater Facet Summary.

Logit:
                               1
                      1 3312693029355934115  1   1
 +----+----+----+----+-Q--+-S--+-M--+-S--+--Q-+----+----+----+
-6   -5   -4   -3   -2   -1    0    1    2    3    4    5    6

Infit MnSq:
        1 1
    1144682785378 21
 +--Q---S--M--S---Q--+---------+---------+---------+---------+---------+---------+---------+
 0         1         2         3         4         5         6         7         8

Outfit MnSq:
     11
  22340779545  23 242    1   1               1    1                      1   1
 +---------+M--------+--S------+----Q----+---------+---------+---------+---------+
 0         1         2         3         4         5         6         7         8

Infit ZStd:
       1   12 11 2336411152241442231332331113122121    1
 +-------Q-+--------S+--------M+---------S---------+Q--------+---------+
-3        -2        -1         0         1         2         3         4

Outfit ZStd:
                        4766445354674222523 1    1  1 1        1   1
 +---------+-----Q---+---S-----+-M-------S-------Q-+---------+---------+
-3        -2        -1         0         1         2         3         4


Rasch Analysis of Expert Ratings 24/02/2014 13:04:14
Table 6.2  Familiarity Level Facet Summary.

Logit:
                             111 1
 +----+----+----+----+----+-QS-M-SQ-+----+----+----+----+----+
-6   -5   -4   -3   -2   -1    0    1    2    3    4    5    6

Infit MnSq:
         1111
 +-------QSMQ--------+---------+---------+---------+---------+---------+---------+---------+
 0         1         2         3         4         5         6         7         8

Outfit MnSq:
       1  11  1
 +---Q--S-M+-S--Q----+---------+---------+---------+---------+---------+---------+---------+
 0         1         2         3         4         5         6         7         8

Infit ZStd:
              1     1         1              1
 +--Q------+---S-----+-----M---+-------S-+---------Q---------+---------+
-3        -2        -1         0         1         2         3         4
```

```
Outfit ZStd:
                      1   1      1         1
 +---------+---Q-----+-S-------M-------S-+-----Q---+---------+---------+
-3        -2        -1         0         1         2         3         4
```

Rasch Analysis of Expert Ratings 24/02/2014 13:04:14
Table 6.3 Item Facet Summary.

```
Logit:
      1     1 1 1     1     31211   121   1  12   1         2
 +----Q----+----+---S+----+----+-M--+----+----+S---+----+----Q
-6   -5   -4   -3   -2   -1    0    1    2    3    4    5    6

Infit MnSq:
         385332
 +------QS-MS-Q------+---------+---------+---------+---------+---------+---------+
 0         1         2         3         4         5         6         7         8

Outfit MnSq:
   1   11743 11   1 3 1
 +Q----S---+M---S----Q---------+---------+---------+---------+---------+---------+
 0         1         2         3         4         5         6         7         8

Infit ZStd:
      1  1       1   1 2  112 131121 1     1 1     1       1
 +-----Q---+-------S-+---------M---------+S--------+--Q------+---------+
-3        -2        -1         0         1         2         3         4

Outfit ZStd:
           1       11 13 2  21 3 11      121       1 2
 +--------Q+---------S---------+M--------+S--------+-Q-------+---------+
-3        -2        -1         0         1         2         3         4
```

Rasch Analysis of Expert Ratings 24/02/2014 13:04:14
Table 7.1.1 rater Measurement Report (arranged by MN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	Exact Agree. Obs %	Exact Agree. Exp %	Nu rater	
7	24	.29	.14	-1.82	.65	.78	-.5	.43	.4	1.29	.71	.66	69.3	66.2	20 No Familiarity	in subset: 3
8	24	.33	.19	-1.42	.62	1.45	1.2	1.04	.6	.58	.58	.67	67.4	67.8	76 No Familiarity	in subset: 3
9	24	.38	.21	-1.33	.60	.93	-.1	.79	.3	1.07	.68	.68	65.5	66.4	9 Limited Familiarity	in subset: 2
9	24	.38	.21	-1.33	.60	1.23	.7	1.07	.5	.72	.61	.68	64.4	66.4	69 Limited Familiarity	in subset: 2
11	24	.46	.23	-1.24	.57	.97	.0	1.03	.3	.98	.66	.68	63.6	63.7	27 Very Familiar	in subset: 1
11	24	.46	.23	-1.24	.57	.82	-.5	.73	.0	1.23	.70	.68	64.4	63.7	33 Very Familiar	in subset: 1
11	24	.46	.23	-1.24	.57	.81	-.5	.54	-.1	1.31	.72	.68	64.1	63.7	44 Very Familiar	in subset: 1
10	24	.42	.27	-.98	.58	1.07	.3	.80	.2	.96	.66	.68	67.0	67.9	43 Limited Familiarity	in subset: 2
11	24	.46	.30	-.85	.57	.63	-1.4	.43	-.3	1.54	.75	.68	68.2	66.1	75 Some Familiarity	in subset: 4
10	24	.42	.33	-.70	.58	1.43	1.3	4.91	2.0	.10	.51	.68	69.7	69.9	61 No Familiarity	in subset: 3
11	24	.46	.34	-.64	.57	1.15	.5	.83	.1	.87	.65	.68	67.6	69.2	8 Limited Familiarity	in subset: 2
11	24	.46	.34	-.64	.57	.88	-.3	1.36	.6	1.00	.67	.68	72.6	69.2	53 Limited Familiarity	in subset: 2
11	24	.46	.34	-.64	.57	1.01	.1	.74	.1	1.04	.67	.68	69.6	69.2	73 Limited Familiarity	in subset: 2
13	24	.54	.36	-.59	.57	1.10	.4	.89	.1	.88	.64	.67	64.6	68.4	22 Very Familiar	in subset: 1
13	24	.54	.36	-.59	.57	.88	-.3	.59	-.2	1.24	.70	.67	66.1	68.4	28 Very Familiar	in subset: 1
13	24	.54	.36	-.59	.57	1.21	.8	.95	.2	.74	.62	.67	68.8	68.4	87 Very Familiar	in subset: 1
11	24	.46	.41	-.37	.57	1.19	.7	.83	.2	.84	.64	.68	71.6	70.3	49 No Familiarity	in subset: 3
11	24	.46	.41	-.37	.57	.69	-1.1	.49	-.2	1.45	.74	.68	76.9	70.3	58 No Familiarity	in subset: 3
11	24	.46	.41	-.37	.57	1.00	.1	.94	.3	.97	.66	.68	71.6	70.3	86 No Familiarity	in subset: 3
12	24	.50	.42	-.32	.57	1.54	1.8	4.37	2.2	-.24	.47	.68	64.4	70.0	7 Limited Familiarity	in subset: 2
12	24	.50	.42	-.32	.57	.94	-.1	.63	-.1	1.17	.69	.68	72.3	70.0	36 Limited Familiarity	in subset: 2
12	24	.50	.42	-.32	.57	.72	-1.0	.49	-.3	1.45	.73	.68	74.3	70.0	56 Limited Familiarity	in subset: 2
12	24	.50	.42	-.32	.57	.68	-1.2	.45	-.3	1.50	.74	.68	74.9	70.0	60 Limited Familiarity	in subset: 2
12	24	.50	.42	-.32	.57	.72	-1.0	.48	-.3	1.45	.73	.68	74.6	70.0	70 Limited Familiarity	in subset: 2
12	24	.50	.42	-.32	.57	.47	-2.3	.32	-.6	1.77	.78	.68	77.0	70.0	80 Limited Familiarity	in subset: 2
14	24	.58	.43	-.27	.57	1.47	1.6	1.78	1.0	.21	.53	.67	67.3	70.4	32 Very Familiar	in subset: 1
14	24	.58	.43	-.27	.57	1.44	1.6	1.89	1.1	.24	.54	.67	62.9	70.4	34 Very Familiar	in subset: 1
14	24	.58	.43	-.27	.57	1.54	1.9	7.18	3.6	-.40	.44	.67	61.2	70.4	83 Very Familiar	in subset: 1
12	24	.50	.49	-.04	.57	1.75	2.3	1.91	1.0	-.16	.48	.68	63.6	70.3	10 No Familiarity	in subset: 3
12	24	.50	.49	-.04	.57	.53	-2.0	.36	-.5	1.69	.77	.68	75.8	70.3	41 No Familiarity	in subset: 3
13	24	.54	.50	.00	.57	1.45	1.6	.99	.2	.47	.58	.67	66.1	70.6	17 Limited Familiarity	in subset: 2
13	24	.54	.50	.00	.57	.58	-1.8	.37	-.6	1.66	.76	.67	76.7	70.6	39 Limited Familiarity	in subset: 2
13	24	.54	.50	.00	.57	.69	-1.2	.45	-.5	1.51	.74	.67	77.0	70.6	42 Limited Familiarity	in subset: 2
13	24	.54	.50	.00	.57	.94	-.1	.78	.0	1.10	.67	.67	72.9	70.6	57 Limited Familiarity	in subset: 2
13	24	.54	.50	.00	.57	.78	-.8	.53	-.3	1.38	.72	.67	75.2	70.6	66 Limited Familiarity	in subset: 2
15	24	.63	.52	.06	.58	.75	-.9	.45	-.5	1.42	.72	.66	74.0	72.2	13 Very Familiar	in subset: 1
15	24	.63	.52	.06	.58	.91	-.2	.63	-.2	1.18	.68	.66	73.7	72.2	26 Very Familiar	in subset: 1
15	24	.63	.52	.06	.58	.71	-1.1	.42	-.6	1.46	.73	.66	73.0	72.2	84 Very Familiar	in subset: 1
13	24	.54	.57	.28	.57	.72	-1.1	.47	-.4	1.46	.73	.67	74.6	70.0	38 No Familiarity	in subset: 3
13	24	.54	.57	.28	.57	.85	-.5	.58	-.2	1.28	.70	.67	77.7	70.0	46 No Familiarity	in subset: 3
14	24	.58	.58	.32	.57	.95	-.1	.64	-.2	1.15	.68	.67	72.3	70.9	23 Limited Familiarity	in subset: 2
14	24	.58	.58	.32	.57	.83	-.6	.51	-.4	1.33	.71	.67	74.0	70.9	64 Limited Familiarity	in subset: 2
14	24	.58	.58	.32	.57	.94	-.1	.63	-.2	1.16	.68	.67	72.6	70.9	79 Limited Familiarity	in subset: 2
14	24	.58	.58	.32	.57	1.36	1.3	1.81	1.0	.30	.55	.67	67.3	70.9	81 Limited Familiarity	in subset: 2
16	24	.67	.60	.41	.60	.71	-.9	.41	-.5	1.41	.72	.66	74.4	73.7	11 Very Familiar	in subset: 1
16	24	.67	.60	.41	.60	.88	-.3	.70	.0	1.16	.67	.66	74.2	73.7	21 Very Familiar	in subset: 1
16	24	.67	.60	.41	.60	.96	.0	.73	.0	1.08	.66	.66	74.4	73.7	77 Very Familiar	in subset: 1
15	24	.63	.61	.45	.58	.66	-1.3	.39	-.6	1.52	.73	.66	77.3	71.6	67 Some Familiarity	in subset: 4
15	24	.63	.61	.45	.58	1.10	.4	.77	.0	.94	.64	.66	71.2	71.6	71 Some Familiarity	in subset: 4
15	24	.63	.66	.65	.58	.77	-.8	.46	-.5	1.39	.71	.66	76.1	70.9	12 Limited Familiarity	in subset: 2
15	24	.63	.66	.65	.58	.75	-.9	.45	-.5	1.42	.72	.66	76.4	70.9	40 Limited Familiarity	in subset: 2
15	24	.63	.66	.65	.58	1.41	1.4	2.80	1.7	.11	.50	.66	65.2	70.9	78 Limited Familiarity	in subset: 2
17	24	.71	.69	.78	.63	1.43	1.2	1.53	.7	.47	.52	.65	71.3	74.9	1 Very Familiar	in subset: 1
17	24	.71	.69	.78	.63	1.19	.6	7.59	3.2	.43	.52	.65	73.7	74.9	5 Very Familiar	in subset: 1
17	24	.71	.69	.78	.63	.67	-.9	.39	-.4	1.39	.71	.65	75.7	74.9	25 Very Familiar	in subset: 1
17	24	.71	.69	.78	.63	1.71	1.8	1.73	.9	.20	.47	.65	69.5	74.9	29 Very Familiar	in subset: 1
17	24	.71	.69	.78	.63	1.19	.6	1.12	.4	.79	.59	.65	73.0	74.9	51 Very Familiar	in subset: 1
15	24	.63	.72	.93	.58	.84	-.5	.50	-.4	1.31	.70	.66	73.1	68.3	45 No Familiarity	in subset: 3
15	24	.63	.72	.93	.58	.62	-1.5	.37	-.7	1.57	.74	.66	70.8	68.3	59 No Familiarity	in subset: 3
16	24	.67	.73	1.00	.60	.88	-.3	.68	-.1	1.16	.67	.66	72.3	70.5	47 Limited Familiarity	in subset: 2
16	24	.67	.73	1.00	.60	.67	-1.1	.39	-.6	1.46	.73	.66	76.7	70.5	65 Limited Familiarity	in subset: 2
16	24	.67	.73	1.00	.60	.69	-1.0	.40	-.5	1.43	.72	.66	73.8	70.5	68 Limited Familiarity	in subset: 2
18	24	.75	.77	1.20	.67	.70	-.6	.89	.3	1.18	.67	.64	76.9	75.7	2 Very Familiar	in subset: 1
18	24	.75	.77	1.20	.67	1.02	.1	1.13	.5	.92	.60	.64	76.4	75.7	15 Very Familiar	in subset: 1
18	24	.75	.77	1.20	.67	.55	-1.1	.32	-.3	1.42	.72	.64	77.9	75.7	16 Very Familiar	in subset: 1
18	24	.75	.77	1.20	.67	.95	.0	1.08	.4	.99	.62	.64	77.1	75.7	30 Very Familiar	in subset: 1
18	24	.75	.77	1.20	.67	1.36	.9	.85	.2	.77	.57	.64	73.2	75.7	31 Very Familiar	in subset: 1
18	24	.75	.77	1.20	.67	1.47	1.1	1.44	.7	.57	.52	.64	72.7	75.7	37 Very Familiar	in subset: 1
18	24	.75	.77	1.20	.67	1.34	.8	.84	.2	.78	.57	.64	71.0	75.7	52 Very Familiar	in subset: 1
18	24	.75	.77	1.20	.67	.93	.0	1.01	.4	1.02	.63	.64	76.4	75.7	72 Very Familiar	in subset: 1
18	24	.75	.77	1.20	.67	1.54	1.2	1.48	.7	.50	.51	.64	72.0	75.7	82 Very Familiar	in subset: 1
17	24	.71	.80	1.37	.63	.84	-.3	.80	.1	1.13	.66	.65	70.2	69.9	35 Limited Familiarity	in subset: 2
17	24	.71	.80	1.37	.63	1.13	.4	2.39	1.3	.70	.57	.65	70.8	69.9	50 Limited Familiarity	in subset: 2
17	24	.71	.80	1.37	.63	1.14	.5	1.08	.4	.84	.60	.65	66.1	69.9	74 Limited Familiarity	in subset: 2
18	24	.75	.83	1.59	.67	1.03	.2	.64	.0	1.03	.63	.64	71.2	69.7	62 Some Familiarity	in subset: 4

19	24	.79	.84	1.68	.72	1.40	.9	1.79	.9	.60	.51	.62	72.0	76.0	4 Very Familiar	in subset: 1
19	24	.79	.84	1.68	.72	1.45	1.0	1.80	.9	.55	.49	.62	71.3	76.0	14 Very Familiar	in subset: 1
19	24	.79	.84	1.68	.72	1.30	.7	1.70	.8	.68	.53	.62	74.2	76.0	85 Very Familiar	in subset: 1
18	24	.75	.86	1.79	.67	.51	-1.3	.29	-.4	1.45	.73	.64	75.5	68.9	63 Limited Familiarity	in subset: 2
18	24	.75	.89	2.07	.67	1.08	.3	.66	.0	1.00	.62	.64	63.6	63.2	54 No Familiarity	in subset: 3
20	24	.83	.90	2.23	.77	1.08	.3	.77	.5	.95	.56	.59	73.5	75.6	3 Very Familiar	in subset: 1
20	24	.83	.90	2.23	.77	1.78	1.4	1.46	.8	.43	.43	.59	69.8	75.6	6 Very Familiar	in subset: 1
20	24	.83	.90	2.23	.77	.30	-1.8	.13	.0	1.51	.71	.59	79.9	75.6	18 Very Familiar	in subset: 1
20	24	.83	.90	2.23	.77	1.33	.7	.68	.4	.86	.54	.59	71.7	75.6	48 Very Familiar	in subset: 1
19	24	.79	.91	2.27	.72	.35	-1.7	.18	-.3	1.51	.73	.62	73.2	67.4	55 Limited Familiarity	in subset: 2
21	24	.88	.95	2.87	.84	.46	-1.2	.15	.4	1.42	.63	.54	76.9	74.3	19 Very Familiar	in subset: 1
22	24	.92	.97	3.65	.94	.69	-.4	.17	1.0	1.29	.52	.46	73.5	72.2	24 Very Familiar	in subset: 1
14.8	24.0	.61	.58	.44	.62	.99	-.1	1.08	.2		.64				Mean (Count: 87)	
3.2	.0	.13	.21	1.05	.07	.34	1.1	1.24	.8		.09				S.D. (Population)	
3.2	.0	.14	.21	1.06	.07	.34	1.1	1.24	.8		.09				S.D. (Sample)	

Model, Populn: RMSE .62 Adj (True) S.D. .85 Separation 1.37 Strata 2.16 Reliability (not inter-rater) .65
Model, Sample: RMSE .62 Adj (True) S.D. .86 Separation 1.38 Strata 2.18 Reliability (not inter-rater) .66
Model, Fixed (all same) chi-square: 216.7 d.f.: 86 significance (probability): .00
Model, Random (normal) chi-square: 66.9 d.f.: 85 significance (probability): .93
Inter-Rater agreement opportunities: 28226 Exact agreements: 20238 = 71.7% Expected: 20204.4 = 71.6%

Rasch Analysis of Expert Ratings 24/02/2014 13:04:14
Table 7.2.1 Familiarity Level Measurement Report (arranged by MN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	N Familiarity Level	
156	312	.50	.51	-.41	.16	.99	.0	1.04	.2	1.02	.68	.68	1 No Familiarity	in subset: 3
435	768	.57	.58	-.13	.10	.91	-1.7	.91	-.4	1.11	.69	.67	2 Limited Familiarity	in subset: 2
59	96	.61	.63	.07	.30	.84	-1.0	.56	-.8	1.25	.71	.67	3 Some Familiarity	in subset: 4
634	912	.70	.71	.46	.10	1.08	1.4	1.29	1.2	.89	.63	.65	4 Very Familiar	in subset: 1
321.0	522.0	.59	.61	.00	.17	.96	-.4	.95	.0		.68		Mean (Count: 4)	
227.4	331.0	.07	.07	.32	.08	.09	1.2	.26	.8		.03		S.D. (Population)	
262.6	382.2	.08	.09	.37	.09	.10	1.4	.30	.9		.03		S.D. (Sample)	

Model, Populn: RMSE .19 Adj (True) S.D. .26 Separation 1.39 Strata 2.19 Reliability .66
Model, Sample: RMSE .19 Adj (True) S.D. .32 Separation 1.71 Strata 2.61 Reliability .74
Model, Fixed (all same) chi-square: 26.6 d.f.: 3 significance (probability): .00
Model, Random (normal) chi-square: 2.7 d.f.: 2 significance (probability): .26

Rasch Analysis of Expert Ratings 24/02/2014 13:04:14
Table 7.3.1 Item Measurement Report (arranged by MN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	Nu Item
2	87	.02	.01	-5.01	.74	1.16	.4	1.94	1.0	.87	.02	.20	1 Speaker B paying
6	87	.07	.04	-3.74	.45	.86	-.3	1.01	.2	1.07	.38	.31	20 Speaker F mine
8	87	.09	.05	-3.38	.40	1.15	.6	1.71	1.1	.88	.23	.34	24 Speaker F infant
10	87	.11	.07	-3.09	.37	.95	-.1	.76	-.3	1.05	.40	.36	14 Speaker D skill
24	87	.28	.21	-1.76	.27	.87	-.9	.81	-.7	1.20	.55	.46	10 Speaker D ladder
41	87	.47	.45	-.66	.24	.78	-2.2	.71	-2.0	1.52	.63	.48	3 Speaker B play
41	87	.47	.45	-.66	.24	.92	-.8	.88	-.7	1.21	.54	.48	17 Speaker E sensed
43	87	.49	.48	-.54	.24	1.30	2.7	1.33	1.9	.29	.26	.48	13 Speaker D cut
45	87	.52	.51	-.42	.24	.92	-.7	.87	-.8	1.21	.53	.47	23 Speaker F lie
48	87	.55	.55	-.24	.24	.86	-1.4	.83	-1.0	1.34	.57	.47	5 Speaker C organism
48	87	.55	.55	-.24	.24	1.13	1.2	1.19	1.1	.67	.36	.47	11 Speaker D across
52	87	.60	.61	.00	.25	1.00	.0	.95	-.2	1.01	.46	.46	2 Speaker B bread
54	87	.62	.64	.12	.25	.76	-2.5	.77	-1.2	1.53	.62	.45	22 Speaker F matches
68	87	.78	.82	1.09	.28	1.31	1.9	1.85	2.1	.46	.09	.38	21 Speaker F took
69	87	.79	.83	1.17	.29	1.22	1.4	1.91	2.1	.57	.13	.37	6 Speaker C wet
69	87	.79	.83	1.17	.29	.91	-.5	.78	-.5	1.12	.44	.37	9 Speaker C reader
71	87	.82	.86	1.34	.30	.95	-.2	.84	-.3	1.07	.40	.36	4 Speaker B rooms
78	87	.90	.93	2.11	.37	.97	.0	.61	-.5	1.07	.34	.28	15 Speaker E union
82	87	.94	.96	2.81	.48	1.02	.1	.76	.0	1.00	.22	.22	18 Speaker E with
83	87	.95	.97	3.06	.53	.77	-.4	.24	-.8	1.16	.40	.20	12 Speaker D door
83	87	.95	.97	3.06	.53	1.08	.3	.88	.1	.96	.15	.20	16 Speaker E onions
85	87	.98	.99	3.81	.73	1.06	.3	2.06	1.0	.92	.04	.14	7 Speaker C dog
87	87	1.00	1.00	(5.75	1.83)	Maximum					.00	.00	8 Speaker C angry
87	87	1.00	1.00	(5.75	1.83)	Maximum					.00	.00	19 Speaker E knife
53.5	87.0	.61	.61	.48	.49	1.00	-.1	1.08	.1		.32		Mean (Count: 24)
27.1	.0	.31	.33	2.71	.43	.16	1.2	.49	1.1		.20		S.D. (Population)
27.7	.0	.32	.34	2.77	.44	.16	1.2	.50	1.2		.20		S.D. (Sample)

With extremes, Model, Populn: RMSE .65 Adj (True) S.D. 2.63 Separation 4.06 Strata 5.74 Reliability .94
With extremes, Model, Sample: RMSE .65 Adj (True) S.D. 2.69 Separation 4.15 Strata 5.86 Reliability .95
Without extremes, Model, Populn: RMSE .39 Adj (True) S.D. 2.26 Separation 5.76 Strata 8.01 Reliability .97
Without extremes, Model, Sample: RMSE .39 Adj (True) S.D. 2.32 Separation 5.90 Strata 8.20 Reliability .97
With extremes, Model, Fixed (all same) chi-square: 568.8 d.f.: 23 significance (probability): .00
With extremes, Model, Random (normal) chi-square: 20.6 d.f.: 22 significance (probability): .55

Rasch Analysis of Expert Ratings 24/02/2014 13:04:14
Table 8.1 Category Statistics.

Model = ?,?,?,D

DATA Score	Category Counts Total	Used	%	Cum. %	QUALITY CONTROL Avge Meas	Exp. Meas	OUTFIT MnSq
0	804	804	42%	42%	-1.51	-1.51	1.0
1	1284	1110	58%	100%	2.02	2.02	1.1

Rasch Analysis of Expert Ratings 24/02/2014 13:04:14
Table 4.1 Unexpected Responses (36 residuals sorted by u).

Cat	Score	Exp.	Resd	StRes	Nu rater	N Familiarity Level	Nu Item
0	0	1.0	-1.0	-9.0	5 Very Familiar	4 Very Familiar	7 Speaker C dog
1	1	.0	1.0	9.0	61 No Familiarity	1 No Familiarity	24 Speaker F infant
1	1	.0	1.0	9.0	83 Very Familiar	4 Very Familiar	1 Speaker B paying
1	1	.0	1.0	8.1	7 Limited Familiarity	2 Limited Familiarity	20 Speaker F mine
1	1	.0	1.0	6.6	50 Limited Familiarity	2 Limited Familiarity	1 Speaker B paying
0	0	1.0	-1.0	-6.0	78 Limited Familiarity	2 Limited Familiarity	16 Speaker E onions
0	0	1.0	-1.0	-5.2	4 Very Familiar	4 Very Familiar	6 Speaker C wet
0	0	1.0	-1.0	-5.2	85 Very Familiar	4 Very Familiar	6 Speaker C wet

Cat	Score	Exp.	Resd	StRes	Nu rater	N Familiarity Level	Nu Item
0	0	1.0	-1.0	-5.0	14 Very Familiar	4 Very Familiar	21 Speaker F took
1	1	.0	1.0	4.9	34 Very Familiar	4 Very Familiar	24 Speaker F infant
0	0	1.0	-1.0	-4.6	53 Limited Familiarity	2 Limited Familiarity	7 Speaker C dog
0	0	1.0	-1.0	-4.5	32 Very Familiar	4 Very Familiar	18 Speaker E with
0	0	1.0	-1.0	-4.5	83 Very Familiar	4 Very Familiar	18 Speaker E with
1	1	.1	.9	4.2	81 Limited Familiarity	2 Limited Familiarity	14 Speaker D skill
0	0	.9	-.9	-4.1	6 Very Familiar	4 Very Familiar	22 Speaker F matches
0	0	.9	-.9	-4.1	30 Very Familiar	4 Very Familiar	6 Speaker C wet
0	0	.9	-.9	-4.1	37 Very Familiar	4 Very Familiar	6 Speaker C wet
0	0	.9	-.9	-4.0	2 Very Familiar	4 Very Familiar	21 Speaker F took
0	0	.9	-.9	-4.0	15 Very Familiar	4 Very Familiar	21 Speaker F took
0	0	.9	-.9	-4.0	72 Very Familiar	4 Very Familiar	21 Speaker F took
0	0	.9	-.9	-4.0	82 Very Familiar	4 Very Familiar	21 Speaker F took
0	0	.9	-.9	-3.7	7 Limited Familiarity	2 Limited Familiarity	16 Speaker E onions
0	0	.9	-.9	-3.7	10 No Familiarity	1 No Familiarity	16 Speaker E onions
0	0	.9	-.9	-3.6	51 Very Familiar	4 Very Familiar	4 Speaker B rooms
0	0	.9	-.9	-3.6	74 Limited Familiarity	2 Limited Familiarity	4 Speaker B rooms
1	1	.1	.9	3.6	78 Limited Familiarity	2 Limited Familiarity	14 Speaker D skill
1	1	.1	.9	3.5	27 Very Familiar	4 Very Familiar	10 Speaker D ladder
0	0	.9	-.9	-3.4	3 Very Familiar	4 Very Familiar	11 Speaker D across
0	0	.9	-.9	-3.3	1 Very Familiar	4 Very Familiar	6 Speaker C wet
0	0	.9	-.9	-3.3	29 Very Familiar	4 Very Familiar	6 Speaker C wet
0	0	.9	-.9	-3.3	29 Very Familiar	4 Very Familiar	9 Speaker C reader
0	0	.9	-.9	-3.3	35 Limited Familiarity	2 Limited Familiarity	6 Speaker C wet
0	0	.9	-.9	-3.2	1 Very Familiar	4 Very Familiar	21 Speaker F took
0	0	.9	-.9	-3.2	10 No Familiarity	1 No Familiarity	18 Speaker E with
0	0	.9	-.9	-3.2	81 Limited Familiarity	2 Limited Familiarity	15 Speaker E union
0	0	.9	-.9	-3.1	86 No Familiarity	1 No Familiarity	16 Speaker E onions
Cat	Score	Exp.	Resd	StRes	Nu rater	N Familiarity Level	Nu Item

Appendix B

Included in this appendix:

- A copy of the main study from Survey Monkey
- IPA transcriptions of all sentences designed for the test in RP and General American pronunciations
- The Facets analyses output of the Spanish-English pronunciation score data
- The Facets analyses output of the Arabic-English pronunciation score data
- The Facets analyses output of the Dhivehi-English pronunciation score data
- The Facets analyses output of bias/interaction - Spanish-English
- The Facets analyses output of bias/interaction - Arabic-English
- The Facets analyses output of bias/interaction - Dhivehi-English
- The Facets analyses output of the Spanish-English intelligibility data
- The Facets analyses output of the Spanish-English intelligibility data
- The Facets analyses output of the Spanish-English intelligibility data

- **A copy of the main study from Survey Monkey**

Main Study

Overview and instructions

Thank you for your interest in joining this study. Your contribution and comments are greatly appreciated, and are vital to the research I am conducting as part of my doctoral studies in Language Testing at the University of Leicester. This research is partly funded by a TOEFL Small Grant for Doctoral Studies in Second or Foreign Language Assessment.

Research Aim

The aim of this research is to examine the effects raters' accent-familiarities have on intelligibility success-rates and pronunciation scores on high stakes tests of English.

There are 3 sections to this survey:

1. Reporting biographical, professional and linguistic experiences

2.Completing intelligibility gap-fill exercises and scoring pronunciation of 13 speakers

3. Comments and suggestions to the researcher

Instructions: This test may take about 20 minutes to complete. Incomplete tests will not be included, so please make sure you have enough time available before you begin for your contributions to be included.

Please share any comments about the test by using the comments boxes on each page.

Note concerning the content of the recorded samples

- The recordings you will listen to and score are comprised of prepared sentences read aloud. The sentences were influenced by BKB-R materials, (Bench, Kowald & Bamford, 1979); however, the contexts may be complex. The reasoning for this choice of stimuli is to attempt to reduce the effect ‘context familiarity’ has on comprehension (Gass & Varonis, 1984), and eliminate correct gap-fill answers resulting from guessing based on context. In short, the listener must be able to cope with the pronunciation of the speaker in order to accurately complete the intelligibility tasks.

Pronunciation scoring should be based solely on the quality of the pronunciation of the speaker. Again, these are not samples of spontaneous speech, so please limit your scoring to only the measure of difficulty you experienced attempting to decipher the phonetic content of the utterances.

Thank you in advance for your participation and cooperation in this research.

Sincerely,

* 1. Do you agree to permitting the researcher to include your answers and comments in the collected data, which could be included in future papers and presentations?

- ◯ Yes
- ◯ No

Main Study

Part 1: Biographical, professional and linguistic experiences

Please answer the following questions according to your personal opinions about your background and experiences.

2. What is your gender?

- ◯ Female
- ◯ Male

3. What is your age?

- ◯ 21-29
- ◯ 30-39
- ◯ 40-49
- ◯ 50-59
- ◯ 60 or older

* 4. What is your home country (the country you were raised in)?

* 5. What is your first, or native language? (if you have more than one native language please list them according to your personal ranking order)

* 6. In what country do you currently reside?

* 7. How long have you lived in the country you currently reside?

- ◯ 0-6 months
- ◯ 7 months to 1 year
- ◯ 1-2 years
- ◯ 2-5 years
- ◯ More than 5 years

* 8. Do you have ESL/EFL teaching experience?

- ◯ Yes, I currently teach ESL/EFL.
- ◯ Yes, I used to teach ESL/EFL.
- ◯ No, I have never taught ESL/EFL

* 9. If teaching, what level of education do you currently teach?(check all that apply)

- ☐ Higher Education - University, College, Junior or Technical College
- ☐ Business - Teaching at a company or business
- ☐ High School, Junior High School
- ☐ Elementary School
- ☐ Private English School
- ☐ No previous teaching experience
- ☐ Other (please specify)

* 10. What is the best description of your level of education?

- ◯ Some university study (no degree awarded)
- ◯ Completed 4 year university degree (BA/BS)
- ◯ Some graduate study (currently studying or previous study)
- ◯ Completed Master's MA/MS
- ◯ Some doctoral study (currently studying or previous study)
- ◯ Completed PhD/EdD
- ◯ Other (please specify)

11. Other than your home country, please list any countries you have lived for one or more years.

* 12. How would you best describe your familiarity with the following World Englishes or non-native accents?

	No Familiarity	Limited Familiarity - You have heard speakers of the accent but without regularity, and/or have not had students with that accent during the last 2 years.	Some Familiarity - You have taught students from the country or region with the accent in the last 2 years; (and/or) have visited the country or region; (and/or) have regular casual contact through encounters in your community, watching TV or movies from the region or any other personal reason to justify a feeling of "some familiarity".	Very Familiar - You are a native speaker of the language of the country or region, have lived in the country or region for 1 or more years, and/or studied the language as a foreign or second language for 1 or more years.
Japanese English	○	○	○	○
Korean English	○	○	○	○
Urdu English (Pakistan)	○	○	○	○
Arabic English	○	○	○	○
Indonesian English	○	○	○	○
Dhivehi English (Maldives)	○	○	○	○
Brazilian English	○	○	○	○
Spanish English	○	○	○	○
German English	○	○	○	○

Comments

Main Study

Part 2: The Speakers - gap-fill and pronunciation scoring

Instructions - In Part 2 of the test you will listen to 13 speakers.

Each speaker has an embedded video and two tasks for you to complete:

4

Task 1: completing an incomplete transcript.

Task 2: scoring the speaker's pronunciation using the provided scale.

Pronunciation Scale:

- **Speech is generally clear and requires little or no listener effort.**
- **Speech is generally clear, with some fluidity of expression, but the speaker exhibits minor difficulties with pronunciation and may require some listener effort at times.**
- **Speech is clear at times, though the speaker exhibits problems with pronunciation and so may require more listener effort. It was necessary to listen more than once before attempting to complete the gap fill.**
- **Consistent pronunciation difficulties cause considerable listener effort throughout the sample. It was necessary to listen several times before attempting to complete the gap fill.**
- **The speaker's pronunciation is unintelligible.**

Since this test involves listening exercises, please work in a quiet room; headphones are suggested.

When completing the gap fill, you may guess, but if you have no idea what the answer is leave it blank

<u>You must score each speaker for pronunciation.</u>

13. **Speaker 1 - Fill in the blanks. Please click on the following link and listen to the recording, and fill in the blanks with the missing words. Each blank has only one missing word.**

___(a)___ ___(b)___ with his ____(c)___.

The ___(d)___ was ____(e)_____.

a.

b.

c.

d.

e.

* 14. Speaker 1 - Pronunciation Score

Please now score Speaker 1 for pronunciation using the scale provided.

- ◯ Speech is generally clear and requires little or no listener effort.
- ◯ Speech is generally clear, with some fluidity of expression, but the speaker exhibits minor difficulties with pronunciation and may require some listener effort at times.
- ◯ Speech is clear at times, though the speaker exhibits problems with pronunciation and so may require more listener effort. It was necessary to listen more than once before attempting to complete the gap fill.
- ◯ Consistent pronunciation difficulties cause considerable listener effort throughout the sample. It was necessary to listen several times before attempting to complete the gap fill.
- ◯ The speaker's pronunciation is unintelligible.

Comments to the researcher

15. **Speaker 2 - Fill in the blanks. Please click on the following link and listen to the recording, and fill in the blanks with the missing words. Each blank has only one missing word.**

The ___(a)___ ___(b)___ at his ____(c)___.

The ___(d)__ ___(e)___ with his ____(f)_____.

a.

b.

c.

d.

e.

f.

* 16. Speaker 2 - Pronunciation Score

Please now score Speaker 2 for pronunciation using the scale provided.

- ◯ Speech is generally clear and requires little or no listener effort.
- ◯ Speech is generally clear, with some fluidity of expression, but the speaker exhibits minor difficulties with pronunciation and may require some listener effort at times.
- ◯ Speech is clear at times, though the speaker exhibits problems with pronunciation and so may require more listener effort. It was necessary to listen more than once before attempting to complete the gap fill.
- ◯ Consistent pronunciation difficulties cause considerable listener effort throughout the sample. It was necessary to listen several times before attempting to complete the gap fill.
- ◯ The speaker's pronunciation is unintelligible.

Comments to the researcher

17. **Speaker 3 - Fill in the blanks. Please click on the following link and listen to the recording, and fill in the blanks with the missing words. Each blank has only one missing word.**

The ___(a)___ _____ (b) _____ is ___(c)___ _____(d)___.

The ___(e)__ ___(f)___ the ____(g)_____.

a.

b.

c.

d.

e.

f.

g.

* 18. Speaker 3 - Pronunciation Score

Please now score Speaker 3 for pronunciation using the scale provided.

- ◯ Speech is generally clear and requires little or no listener effort.
- ◯ Speech is generally clear, with some fluidity of expression, but the speaker exhibits minor difficulties with pronunciation and may require some listener effort at times.
- ◯ Speech is clear at times, though the speaker exhibits problems with pronunciation and so may require more listener effort. It was necessary to listen more than once before attempting to complete the gap fill.
- ◯ Consistent pronunciation difficulties cause considerable listener effort throughout the sample. It was necessary to listen several times before attempting to complete the gap fill.
- ◯ The speaker's pronunciation is unintelligible.

Comments to the researcher

19. **Speaker 4 - Fill in the blanks. Please click on the following link and listen to the recording, and fill in the blanks with the missing words. Each blank has only one missing word.**

The ___(a)___ ___(b)___ ____(c)___.

The ___(d)__ ___(e)___ in a ____(f)_____.

a.

b.

c.

d.

e.

f.

* 20. Speaker 4 - Pronunciation Score

Please now score Speaker 4 for pronunciation using the scale provided.

- ◯ Speech is generally clear and requires little or no listener effort.
- ◯ Speech is generally clear, with some fluidity of expression, but the speaker exhibits minor difficulties with pronunciation and may require some listener effort at times.
- ◯ Speech is clear at times, though the speaker exhibits problems with pronunciation and so may require more listener effort. It was necessary to listen more than once before attempting to complete the gap fill.
- ◯ Consistent pronunciation difficulties cause considerable listener effort throughout the sample. It was necessary to listen several times before attempting to complete the gap fill.
- ◯ The speaker's pronunciation is unintelligible.

Comments to the researcher

21. **Speaker 5 - Fill in the blanks. Please click on the following link and listen to the recording, and fill in the blanks with the missing words. Each blank has only one missing word.**

The ___(a)___ ___(b)___ on the ____(c)___.

The ___(d)__ ___(e)___ ____(f)_____.

a.

b.

c.

d.

e.

f.

* 22. Speaker 5 - Pronunciation Score

Please now score Speaker 5 for pronunciation using the scale provided.

- ◯ Speech is generally clear and requires little or no listener effort.
- ◯ Speech is generally clear, with some fluidity of expression, but the speaker exhibits minor difficulties with pronunciation and may require some listener effort at times.
- ◯ Speech is clear at times, though the speaker exhibits problems with pronunciation and so may require more listener effort. It was necessary to listen more than once before attempting to complete the gap fill.
- ◯ Consistent pronunciation difficulties cause considerable listener effort throughout the sample. It was necessary to listen several times before attempting to complete the gap fill.
- ◯ The speaker's pronunciation is unintelligible.

Comments to the researcher

23. **Speaker 6 - Fill in the blanks. Please click on the following link and listen to the recording, and fill in the blanks with the missing words. Each blank has only one missing word.**

He ___(a)___ to ___(b)___ his ____(c)___.

The ___(d)__ was ___(e)___ ____(f)_____.

a.

b.

c.

d.

e.

f.

* 24. Speaker 6 - Pronunciation Score

Please now score Speaker 6 for pronunciation using the scale provided.

- ◯ Speech is generally clear and requires little or no listener effort.
- ◯ Speech is generally clear, with some fluidity of expression, but the speaker exhibits minor difficulties with pronunciation and may require some listener effort at times.
- ◯ Speech is clear at times, though the speaker exhibits problems with pronunciation and so may require more listener effort. It was necessary to listen more than once before attempting to complete the gap fill.
- ◯ Consistent pronunciation difficulties cause considerable listener effort throughout the sample. It was necessary to listen several times before attempting to complete the gap fill.
- ◯ The speaker's pronunciation is unintelligible.

Comments to the researcher

25. **Speaker 7 - Fill in the blanks. Please click on the following link and listen to the recording, and fill in the blanks with the missing words. Each blank has only one missing word.**

The ___(a)___ was ___(b)___ ____(c)___.

___(d)___ ____(e)_____ for ____(f)_____.

a.

b.

c.

d.

e.

f.

* 26. Speaker 7 - Pronunciation Score

Please now score Speaker 7 for pronunciation using the scale provided.

- ◯ Speech is generally clear and requires little or no listener effort.
- ◯ Speech is generally clear, with some fluidity of expression, but the speaker exhibits minor difficulties with pronunciation and may require some listener effort at times.
- ◯ Speech is clear at times, though the speaker exhibits problems with pronunciation and so may require more listener effort. It was necessary to listen more than once before attempting to complete the gap fill.
- ◯ Consistent pronunciation difficulties cause considerable listener effort throughout the sample. It was necessary to listen several times before attempting to complete the gap fill.
- ◯ The speaker's pronunciation is unintelligible.

Comments to the researcher

27. **Speaker 8 - Fill in the blanks. Please click on the following link and listen to the recording, and fill in the blanks with the missing words. Each blank has only one missing word.**

The old ___(a)___ ___(b)___ the ____(c)___.

The ___(d)___ ____(e)_____ is ____(f)_____.

a.

b.

c.

d.

e.

f.

* 28. Speaker 8 - Pronunciation Score

Please now score Speaker 8 for pronunciation using the scale provided.

- ◯ Speech is generally clear and requires little or no listener effort.
- ◯ Speech is generally clear, with some fluidity of expression, but the speaker exhibits minor difficulties with pronunciation and may require some listener effort at times.
- ◯ Speech is clear at times, though the speaker exhibits problems with pronunciation and so may require more listener effort. It was necessary to listen more than once before attempting to complete the gap fill.
- ◯ Consistent pronunciation difficulties cause considerable listener effort throughout the sample. It was necessary to listen several times before attempting to complete the gap fill.
- ◯ The speaker's pronunciation is unintelligible.

Comments to the researcher

29. **Speaker 9 - Fill in the blanks. Please click on the following link and listen to the recording, and fill in the blanks with the missing words. Each blank has only one missing word.**

The ___(a)___ ___(b)___ some ____(c)___.

The ___(d)___ is ____(e)_____ ____(f)_____.

a.

b.

c.

d.

e.

f.

* 30. Speaker 9 - Pronunciation Score

Please now score Speaker 9 for pronunciation using the scale provided.

- ◯ Speech is generally clear and requires little or no listener effort.
- ◯ Speech is generally clear, with some fluidity of expression, but the speaker exhibits minor difficulties with pronunciation and may require some listener effort at times.
- ◯ Speech is clear at times, though the speaker exhibits problems with pronunciation and so may require more listener effort. It was necessary to listen more than once before attempting to complete the gap fill.
- ◯ Consistent pronunciation difficulties cause considerable listener effort throughout the sample. It was necessary to listen several times before attempting to complete the gap fill.
- ◯ The speaker's pronunciation is unintelligible.

Comments to the researcher

31. **Speaker 10 - Fill in the blanks. Please click on the following link and listen to the recording, and fill in the blanks with the missing words. Each blank has only one missing word.**

The ___(a)___ is ___(b)___ ____(c)___.

The ___(d)___ ____(e)_____ the ____(f)_____.

a.

b.

c.

d.

e.

f.

* 32. Speaker 10 - Pronunciation Score

Please now score Speaker 10 for pronunciation using the scale provided.

- ◯ Speech is generally clear and requires little or no listener effort.
- ◯ Speech is generally clear, with some fluidity of expression, but the speaker exhibits minor difficulties with pronunciation and may require some listener effort at times.
- ◯ Speech is clear at times, though the speaker exhibits problems with pronunciation and so may require more listener effort. It was necessary to listen more than once before attempting to complete the gap fill.
- ◯ Consistent pronunciation difficulties cause considerable listener effort throughout the sample. It was necessary to listen several times before attempting to complete the gap fill.
- ◯ The speaker's pronunciation is unintelligible.

Comments to the researcher

33. **Speaker 11 - Fill in the blanks. Please click on the following link and listen to the recording, and fill in the blanks with the missing words. Each blank has only one missing word.**

The ___(a)___ ___(b)___ into the ____(c)___.

The ___(d)___ ____(e)_____ a ____(f)_____.

a.

b.

c.

d.

e.

f.

* 34. Speaker 11 - Pronunciation Score

Please now score Speaker 11 for pronunciation using the scale provided.

- ◯ Speech is generally clear and requires little or no listener effort.
- ◯ Speech is generally clear, with some fluidity of expression, but the speaker exhibits minor difficulties with pronunciation and may require some listener effort at times.
- ◯ Speech is clear at times, though the speaker exhibits problems with pronunciation and so may require more listener effort. It was necessary to listen more than once before attempting to complete the gap fill.
- ◯ Consistent pronunciation difficulties cause considerable listener effort throughout the sample. It was necessary to listen several times before attempting to complete the gap fill.
- ◯ The speaker's pronunciation is unintelligible.

Comments to the researcher

35. **Speaker 12 - Fill in the blanks. Please click on the following link and listen to the recording, and fill in the blanks with the missing words. Each blank has only one missing word.**

___(a)___ ___(b)___ her ____(c)___.

The ___(d)___ ____(e)_____ a ____(f)_____.

a.

b.

c.

d.

e.

f.

* 36. Speaker 12 - Pronunciation Score

Please now score Speaker 12 for pronunciation using the scale provided.

- ◯ Speech is generally clear and requires little or no listener effort.
- ◯ Speech is generally clear, with some fluidity of expression, but the speaker exhibits minor difficulties with pronunciation and may require some listener effort at times.
- ◯ Speech is clear at times, though the speaker exhibits problems with pronunciation and so may require more listener effort. It was necessary to listen more than once before attempting to complete the gap fill.
- ◯ Consistent pronunciation difficulties cause considerable listener effort throughout the sample. It was necessary to listen several times before attempting to complete the gap fill.
- ◯ The speaker's pronunciation is unintelligible.

Comments to the researcher

37. **Speaker 13 - Fill in the blanks. Please click on the following link and listen to the recording, and fill in the blanks with the missing words. Each blank has only one missing word.**

___(a)___ ___(b)___ at the ____(c)___.

The ___(d)___ ____(e)_____ the ____(f)_____.

a.

b.

c.

d.

e.

f.

* 38. Speaker 13 - Pronunciation Score

Please now score Speaker 13 for pronunciation using the scale provided.

- Speech is generally clear and requires little or no listener effort.
- Speech is generally clear, with some fluidity of expression, but the speaker exhibits minor difficulties with pronunciation and may require some listener effort at times.
- Speech is clear at times, though the speaker exhibits problems with pronunciation and so may require more listener effort. It was necessary to listen more than once before attempting to complete the gap fill.
- Consistent pronunciation difficulties cause considerable listener effort throughout the sample. It was necessary to listen several times before attempting to complete the gap fill.
- The speaker's pronunciation is unintelligible.

Comments to the researcher

Main Study

Final Page

Thank you for your time and efforts participating in this research, it is greatly appreciated.

Sincerely,

Kevin C. Browne

39. Please share any additional comments you may have concerning the instrument or your experience participating in this research.

40. If you are willing to contribute further to this research by possibly answering some questions or completing another survey please leave your email address. (It will not be shared or used for any other purposes than the current research)

IPA transcriptions of all sentences designed for the test in RP and General American pronunciations

Sentences for Spanish-English	General American (GA)	Received Pronunciaiton (RP)
The jug stood on the shelf.	ðə ˈdʒəg ˈstʊd ɑːn ðə ˈʃelf	ðə dʒʌg stʊd ɒn ðə ʃelf
The shoes were very yellow.	ðə ˈʃuːz wər ˈveri ˈjeloʊ	ðə ʃuːz wə ˈveri ˈjeləʊ
The machine was awkwardly noisy.	ðə mɪˈʃiːn wəz ˈɑːkwərdli ˈnɔɪzi	ðə məˈʃiːn wəz ˈɔːkwədli ˈnɔɪzi
The change demanded some strength.	ðə ˈtʃeɪndʒ ˌdɪˈmændəd səm ˈstreŋkθ	ðə tʃeɪndʒ dɪˈmɑːndɪd səm streŋθ
They're staying for improvement.	ˈðer ˈsteɪɪŋ fər ˌɪmˈpruːvmənt	ˈðeə ˈsteɪɪŋ fər ɪmˈpruːvmənt
The agency is transferring along.	ði ˈeɪdʒənsi z trænsˈfɜːrɪŋ əˈlɔːŋ	ði ˈeɪdʒənsi z trænsˈfɜːrɪŋ əˈlɒŋ
She threw her toy.	ʃi ˈθruː hər ˌtɔɪ	ʃi θruː hə tɔɪ
The soldiers swept into the attack.	ðə ˈsoʊldʒəz ˈswept ˌɪnˈtuː ði əˈtæk	ðə ˈsəʊldʒəz swept ˈɪntə ði əˈtæk
The washing function broke.	ðə ˈwɑːʃɪŋ ˈfʌŋkʃn̩ broʊk	ðə ˈwɒʃɪŋ ˈfʌŋkʃn̩ brəʊk
The patient suffered a seizure.	ðə ˈpeɪʃənt ˈsʌfərd ə ˈsiːʒər	ðə ˈpeɪʃnt ˈsʌfəd ə ˈsiːʒə

Sentences for Arabic-English	General American (GA)	Received Pronunciaiton (RP)
The tea cloth is quite wet.	ðə ˈtiː ˈklɒθ s ˈkwaɪt ˈwet	ðə ˈtiː ˈklɒθ s ˈkwaɪt ˈwet
They're shopping for cheese.	ˈðer ˈʃɑːpɪŋ fər ˈtʃiːz	ˈðer ˈʃɑːpɪŋ fər ˈtʃiːz
The jury learned the law.	ðə ˈdʒʊri ˈlɝːnd ðə ˈlɑː	ðə ˈdʒʊri ˈlɝːnd ðə ˈlɑː
Playing involved her hands.	ˈpleɪɪŋ ˌɪnˈvɑːlvd hər ˈhændz	ˈpleɪɪŋ ˌɪnˈvɑːlvd hər ˈhændz
The blend is rather different.	ðə ˈblend z ˈræðər ˈdɪfərənt	ðə ˈblend z ˈræðər ˈdɪfərənt
He paid to change his server.	hi ˈpeɪd tə ˈtʃeɪndʒ ɪz ˈsɝːvər	hi ˈpeɪd tə ˈtʃeɪndʒ ɪz ˈsɝːvər
The group dreamt of treasure.	ðə ˈgruːp ˈdremt əv ˈtreʒər	ðə ˈgruːp ˈdremt əv ˈtreʒər
The boy wore a patch.	ðə ˌbɔɪ ˈwɔːr ə ˈpætʃ	ðə ˌbɔɪ ˈwɔːr ə ˈpætʃ
The credit was quite usual.	ðə ˈkredət wəz ˈkwaɪt ˈjuːʒəwəl	ðə ˈkredət wəz ˈkwaɪt ˈjuːʒəwəl
The children feared the bridge.	ðə ˈtʃɪldrən ˈfɪrd ðə ˈbrɪdʒ	ðə ˈtʃɪldrən ˈfɪrd ðə ˈbrɪdʒ

Sentences for Dhivehi-English	General American (GA)	Received Pronunciaiton (RP)
The judge laughed at his mother.	ðə ˈdʒədʒ ˈlæft ət ɪz ˈmʌðr̩	ðə dʒʌdʒ lɑːft ət ɪz ˈmʌðə
Father looked at the chief.	ˈfɑːðr̩ ˈlʊkt ət ðə ˈtʃiːf	ˈfɑːðə lʊkt ət ðə tʃiːf
The farmer managed the bulls.	ðə ˈfɑːrmər ˈmænədʒd ðə ˈbʊlz	ðə ˈfɑːmə ˈmænɪdʒd ðə bʊlz
Some structures were under the tree.	səm ˈstrʌktʃərz wər ˈʌndr̩ ðə ˈtriː	səm ˈstrʌktʃəz wər ˈʌndə ðə triː
The creature travelled quietly.	ðə ˈkriːtʃr̩ ˈtrævl̩d ˈkwaɪətli	ðə ˈkriːtʃə ˈtrævl̩d ˈkwaɪətli
The cat parted with his stick.	ðə kæt ˈpɑːrtəd wɪθ ɪz ˈstɪk	ðə kæt ˈpɑːtɪd wɪð ɪz stɪk
They're minding society.	ˈðer ˈmaɪndɪŋ səˈsaɪəti	ˈðeə ˈmaɪndɪŋ səˈsaɪəti
The passenger stood in a bath.	ðə ˈpæsəndʒər ˈstʊd ɪn ə ˈbæθ	ðə ˈpæsɪndʒə stʊd ɪn ə bɑːθ
The old hag challenged the theater.	ði oʊld ˈhæg ˈtʃælədʒd ðə ˈθiːətər	ði əʊld hæg ˈtʃæləndʒd ðə ˈθiːətə
The huge girl is shouting.	ðə ˈhjuːdʒ ˈgɝːl z ˈʃaʊtɪŋ	ðə hjuːdʒ gɜːl z ˈʃaʊtɪŋ

The Facets analyses output of the Spanish-English pronunciation score data

```
Facets (Many-Facet Rasch Measurement) Version No. 3.71.1  Copyright �(c) 1987-2013, John M. Linacre. All rights reserved.
22/12/2014 16:52:26

Rasch Analysis of Expert Ratings 22/12/2014 16:52:26
Table 1. Specifications from file "C:\Users\kevin\Desktop\Main Study Data\Spanish English Pronunciaton Operating File.txt.txt".

Title = Rasch Analysis of Expert Ratings 22/12/2014 16:52:26
Data file = "C:\Users\kevin\Desktop\Main Study Data\Pronunciation Score Data Spanish Familiarity.xlsx"
Output file = C:\Users\kevin\Desktop\Main Study Data\Spanish English Pronunciaton Operating File.txt.out.txt

; Data specification
Facets = 3
Delements = N
Non-centered = 1
Positive =  1, 2, 3
Labels =
 1,rater ; (elements = 190)
 2,Spanish English Familiarity Level ; (elements = 4)
 3,Speaker ; (elements = 12)
Model = ?,?,?,R5,1

; Output description
Arrange tables in order = MN
Bias/Interaction direction = plus ; ability, easiness, leniency: higher score = positive logit
Fair score = Mean
Pt-biserial = Measure
Heading lines in output data files = Y
Inter-rater coefficients reported for facet = 1
Barchart = Yes
Total score for elements = Yes
T3onscreen show only one line on screen iteration report = Y
T4MAX maximum number of unexpected observations reported in Table 4 = 100
T8NBC show table 8 numbers-barcharts-curves = NBC
Unexpected observations reported if standardized residual >= 3
Usort unexpected observations sort order = u
WHexact - Wilson-Hilferty standardization = Y

; Convergence control
Convergence = .1, .01
Iterations (maximum) = 0 ; unlimited
Xtreme scores adjusted by = .3, .5  ;(estimation, bias)

Rasch Analysis of Expert Ratings 22/12/2014 16:52:26
Table 2. Data Summary Report.

Assigning models to Data= "C:\Users\kevin\Desktop\Main Study Data\Pronunciation Score Data Spanish Familiarity.xlsx"
Total lines in data file = 194
Total data lines = 190
Responses matched to model: ?,?,?,R5,1 = 2280
    Total non-blank responses found = 2280
              Number of blank lines = 3
Valid responses used for estimation = 2280

Rasch Analysis of Expert Ratings 22/12/2014 16:52:26
Table 3. Iteration Report.
```

Iteration	Max. Score Residual Elements	%	Categories	Max. Logit Change Elements	Steps
PROX 1				-1.6094	
PROX 2				-.3098	
JMLE 3	-498.9658	29.2	532.8774	.8917	-2.1512
JMLE 4	-50.9991	-13.0	-40.0207	-.3720	.1802
JMLE 5	-26.3881	-5.4	34.8404	-.1744	-.1497
JMLE 6	-23.0878	-4.3	21.0636	-.1414	-.0858
JMLE 7	-18.8652	-3.4	17.2404	-.1106	-.0743
JMLE 8	-15.7680	-2.8	14.1856	-.0894	-.0614
JMLE 9	-13.2930	-2.3	11.8519	-.0734	-.0518
JMLE 10	-11.3046	-1.9	10.0148	-.0610	-.0441
JMLE 11	-9.6803	-1.6	8.5408	-.0513	-.0378
JMLE 12	-8.3366	-1.4	7.3376	-.0436	-.0326
JMLE 13	-7.2133	-1.2	6.3411	-.0372	-.0283
JMLE 14	-6.2657	-1.0	5.5059	-.0320	-.0246
JMLE 15	-5.4606	-.9	4.7991	-.0277	-.0214
JMLE 16	-4.7722	-.8	-4.1966	-.0240	-.0187
JMLE 17	-4.1805	-.7	-3.6834	-.0210	-.0164
JMLE 18	-3.6696	-.6	-3.2395	-.0183	-.0144
JMLE 19	-3.2269	-.5	-2.8538	-.0160	-.0127
JMLE 20	-2.8419	-.4	-2.5176	-.0141	-.0112
JMLE 21	-2.5061	-.4	-2.2237	-.0124	-.0099
JMLE 22	-2.2126	-.3	-1.9661	-.0109	-.0087
JMLE 23	-1.9554	-.3	-1.7398	-.0096	-.0077
JMLE 24	-1.7296	-.3	-1.5408	-.0085	-.0068
JMLE 25	-1.5311	-.2	-1.3654	-.0075	-.0060
JMLE 26	-1.3563	-.2	-1.2107	-.0067	-.0054
JMLE 27	-1.2021	-.2	-1.0741	-.0059	-.0047
JMLE 28	-1.0661	-.2	-.9533	-.0052	-.0042
JMLE 29	-.9459	-.1	-.8464	-.0046	-.0037
JMLE 30	-.8396	-.1	-.7517	-.0041	-.0033
JMLE 31	-.7455	-.1	-.6679	-.0036	-.0029
JMLE 32	-.6621	-.1	-.5935	-.0032	-.0026
JMLE 33	-.5883	-.1	-.5276	-.0029	-.0023
JMLE 34	-.5228	-.1	-.4690	-.0026	-.0021
JMLE 35	-.4647	-.1	-.4171	-.0023	-.0018
JMLE 36	-.4132	-.1	-.3709	-.0020	-.0016
JMLE 37	-.3674	-.1	-.3300	-.0018	-.0015
JMLE 38	-.3268	-.1	-.2935	-.0016	-.0013
JMLE 39	-.2907	.0	-.2611	-.0014	-.0011
JMLE 40	-.2586	.0	-.2324	-.0013	-.0010
JMLE 41	-.2301	.0	-.2068	-.0011	-.0009
JMLE 42	-.2047	.0	-.1840	-.0010	-.0008
JMLE 43	-.1822	.0	-.1638	-.0009	-.0007
JMLE 44	-.1621	.0	-.1458	-.0008	-.0006
JMLE 45	-.1443	.0	-.1298	-.0007	-.0006
JMLE 46	-.1284	.0	-.1154	-.0006	-.0005
JMLE 47	-.1143	.0	-.1028	-.0006	-.0005
JMLE 48	-.1017	.0	-.0915	-.0005	-.0004
JMLE 49	-.0906	.0	-.0815	-.0004	-.0004

```
Warning (6)! There may be 4 disjoint subsets


Rasch Analysis of Expert Ratings 22/12/2014 16:52:26
Table 4. Unexpected Responses - appears after Table 8.


Rasch Analysis of Expert Ratings 22/12/2014 16:52:26
Table 5. Measurable Data Summary.
```

Cat	Score	Exp.	Resd	StRes	
2.83	2.83	2.83	.00	.00	Mean (Count: 2280)
.97	.97	.63	.74	1.00	S.D. (Population)
.97	.97	.63	.74	1.00	S.D. (Sample)

```
+--------------------------------------------------+
Data log-likelihood chi-square = 5051.5649
Approximate degrees of freedom = 2073
Chi-square significance prob.  = .0000
                                          Count   Mean   S.D.   Params
Responses used for estimation      =       2280   2.83   0.97      207
Count of measurable responses      =       2280
Raw-score variance of observations   =   0.94 100.00%
Variance explained by Rasch measures =   0.39  42.00%
Variance of residuals                =   0.55  58.00%

Rasch Analysis of Expert Ratings 22/12/2014 16:52:26
Table 6.0  All Facet Vertical "Rulers".

Vertical = (1*,2A,3A,S) Yardstick (columns lines low high extreme)= 0,8,-3,3,End
+---------------------------------------------------------------------------------------------------------------------------------------+
| (1)|  (2)    |                (3)                      |                                    (4)                                      | (5) |
|Measr|+Rater   |+Spanish English Familiarity Level       |+Speaker Participants                                                        |Rating|
|Logit|         |                                         |                                                                             |Scale|
|-----+---------+-----------------------------------------+-----------------------------------------------------------------------------+-----|
|     |(lenient)|   (lenient)                             |   (high scoring)                                                            |     |
|   3 +         +                                         +                                                                             + (5) |
|     |         |                                         |                                                                             | --- |
|     |         |                                         |                                                                             |     |
|     |         |                                         |                                                                             |     |
|     |         |                                         |                                                                             |     |
|     |         |                                         |                                                                             |     |
|     |         |                                         |                                                                             |     |
|     |         |                                         |                                                                             |     |
|   2 + .       +                                         +                                                                             +     |
|     | .       |                                         |                                                                             |     |
|     |         |                                         |                                                                             |  4  |
|     | .       |                                         |                                                                             |     |
|     | .       |                                         |                                                                             |     |
|     | .       |                                         |                                                                             |     |
|     | *.      |                                         |                                                                             |     |
|     | .       |                                         | Speaker 10 Arabic-English 2         Speaker 11 Spanish-English Female 2     |     |
|   1 + .       +                                         +                                                                             +     |
|     | .       |                                         |                                                                             | --- |
|     | **.     |                                         | Speaker 13 Dhivehi-English Male 2                                           |     |
|     | *.      |                                         | Speaker 12 Arabic-English Female 2                                          |     |
|     | **.     |                                         | Speaker 6 Arabic-English Male 1                                             |     |
|     | **.     |                                         |                                                                             |     |
|     | ****    |                                         |                                                                             |     |
|     | **.     | Some Familiarity     Very Familiar      |                                                                             |     |
*   0 * ***     * Limited Familiarity                     * Speaker 8 Dhivehi-English Male 1                                            *  3  *
|     | ****.   |                                         |                                                                             |     |
|     | ******* | No Familiarity                          | Speaker 4 Dhivehi-English Female 2                                          |     |
|     | ****.   |                                         |                                                                             |     |
|     | **.     |                                         | Speaker 2 Dhivehi-English Female 1   Speaker 7 Spanish-English Male 2       |     |
|     | *.      |                                         | Speaker 9 Spanish-English Female 1                                          |     |
|     | *       |                                         |                                                                             | --- |
|     | ***.    |                                         |                                                                             |     |
|  -1 + ****.   +                                         + Speaker 3 Arabic-English Female 1                                           +     |
|     | ***     |                                         | Speaker 5 Spanish-English Male 1                                            |     |
|     | *.      |                                         |                                                                             |     |
|     | .       |                                         |                                                                             |     |
|     | *       |                                         |                                                                             |     |
|     | **      |                                         |                                                                             |     |
|     |         |                                         |                                                                             |  2  |
|     | .       |                                         |                                                                             |     |
|  -2 + *       +                                         +                                                                             +     |
|     | .       |                                         |                                                                             |     |
|     |         |                                         |                                                                             |     |
|     | *       |                                         |                                                                             |     |
|     |         |                                         |                                                                             |     |
|     | .       |                                         |                                                                             |     |
|     |         |                                         |                                                                             |     |
|     |         |                                         |                                                                             |     |
|  -3 +         +                                         +                                                                             + (1) |
|     |(severe) |    (severe)                             |  (low scoring)                                                              |     |
|-----+---------+-----------------------------------------+-----------------------------------------------------------------------------+-----|
|Measr| * = 3   |+Spanish English Familiarity Level       |+Speaker                                                                     |Scale|
+---------------------------------------------------------------------------------------------------------------------------------------+

S.1: Model = ?,?,?,R5
There are 4 disconnected subsets identified in Table 7.

Rasch Analysis of Expert Ratings 22/12/2014 16:52:26
Table 6.1  rater Facet Summary.

Logit:
                    1     1 11  1
     1 12 13 115224554658 759249552664 71 25 111  11
 +---------Q-------S-+------M--+----S----+--Q------+---------+
-3        -2        -1         0         1         2         3

Infit MnSq:
     112111111 1
   4467524553273431 321 1221   1                         1  1
 +--S------M------S--+---Q-----+---------+---------+---------+
 0         1         2         3         4         5         6

Outfit MnSq:
     122111111 1
   4430617090292432 221 2211   1                         1 1
 +--S------M------S--+---Q-----+---------+---------+---------+
 0         1         2         3         4         5         6

Infit ZStd:
                  11  1 1111 1
         121  2258885479066011627311231 32  1            2
 +----+----Q----+-S--+---M+----+S---+---Q+----+----+----+----+
-5   -4   -3   -2   -1    0    1    2    3    4    5    6    7

Outfit ZStd:
                1 11    11 1 1
         121 1133718638895827182533 23 231  1           11
 +----+----Q----+-S--+---M+----+S---+--Q-+----+----+----+----+
-5   -4   -3   -2   -1    0    1    2    3    4    5    6    7


Rasch Analysis of Expert Ratings 22/12/2014 16:52:26
Table 6.2  Spanish English Familiarity Level Facet Summary.

Logit:
                           1  12
 +---------+---------+------QS-M-SQ------+---------+---------+
-3        -2        -1         0         1         2         3

Infit MnSq:
         1 11 1
 +------Q-S+MS-Q-----+---------+---------+---------+---------+
 0         1         2         3         4         5         6

Outfit MnSq:
         1 11 1
 +------Q-S+MS-Q-----+---------+---------+---------+---------+
 0         1         2         3         4         5         6
```

```
Infit ZStd:
    1                     1            1  1
 +----+----+-S--+----+----+-M--+----+----+-S--+----+----+-Q--+
-5   -4   -3   -2   -1    0    1    2    3    4    5    6    7

Outfit ZStd:
    1                     1           1   1
 +----+----+-S--+----+----+-M--+----+----+S---+----+----+Q---+
-5   -4   -3   -2   -1    0    1    2    3    4    5    6    7
```

Rasch Analysis of Expert Ratings 22/12/2014 16:52:26
Table 6.3 Speaker Facet Summary.

```
Logit:
                    11   1111 1     11 1  2
 +---------+----Q----+--S------M------S--+----Q----+---------+
-3        -2        -1         0         1         2         3

Infit MnSq:
        1512  11 1
 +----Q-S--M--S-Q----+---------+---------+---------+---------+
 0         1         2         3         4         5         6

Outfit MnSq:
        1512  11 1
 +----Q-S--M--S-Q----+---------+---------+---------+---------+
 0         1         2         3         4         5         6

Infit ZStd:
            1 2  21   1   11            1 1       1
 Q----+----+-S--+----+---M+----+----+S---+----+--Q-+----+----+
-5   -4   -3   -2   -1    0    1    2    3    4    5    6    7

Outfit ZStd:
            11 1 21   1   11            1  1      1
 +----+----+-S--+----+---M+----+----+S---+----+--Q-+----+----+
-5   -4   -3   -2   -1    0    1    2    3    4    5    6    7
```

Rasch Analysis of Expert Ratings 22/12/2014 16:52:26
Table 7.1.1 rater Measurement Report (arranged by MN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	Exact Agree. Obs %	Exact Agree. Exp %	Num rater	
20	12	1.67	1.63	-2.64	.47	.37	-1.9	.43	-1.7	1.70	.63	.46	22.2	21.7	138 138	in subset: 1
20	12	1.67	1.74	-2.38	.47	.57	-1.1	.59	-1.0	1.51	.71	.46	25.4	26.9	24 24	in subset: 4
22	12	1.83	1.75	-2.35	.44	.31	-2.3	.35	-2.1	1.79	.49	.48	21.1	22.0	145 145	in subset: 2
22	12	1.83	1.75	-2.35	.44	1.10	.3	1.18	.5	.80	.23	.48	23.3	22.0	183 183	in subset: 2
23	12	1.92	1.84	-2.15	.43	2.25	2.4	2.30	2.6	-.69	-.60	.48	25.1	24.0	75 75	in subset: 3
24	12	2.00	1.92	-1.96	.43	.44	-1.6	.45	-1.6	1.65	.63	.49	27.3	25.6	65 65	in subset: 3
22	12	1.83	1.92	-1.96	.44	2.08	2.2	2.10	2.2	-.34	.08	.48	18.0	29.8	72 72	in subset: 4
22	12	1.83	1.92	-1.96	.44	1.00	.1	1.03	.1	.97	.35	.48	28.5	29.8	162 162	in subset: 4
24	12	2.00	1.97	-1.84	.43	.80	-.4	.77	-.5	1.21	.52	.49	29.0	28.4	147 147	in subset: 1
25	12	2.08	2.06	-1.67	.42	.87	-.2	.94	.0	1.11	.53	.50	27.5	29.8	111 111	in subset: 1
26	12	2.17	2.08	-1.62	.41	1.56	1.3	1.65	1.5	.21	-.01	.50	27.7	27.9	5 5	in subset: 2
26	12	2.17	2.08	-1.61	.41	.93	.0	.94	.0	1.03	.30	.50	29.1	28.6	53 53	in subset: 3
26	12	2.17	2.08	-1.61	.41	1.04	.2	1.03	.1	.97	.64	.50	31.0	28.6	151 151	in subset: 3
26	12	2.17	2.08	-1.61	.41	.47	-1.5	.47	-1.5	1.61	.16	.50	31.2	28.6	180 180	in subset: 3
24	12	2.00	2.10	-1.59	.43	1.08	.3	1.04	.2	.97	.70	.49	30.7	31.9	40 40	in subset: 4
26	12	2.17	2.14	-1.50	.41	1.19	.5	1.20	.6	.71	.33	.50	33.0	31.0	69 69	in subset: 1
27	12	2.25	2.16	-1.45	.41	1.07	.3	1.13	.4	.91	.45	.51	27.0	29.1	165 165	in subset: 2
27	12	2.25	2.16	-1.45	.41	.18	-3.1	.17	-3.2	1.94	.72	.51	36.5	29.8	33 33	in subset: 3
25	12	2.08	2.18	-1.41	.42	1.21	.6	1.21	.6	.67	-.09	.50	32.5	32.7	85 85	in subset: 4
27	12	2.25	2.22	-1.33	.41	1.39	1.0	1.35	.9	.60	.75	.51	29.6	32.0	167 167	in subset: 1
28	12	2.33	2.25	-1.29	.40	1.13	.4	1.13	.4	.82	.42	.51	31.9	31.0	13 13	in subset: 3
28	12	2.33	2.25	-1.29	.40	.50	-1.4	.49	-1.4	1.52	.57	.51	38.2	31.0	14 14	in subset: 3
28	12	2.33	2.25	-1.29	.40	.42	-1.7	.45	-1.6	1.63	.40	.51	37.1	31.0	125 125	in subset: 3
26	12	2.17	2.27	-1.24	.41	1.25	.7	1.24	.6	.72	.67	.50	30.7	33.4	108 108	in subset: 4
28	12	2.33	2.31	-1.17	.40	.63	-.9	.63	-.9	1.40	.46	.51	36.6	32.9	81 81	in subset: 1
28	12	2.33	2.31	-1.17	.40	.66	-.8	.64	-.9	1.41	.62	.51	34.1	32.9	99 99	in subset: 1
28	12	2.33	2.31	-1.17	.40	.96	.0	1.00	.1	.95	.35	.51	35.1	32.9	136 136	in subset: 1
28	12	2.33	2.31	-1.17	.40	.92	.0	.97	.0	.98	.60	.51	35.9	32.9	142 142	in subset: 1
29	12	2.42	2.33	-1.13	.40	.28	-2.4	.26	-2.6	1.85	.81	.52	36.5	31.1	39 39	in subset: 2
29	12	2.42	2.33	-1.13	.40	.73	-.6	.77	-.5	1.24	.35	.52	37.0	31.1	164 164	in subset: 2
29	12	2.42	2.33	-1.13	.40	.65	-.8	.68	-.7	1.34	.45	.52	37.3	32.0	43 43	in subset: 3
29	12	2.42	2.33	-1.13	.40	.86	-.2	.85	-.2	1.19	.80	.52	30.7	32.0	135 135	in subset: 3
27	12	2.25	2.36	-1.07	.41	.94	.0	.98	.0	1.01	.04	.51	29.8	33.8	105 105	in subset: 4
29	12	2.42	2.39	-1.01	.40	.66	-.8	.69	-.7	1.35	.80	.52	39.1	33.6	1 1	in subset: 1
29	12	2.42	2.39	-1.01	.40	1.72	1.6	1.72	1.6	.18	.40	.52	30.4	33.6	64 64	in subset: 1
29	12	2.42	2.39	-1.01	.40	.33	-2.2	.30	-2.3	1.80	.76	.52	40.0	33.6	88 88	in subset: 1
29	12	2.42	2.39	-1.01	.40	1.07	.3	1.14	.4	.88	.48	.52	31.3	33.6	93 93	in subset: 1
29	12	2.42	2.39	-1.01	.40	1.48	1.1	1.47	1.1	.34	-.08	.52	33.6	33.6	121 121	in subset: 1
29	12	2.42	2.39	-1.01	.40	.87	-.2	.85	-.2	1.14	.44	.52	32.6	33.6	184 184	in subset: 1
30	12	2.50	2.41	-.97	.39	1.35	.9	1.36	.9	.62	.76	.52	34.3	31.8	155 155	in subset: 2
30	12	2.50	2.41	-.97	.39	.18	-3.1	.19	-3.1	1.89	.78	.52	39.6	32.8	28 28	in subset: 3
30	12	2.50	2.41	-.97	.39	.54	-1.2	.56	-1.2	1.48	.30	.52	38.9	32.8	35 35	in subset: 3
30	12	2.50	2.41	-.97	.39	1.21	.6	1.27	.7	.69	.36	.52	31.9	32.8	44 44	in subset: 3
30	12	2.50	2.41	-.97	.39	.56	-1.2	.53	-1.3	1.52	.53	.52	38.6	32.8	68 68	in subset: 3
30	12	2.50	2.41	-.97	.39	.43	-1.7	.46	-1.5	1.59	.43	.52	39.5	32.8	83 83	in subset: 3
30	12	2.50	2.41	-.97	.39	.79	-.4	.78	-.4	1.27	.68	.52	36.1	32.8	94 94	in subset: 3
30	12	2.50	2.41	-.97	.39	1.61	1.4	1.64	1.4	.33	.46	.52	31.7	32.8	118 118	in subset: 3
28	12	2.33	2.45	-.91	.40	1.07	.3	1.02	.1	1.03	.49	.51	29.8	34.1	181 181	in subset: 4
30	12	2.50	2.48	-.85	.39	1.39	1.0	1.34	.9	.55	.42	.52	34.2	34.1	3 3	in subset: 1
30	12	2.50	2.48	-.85	.39	1.42	1.0	1.42	1.0	.51	-.02	.52	33.5	34.1	4 4	in subset: 1
30	12	2.50	2.48	-.85	.39	.19	-3.0	.20	-3.0	1.87	.76	.52	40.6	34.1	82 82	in subset: 1
30	12	2.50	2.48	-.85	.39	1.02	.1	.96	.0	1.02	.52	.52	37.2	34.1	168 168	in subset: 1
30	12	2.50	2.48	-.85	.39	.59	-1.0	.62	-.9	1.44	.69	.52	35.4	34.1	185 185	in subset: 1
31	12	2.58	2.49	-.82	.39	1.23	.6	1.24	.6	.82	.80	.52	29.2	32.4	56 56	in subset: 2
31	12	2.58	2.49	-.82	.39	.62	-.9	.63	-.9	1.38	.65	.52	35.8	32.4	146 146	in subset: 2
31	12	2.58	2.50	-.82	.39	.83	-.3	.82	-.3	1.21	.65	.52	35.1	33.5	15 15	in subset: 3
31	12	2.58	2.50	-.82	.39	.71	-.6	.71	-.7	1.29	.07	.52	34.6	33.5	26 26	in subset: 3
31	12	2.58	2.50	-.82	.39	.88	-.1	.95	.0	1.13	.18	.52	36.6	33.5	63 63	in subset: 3
29	12	2.42	2.54	-.75	.40	1.32	.8	1.25	.7	.76	.08	.52	30.7	34.1	45 45	in subset: 4
29	12	2.42	2.54	-.75	.40	.82	-.3	.83	-.3	1.13	.25	.52	28.1	34.1	76 76	in subset: 4
31	12	2.58	2.56	-.70	.39	1.16	.5	1.19	.5	.82	.70	.52	31.5	34.4	144 144	in subset: 1
32	12	2.67	2.58	-.67	.39	.54	-1.2	.55	-1.2	1.46	.25	.53	34.1	32.9	66 66	in subset: 2
32	12	2.67	2.58	-.67	.39	.44	-1.6	.43	-1.7	1.61	.61	.53	40.0	32.9	87 87	in subset: 2
32	12	2.67	2.58	-.67	.39	2.39	2.7	2.31	2.5	-.57	.08	.53	31.6	32.9	173 173	in subset: 2
32	12	2.67	2.58	-.67	.39	.55	-1.2	.55	-1.2	1.52	.67	.53	39.0	34.0	57 57	in subset: 3
32	12	2.67	2.58	-.67	.39	.75	-.5	.74	-.5	1.27	.29	.53	37.7	34.0	154 154	in subset: 3
32	12	2.67	2.65	-.55	.39	.61	-1.0	.59	-1.1	1.44	.44	.53	37.1	34.6	2 2	in subset: 1
32	12	2.67	2.65	-.55	.39	.95	.0	.92	.0	1.11	.66	.53	38.2	34.6	52 52	in subset: 1
32	12	2.67	2.65	-.55	.39	1.10	.3	1.09	.3	.91	.84	.53	33.2	34.6	140 140	in subset: 1
33	12	2.75	2.66	-.52	.38	1.59	1.4	1.58	1.3	.36	.18	.53	28.7	33.2	36 36	in subset: 2
33	12	2.75	2.67	-.52	.38	1.34	.9	1.37	.9	.65	.51	.53	37.4	34.3	91 91	in subset: 3
33	12	2.75	2.67	-.52	.38	.95	.0	.91	-.1	1.16	.64	.53	40.0	34.3	127 127	in subset: 3
33	12	2.75	2.67	-.52	.38	.71	-.6	.70	-.7	1.30	.27	.53	39.1	34.3	176 176	in subset: 3
31	12	2.58	2.71	-.44	.39	1.47	1.1	1.45	1.1	.45	.36	.52	29.4	33.8	112 112	in subset: 4
33	12	2.75	2.73	-.40	.38	1.08	.3	1.08	.3	.91	.54	.53	36.2	34.6	37 37	in subset: 1
33	12	2.75	2.73	-.40	.38	.93	.0	.94	.0	1.09	.29	.53	37.9	34.6	48 48	in subset: 1
33	12	2.75	2.73	-.40	.38	.97	.0	.96	.0	1.03	.76	.53	35.4	34.6	126 126	in subset: 1
33	12	2.75	2.73	-.40	.38	1.37	.9	1.36	.9	.55	.14	.53	32.6	34.6	182 182	in subset: 1
34	12	2.83	2.75	-.38	.38	.57	-1.1	.57	-1.1	1.47	.56	.53	38.7	33.4	59 59	in subset: 2
34	12	2.83	2.75	-.38	.38	.84	-.3	.83	-.3	1.19	.31	.53	36.3	33.4	137 137	in subset: 2
34	12	2.83	2.75	-.38	.38	.61	-1.0	.60	-1.0	1.43	.52	.53	41.9	33.4	139 139	in subset: 2
34	12	2.83	2.75	-.37	.38	1.39	1.0	1.40	1.0	.57	.58	.53	30.3	34.5	6 6	in subset: 3
34	12	2.83	2.75	-.37	.38	.63	-.9	.63	-.9	1.44	.66	.53	41.2	34.5	32 32	in subset: 3

34	12	2.83	2.75	-.37	.38	1.29	.8	1.25	.7	.75	.50	.53	35.7	34.5	47 47	in subset: 3
34	12	2.83	2.75	-.37	.38	1.50	1.2	1.55	1.3	.42	.18	.53	32.1	34.5	50 50	in subset: 3
34	12	2.83	2.75	-.37	.38	1.16	.5	1.19	.5	.81	-.01	.53	32.8	34.5	86 86	in subset: 3
34	12	2.83	2.75	-.37	.38	.45	-1.6	.46	-1.6	1.57	.49	.53	41.1	34.5	128 128	in subset: 3
34	12	2.83	2.75	-.37	.38	1.02	.1	1.02	.1	.98	.13	.53	36.7	34.5	131 131	in subset: 3
32	12	2.67	2.80	-.29	.39	.94	.0	.98	.0	1.02	.69	.53	26.8	33.3	55 55	in subset: 4
34	12	2.83	2.82	-.25	.38	.59	-1.1	.57	-1.1	1.46	.55	.53	35.4	34.4	8 8	in subset: 1
34	12	2.83	2.82	-.25	.38	.41	-1.8	.40	-1.8	1.66	.72	.53	38.8	34.4	25 25	in subset: 1
34	12	2.83	2.82	-.25	.38	1.39	1.0	1.40	1.0	.52	.58	.53	33.6	34.4	61 61	in subset: 1
34	12	2.83	2.82	-.25	.38	.80	-.4	.79	-.4	1.19	.69	.53	38.1	34.4	104 104	in subset: 1
34	12	2.83	2.82	-.25	.38	.68	-.7	.71	-.7	1.33	.78	.53	37.4	34.4	106 106	in subset: 1
34	12	2.83	2.82	-.25	.38	.64	-.9	.63	-.9	1.38	.66	.53	36.3	34.4	178 178	in subset: 1
34	12	2.83	2.82	-.25	.38	5.85	6.2	5.81	6.2	-4.42	.22	.53	9.7	34.4	188 188	in subset: 1
34	12	2.83	2.82	-.25	.38	5.62	6.1	5.55	6.0	-4.13	-.01	.53	16.8	34.4	189 189	in subset: 1
35	12	2.92	2.83	-.23	.38	.56	-1.2	.56	-1.2	1.51	.67	.53	38.7	33.4	30 30	in subset: 2
35	12	2.92	2.83	-.23	.38	.55	-1.2	.55	-1.2	1.52	.68	.53	36.8	33.4	98 98	in subset: 2
35	12	2.92	2.83	-.23	.38	.54	-1.2	.54	-1.2	1.49	.52	.53	40.2	33.4	122 122	in subset: 2
35	12	2.92	2.83	-.23	.38	.51	-1.4	.53	-1.3	1.47	.72	.53	33.3	33.4	143 143	in subset: 2
35	12	2.92	2.84	-.23	.38	1.43	1.1	1.48	1.2	.44	.49	.53	33.2	34.5	11 11	in subset: 3
35	12	2.92	2.84	-.23	.38	.45	-1.6	.45	-1.6	1.60	.61	.53	41.3	34.5	12 12	in subset: 3
35	12	2.92	2.84	-.23	.38	.54	-1.3	.53	-1.3	1.57	.83	.53	43.4	34.5	18 18	in subset: 3
35	12	2.92	2.84	-.23	.38	.73	-.6	.73	-.6	1.31	.52	.53	36.7	34.5	96 96	in subset: 3
35	12	2.92	2.84	-.23	.38	.41	-1.8	.42	-1.8	1.60	.46	.53	40.5	34.5	110 110	in subset: 3
35	12	2.92	2.84	-.23	.38	.77	-.5	.77	-.5	1.23	.28	.53	39.5	34.5	116 116	in subset: 3
35	12	2.92	2.84	-.23	.38	.52	-1.3	.52	-1.3	1.52	.54	.53	41.0	34.5	119 119	in subset: 3
35	12	2.92	2.84	-.23	.38	.79	-.4	.82	-.3	1.25	.47	.53	38.5	34.5	156 156	in subset: 3
33	12	2.75	2.89	-.14	.38	2.57	2.9	2.65	3.0	-.75	-.08	.53	22.4	32.8	186 186	in subset: 4
35	12	2.92	2.91	-.11	.38	.51	-1.4	.51	-1.4	1.57	.71	.53	41.4	34.0	19 19	in subset: 1
35	12	2.92	2.91	-.11	.38	1.26	.7	1.26	.7	.73	.73	.53	30.8	34.0	89 89	in subset: 1
35	12	2.92	2.91	-.11	.38	1.41	1.0	1.41	1.0	.47	.35	.53	34.5	34.0	101 101	in subset: 1
35	12	2.92	2.91	-.11	.38	1.35	.9	1.37	.9	.61	.39	.53	28.4	34.0	102 102	in subset: 1
36	12	3.00	2.92	-.08	.38	.70	-.7	.69	-.7	1.34	.47	.53	36.0	33.2	27 27	in subset: 2
36	12	3.00	2.92	-.08	.38	1.14	.4	1.13	.4	.88	.74	.53	29.4	33.2	166 166	in subset: 2
36	12	3.00	2.92	-.08	.38	1.21	.6	1.22	.6	.76	.70	.53	31.7	34.3	17 17	in subset: 3
36	12	3.00	2.92	-.08	.38	.57	-1.1	.56	-1.2	1.52	.74	.53	41.5	34.3	54 54	in subset: 3
36	12	3.00	2.92	-.08	.38	.96	.0	.96	.0	.98	.43	.53	36.9	34.3	74 74	in subset: 3
36	12	3.00	2.92	-.08	.38	.38	-1.9	.38	-1.9	1.70	.76	.53	41.9	34.3	129 129	in subset: 3
36	12	3.00	2.92	-.08	.38	.47	-1.5	.47	-1.5	1.63	.82	.53	40.0	34.3	133 133	in subset: 3
36	12	3.00	2.92	-.08	.38	.60	-1.0	.60	-1.0	1.49	.72	.53	38.6	34.3	141 141	in subset: 3
36	12	3.00	2.92	-.08	.38	.66	-.8	.65	-.8	1.38	.50	.53	39.9	34.3	150 150	in subset: 3
34	12	2.83	2.97	.01	.38	2.43	2.7	2.50	2.8	-.63	-.06	.53	21.1	32.0	10 10	in subset: 4
34	12	2.83	2.97	.01	.38	.49	-1.4	.48	-1.5	1.57	.64	.53	32.0	32.0	16 16	in subset: 4
34	12	2.83	2.97	.01	.38	1.07	.2	1.10	.3	.96	.31	.53	26.8	32.0	23 23	in subset: 4
34	12	2.83	2.97	.01	.38	1.23	.6	1.28	.7	.71	.37	.53	29.8	32.0	117 117	in subset: 4
36	12	3.00	2.99	.04	.38	1.05	.2	1.04	.2	.97	.53	.53	33.5	33.5	78 78	in subset: 1
36	12	3.00	2.99	.04	.38	.61	-1.0	.61	-1.0	1.44	.55	.53	40.0	33.5	92 92	in subset: 1
36	12	3.00	2.99	.04	.38	.33	-2.2	.33	-2.2	1.76	.81	.53	39.4	33.5	124 124	in subset: 1
36	12	3.00	2.99	.04	.38	.92	.0	.95	.0	1.11	.45	.53	34.2	33.5	134 134	in subset: 1
36	12	3.00	2.99	.04	.38	1.06	.2	1.06	.2	.94	.14	.53	29.8	33.5	170 170	in subset: 1
37	12	3.08	3.01	.06	.38	.50	-1.4	.50	-1.4	1.54	.55	.53	39.3	33.9	49 49	in subset: 3
37	12	3.08	3.01	.06	.38	.96	.0	.96	.0	1.13	.65	.53	38.3	33.9	58 58	in subset: 3
37	12	3.08	3.01	.06	.38	1.01	.1	1.02	.1	.97	.62	.53	35.5	33.9	79 79	in subset: 3
37	12	3.08	3.01	.06	.38	.48	-1.5	.48	-1.5	1.54	.37	.53	39.0	33.9	103 103	in subset: 3
37	12	3.08	3.01	.06	.38	1.44	1.1	1.43	1.1	.42	.48	.53	34.2	33.9	115 115	in subset: 3
37	12	3.08	3.08	.18	.38	.90	-.1	.88	-.1	1.15	.70	.53	39.0	32.8	20 20	in subset: 1
37	12	3.08	3.08	.18	.38	.63	-.9	.62	-1.0	1.47	.75	.53	38.2	32.8	77 77	in subset: 1
37	12	3.08	3.08	.18	.38	.57	-1.2	.56	-1.2	1.47	.49	.53	32.9	32.8	80 80	in subset: 1
38	12	3.17	3.09	.20	.38	1.42	1.0	1.39	1.0	.53	.67	.53	29.7	32.4	46 46	in subset: 2
38	12	3.17	3.09	.20	.38	1.50	1.2	1.49	1.2	.43	.62	.53	29.7	32.4	84 84	in subset: 2
38	12	3.17	3.09	.20	.38	1.95	2.0	1.93	2.0	-.15	.63	.53	21.8	32.4	90 90	in subset: 2
38	12	3.17	3.09	.20	.38	.38	-1.9	.39	-1.9	1.66	.57	.53	36.3	32.4	107 107	in subset: 2
38	12	3.17	3.09	.20	.38	.57	-1.1	.57	-1.1	1.48	.55	.53	37.3	32.4	174 174	in subset: 2
38	12	3.17	3.09	.21	.38	.73	-.6	.73	-.6	1.27	.73	.53	37.3	33.4	34 34	in subset: 3
38	12	3.17	3.09	.21	.38	1.23	.6	1.23	.6	.72	.52	.53	33.8	33.4	38 38	in subset: 3
38	12	3.17	3.09	.21	.38	.64	-.9	.63	-.9	1.41	.49	.53	36.6	33.4	62 62	in subset: 3
38	12	3.17	3.09	.21	.38	.48	-1.5	.47	-1.5	1.62	.79	.53	39.8	33.4	120 120	in subset: 3
38	12	3.17	3.09	.21	.38	.76	-.5	.75	-.5	1.31	.56	.53	37.8	33.4	160 160	in subset: 3
38	12	3.17	3.09	.21	.38	.77	-.5	.75	-.5	1.33	.70	.53	37.4	33.4	161 161	in subset: 3
38	12	3.17	3.09	.21	.38	1.17	.5	1.16	.5	.77	.22	.53	35.4	33.4	169 169	in subset: 3
38	12	3.17	3.17	.33	.38	2.47	2.8	2.45	2.8	-.71	.36	.53	21.9	32.0	113 113	in subset: 1
39	12	3.25	3.18	.35	.38	.59	-1.1	.58	-1.1	1.48	.59	.53	34.6	31.8	71 71	in subset: 2
39	12	3.25	3.18	.35	.38	.60	-1.0	.59	-1.1	1.47	.58	.53	39.5	32.8	29 29	in subset: 3
39	12	3.25	3.18	.35	.38	.56	-1.2	.56	-1.2	1.44	.15	.53	34.7	32.8	51 51	in subset: 3
39	12	3.25	3.18	.35	.38	.60	-1.0	.59	-1.1	1.47	.58	.53	36.4	32.8	100 100	in subset: 3
39	12	3.25	3.18	.35	.38	.39	-1.9	.40	-1.9	1.67	.61	.53	40.7	32.8	130 130	in subset: 3
39	12	3.25	3.18	.35	.38	.79	-.4	.79	-.4	1.28	.58	.53	36.7	32.8	159 159	in subset: 3
37	12	3.08	3.24	.44	.38	2.02	2.1	2.01	2.1	-.16	.28	.53	24.1	29.1	190 190	in subset: 4
39	12	3.25	3.25	.47	.38	.52	-1.4	.50	-1.4	1.60	.80	.53	31.4	31.1	42 42	in subset: 1
40	12	3.33	3.27	.49	.38	1.00	.1	1.00	.1	.98	.01	.53	30.4	31.1	153 153	in subset: 2
40	12	3.33	3.27	.49	.38	.67	-.8	.68	-.8	1.40	.72	.53	35.3	31.1	163 163	in subset: 2
40	12	3.33	3.27	.50	.38	1.23	.6	1.20	.6	.71	.61	.53	32.1	32.0	9 9	in subset: 3
40	12	3.33	3.27	.50	.38	.45	-1.7	.45	-1.6	1.63	.59	.53	36.9	32.0	67 67	in subset: 3
40	12	3.33	3.27	.50	.38	.94	.0	.94	.0	1.09	.79	.53	35.8	32.0	114 114	in subset: 3
40	12	3.33	3.34	.62	.38	.41	-1.8	.42	-1.8	1.67	.63	.53	33.5	30.0	172 172	in subset: 1
41	12	3.42	3.35	.64	.38	1.89	1.9	1.89	1.9	.03	.52	.53	26.1	31.1	7 7	in subset: 3
41	12	3.42	3.35	.64	.38	.94	.0	.93	.0	1.07	.52	.53	36.3	31.1	41 41	in subset: 3
41	12	3.42	3.35	.64	.38	.64	-.9	.64	-.9	1.45	.76	.53	33.5	31.1	148 148	in subset: 3
41	12	3.42	3.43	.76	.38	1.09	.3	1.06	.2	.88	.71	.53	26.3	28.8	73 73	in subset: 1
41	12	3.42	3.43	.76	.38	1.31	.8	1.28	.7	.62	.58	.53	22.9	28.8	132 132	in subset: 1
42	12	3.50	3.44	.79	.38	.41	-1.8	.42	-1.8	1.70	.82	.53	30.6	29.2	95 95	in subset: 2
42	12	3.50	3.44	.79	.38	.78	-.4	.78	-.4	1.25	.50	.53	28.8	30.0	60 60	in subset: 3
42	12	3.50	3.44	.79	.38	.37	-2.0	.35	-2.1	1.76	.71	.53	39.0	30.0	70 70	in subset: 3
42	12	3.50	3.44	.79	.38	.85	-.2	.85	-.2	1.16	.61	.53	33.7	30.0	123 123	in subset: 3
42	12	3.50	3.44	.79	.38	.96	.0	.96	.0	1.07	.68	.53	33.5	30.0	149 149	in subset: 3
43	12	3.58	3.53	.94	.39	.97	.0	.96	.0	1.04	.34	.52	24.8	28.1	179 179	in subset: 2
43	12	3.58	3.60	1.06	.39	.17	-3.3	.16	-3.4	1.96	.78	.52	29.0	26.2	157 157	in subset: 1
44	12	3.67	3.62	1.09	.39	1.00	.1	.98	.0	1.01	.66	.52	31.0	27.5	21 21	in subset: 3
42	12	3.50	3.66	1.17	.38	1.27	.7	1.24	.7	.74	.62	.53	19.7	22.5	177 177	in subset: 4
44	12	3.67	3.68	1.21	.39	1.39	1.0	1.32	.8	.55	.68	.52	23.5	24.8	31 31	in subset: 1
44	12	3.67	3.68	1.21	.39	2.04	2.2	1.99	2.1	-.23	.44	.52	27.7	24.8	187 187	in subset: 1
45	12	3.75	3.70	1.24	.39	.78	-.4	.82	-.3	1.19	.19	.52	27.0	25.6	152 152	in subset: 2
45	12	3.75	3.70	1.24	.39	.43	-1.7	.42	-1.8	1.66	.59	.52	31.1	26.2	109 109	in subset: 3
46	12	3.83	3.79	1.40	.40	3.03	3.5	2.98	3.5	-1.36	.49	.51	14.7	24.3	158 158	in subset: 2
44	12	3.67	3.83	1.47	.39	1.10	.3	1.05	.2	.92	.58	.52	19.7	19.5	175 175	in subset: 4
47	12	3.92	3.88	1.56	.41	1.45	1.1	1.40	1.0	.62	.77	.51	19.4	23.2	97 97	in subset: 3
49	12	4.08	4.05	1.90	.42	2.54	2.9	2.42	2.7	-.67	.40	.49	17.4	19.8	171 171	in subset: 2
49	12	4.08	4.11	2.02	.42	1.86	1.8	1.71	1.6	.21	.63	.49	16.2	17.1	22 22	in subset: 1
34.0	12.0	2.83	2.79	-.32	.39	1.00	-.2	1.00	-.2		.50				Mean (Count: 190)	
5.6	.0	.47	.47	.83	.02	.70	1.4	.69	1.4		.24				S.D. (Population)	
5.6	.0	.47	.47	.84	.02	.70	1.4	.70	1.4		.24				S.D. (Sample)	

Model, Populn: RMSE .39 Adj (True) S.D. .74 Separation 1.88 Strata 2.85 Reliability (not inter-rater) .78
Model, Sample: RMSE .39 Adj (True) S.D. .74 Separation 1.89 Strata 2.85 Reliability (not inter-rater) .78
Model, Fixed (all same) chi-square: 796.6 d.f.: 189 significance (probability): .00
Model, Random (normal) chi-square: 156.7 d.f.: 188 significance (probability): .95
Inter-Rater agreement opportunities: 64608 Exact agreements: 22013 = 34.1% Expected: 20798.6 = 32.2%

Rasch Analysis of Expert Ratings 22/12/2014 16:52:26
Table 7.2.1 Spanish English Familiarity Level Measurement Report (arranged by MN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	N Spanish English Familiarity Level	
607	240	2.53	2.63	-.26	.09	1.31	3.2	1.30	3.1	.65	.59	.67	1 No Familiarity	in subset: 4
1899	684	2.78	2.78	.00	.05	1.14	2.5	1.13	2.3	.84	.63	.63	2 Limited Familiarity	in subset: 1
2723	936	2.91	2.85	.12	.04	.82	-4.4	.82	-4.3	1.20	.65	.62	3 Some Familiarity	in subset: 3
1224	420	2.91	2.85	.13	.07	.99	.0	1.00	.0	1.00	.66	.66	4 Very Familiar	in subset: 2

```
|-------------------------------+---------------+-----------------------+------+-------------+-----------------------------------|
|  1613.3   570.0    2.78  2.78 |    .00   .06  | 1.07   .3  1.06   .3  |      |  .63        | Mean (Count: 4)                   |
|   787.0   263.8     .16   .09 |    .16   .02  |  .18  3.0   .18  2.9  |      |  .03        | S.D. (Population)                 |
|   908.7   304.6     .18   .10 |    .18   .02  |  .21  3.5   .20  3.4  |      |  .03        | S.D. (Sample)                     |
+--------------------------------------------------------------------------------------------------------------------------------+
Model, Populn: RMSE .07  Adj (True) S.D. .14  Separation 2.17  Strata 3.22  Reliability .82
Model, Sample: RMSE .07  Adj (True) S.D. .17  Separation 2.57  Strata 3.76  Reliability .87
Model, Fixed (all same) chi-square:  16.6  d.f.: 3  significance (probability): .00
Model,  Random (normal) chi-square:  2.5  d.f.: 2  significance (probability): .28
----------------------------------------------------------------------------------------------------------------------------------
```

```
Rasch Analysis of Expert Ratings 22/12/2014 16:52:26
Table 7.3.1  Speaker Measurement Report  (arranged by MN).

+-------------------------------------------------------------------------------------------------------------------------------------+
|  Total   Total   Obsvd  Fair(M)|        Model | Infit      Outfit    |Estim.| Correlation |                                         |
|  Score   Count  Average Average|Measure  S.E. | MnSq ZStd  MnSq ZStd |Discrm| PtMea PtExp | Nu Speaker                              |
|--------------------------------+--------------+----------------------+------+-------------+-----------------------------------------|
|   424     190     2.23   2.18  |  -1.10   .10 |  .74 -2.7   .74 -2.8 | 1.30 |   .52   .51 |  4 Speaker 5 Spanish-English Male 1     |
|   436     190     2.29   2.24  |   -.98   .10 | 1.35  3.2  1.37  3.3 |  .58 |   .35   .52 |  2 Speaker 3 Arabic-English Female 1    |
|   471     190     2.48   2.43  |   -.62   .10 | 1.00   .0  1.01   .0 |  .97 |   .44   .53 |  8 Speaker 9 Spanish-English Female 1   |
|   479     190     2.52   2.47  |   -.54   .10 | 1.30  2.7  1.29  2.7 |  .68 |   .40   .53 |  1 Speaker 2 Dhivehi-English Female 1   |
|   490     190     2.58   2.53  |   -.44   .10 |  .82 -1.8   .83 -1.8 | 1.18 |   .53   .53 |  6 Speaker 7 Spanish-English Male 2     |
|   505     190     2.66   2.61  |   -.29   .10 |  .85 -1.6   .84 -1.6 | 1.19 |   .51   .54 |  3 Speaker 4 Dhivehi-English Female 2   |
|   530     190     2.79   2.75  |   -.06   .10 | 1.56  4.8  1.55  4.8 |  .38 |   .59   .54 |  7 Speaker 8 Dhivehi-English Male 1     |
|   590     190     3.11   3.07  |    .50   .10 |  .84 -1.7   .83 -1.8 | 1.19 |   .65   .54 |  5 Speaker 6 Arabic-English Male 1      |
|   602     190     3.17   3.14  |    .61   .10 | 1.02   .2  1.02   .2 |  .99 |   .63   .54 | 11 Speaker 12 Arabic-English Female 2   |
|   618     190     3.25   3.23  |    .75   .10 |  .79 -2.3   .79 -2.2 | 1.24 |   .64   .54 | 12 Speaker 13 Dhivehi-English Male 2    |
|   654     190     3.44   3.43  |   1.09   .10 |  .78 -2.4   .77 -2.5 | 1.25 |   .60   .54 |  9 Speaker 10 Arabic-English 2          |
|   654     190     3.44   3.43  |   1.09   .10 |  .92  -.7   .91  -.8 | 1.10 |   .56   .54 | 10 Speaker 11 Spanish-English Female 2  |
|--------------------------------+--------------+----------------------+------+-------------+-----------------------------------------|
|   537.8   190.0    2.83   2.79 |    .00   .10 | 1.00  -.2  1.00  -.2 |      |   .53       | Mean (Count: 12)                        |
|    79.0      .0     .42    .43 |    .75   .00 |  .25  2.4   .25  2.4 |      |   .09       | S.D. (Population)                       |
|    82.5      .0     .43    .45 |    .78   .00 |  .27  2.5   .27  2.6 |      |   .10       | S.D. (Sample)                           |
+-------------------------------------------------------------------------------------------------------------------------------------+
Model, Populn: RMSE .10  Adj (True) S.D. .74  Separation 7.54  Strata 10.39  Reliability .98
Model, Sample: RMSE .10  Adj (True) S.D. .77  Separation 7.88  Strata 10.84  Reliability .98
Model, Fixed (all same) chi-square:  685.8  d.f.: 11  significance (probability): .00
Model,  Random (normal) chi-square:  10.8  d.f.: 10  significance (probability): .37
---------------------------------------------------------------------------------------------------------------------------------------
```

```
Rasch Analysis of Expert Ratings 22/12/2014 16:52:26
Table 8.1  Category Statistics.

 Model = ?,?,?,R5
+-------------------------------------------------------------------------------------------------------------+
|           DATA                    |  QUALITY CONTROL |RASCH-ANDRICH| EXPECTATION    |  MOST  | RASCH-    | Cat|
|      Category Counts          Cum.|  Avge  Exp. OUTFIT| Thresholds  |  Measure at    |PROBABLE| THURSTONE |PEAK|
|Score Total      Used     %      % |  Meas  Meas  MnSq |Measure  S.E.|Category  -0.5  |  from  |Thresholds |Prob|
|-----------------------------------+------------------+-------------+----------------+--------+-----------+----|
|  1     166       166     7%    7% | -1.48  -1.67  1.2 |             |( -3.90)        |  low   |   low     |100%|
|  2     709       709    31%   38% |  -.94   -.91  1.0 |  -2.74   .09|  -1.78   -2.99 | -2.74  |  -2.85    | 57%|
|  3     842       842    37%   75% |  -.19   -.14  1.0 |   -.69   .05|    .05    -.81 |  -.69  |   -.76    | 50%|
|  4     472       472    21%   96% |   .69    .61   .9 |    .82   .06|   1.80     .89 |   .82  |    .85    | 54%|
|  5      91        91     4%  100% |  1.33   1.31  1.0 |   2.61   .12|(  3.80)   2.93 |  2.61  |   2.74    |100%|
+---------------------------------------------------------------------------(Mean)---------(Modal)--(Median)------+
```

```
Scale structure

Measr:-4.0            -2.0             0.0             2.0             4.0
        +               +               +               +               +
  Mode:<(^)-------12-------^-------23-----^-----34-------^-----45--------(^)5>

Median:<(^)------12--------^-------23-----^-----34-------^------45-------(^)5>

  Mean:<(^)-----12---------^------23------^------34------^--------45-----(^)5>
        +               +               +               +               +
Measr:-4.0            -2.0             0.0             2.0             4.0
```

```
Probability Curves
      -4.0            -2.0             0.0             2.0             4.0
       ++---------------+---------------+---------------+---------------++
     1 |                                                                 |
       |                                                                 |
       |                                                                 |
       |                                                                 |
       |                                                                5|
     P |111                                                          555 |
     r |   11                                                       5    |
     o |     1                                                    55     |
     b |      11                                                55       |
     a |       1     2222222222                     444444     5         |
     b |        11222         222    3333333     444      444 55         |
     i |        2211             2233       333 44           *4          |
     l |       22   1            332          4*3          55  44        |
     i |     22      11        33   22      44   33       5     444      |
     t |   22          1      33      22  44      33    55         44    |
     y | 22             11  33          24          3355             44  |
       |2                 **           4422         5533               44|
       |               333  111     444    222    55    333              |
       |            333       111444          22*55        333           |
       |      3333333        44444111111  555555 22222        333333     |
     0 |*********************5555555555***11111111111********************|
       ++---------------+---------------+---------------+---------------++
      -4.0            -2.0             0.0             2.0             4.0
```

```
Expected Score Ogive (Model ICC)
      -4.0            -2.0             0.0             2.0             4.0
       ++---------------+---------------+---------------+---------------++
     5 |                                                                 |
       |                                                           555555|
       |                                                      445555     |
       |                                                 44444           |
     4 |                                             4444                |
       |                                         4444                    |
       |                                      3344                       |
       |                                   333                           |
     3 |                               3333                              |
       |                           3333                                  |
       |                        223                                      |
       |                     2222                                        |
     2 |                22222                                            |
       |           22222                                                 |
       |     1111222                                                     |
       |11111                                                            |
     1 |                                                                 |
       ++---------------+---------------+---------------+---------------++
      -4.0            -2.0             0.0             2.0             4.0
```

```
Rasch Analysis of Expert Ratings 22/12/2014 16:52:26
Table 4.1 Unexpected Responses (9 residuals sorted by u).

+-----------------------------------------------------------------------------------------------+
| Cat  Score  Exp.  Resd StRes| Num rat N Spanish English Fam Nu Speaker                        |
|-----------------------------+-----------------------------------------------------------------|
|  5     5     2.4   2.6  3.5 | 118 118 3 Some Familiarity     7 Speaker 8 Dhivehi-English Male 1   |
|  1     1     3.6  -2.6 -3.4 | 158 158 4 Very Familiar        1 Speaker 2 Dhivehi-English Female 1 |
|  5     5     2.5   2.5  3.3 | 188 188 2 Limited Familiarity  1 Speaker 2 Dhivehi-English Female 1 |
|  5     5     2.5   2.5  3.3 | 189 189 2 Limited Familiarity  1 Speaker 2 Dhivehi-English Female 1 |
```

```
|  2     2     4.2  -2.2 -3.2 | 158 158 4 Very Familiar       11 Speaker 12 Arabic-English Female 2  |
|  1     1     3.5  -2.5 -3.2 | 189 189 2 Limited Familiarity  9 Speaker 10 Arabic-English 2         |
|  1     1     3.5  -2.5 -3.2 | 189 189 2 Limited Familiarity 10 Speaker 11 Spanish-English Female 2 |
|  5     5     2.7   2.3  3.1 | 188 188 2 Limited Familiarity  3 Speaker 4 Dhivehi-English Female 2  |
|  5     5     2.7   2.3  3.1 | 189 189 2 Limited Familiarity  3 Speaker 4 Dhivehi-English Female 2  |
|-----------------------------+----------------------------------------------------------------------|
| Cat  Score  Exp.  Resd StRes| Num rat N Spanish English Fam Nu Speaker                             |
+----------------------------------------------------------------------------------------------------+
```

The Facets analyses output of the Arabic-English pronunciation score data

```
Facets (Many-Facet Rasch Measurement) Version No. 3.71.1  Copyright �(c) 1987-2013, John M. Linacre. All rights reserved.
22/12/2014 16:49:51

Rasch Analysis of Expert Ratings 22/12/2014 16:49:51
Table 1. Specifications from file "C:\Users\kevin\Desktop\Main Study Data\Arabic English Pronunciaton Operating File.txt.txt".

Title = Rasch Analysis of Expert Ratings 22/12/2014 16:49:51
Data file = "C:\Users\kevin\Desktop\Main Study Data\Pronunciation Score Data Arabic Familiarity.xlsx"
Output file = C:\Users\kevin\Desktop\Main Study Data\Arabic English Pronunciaton Operating File.txt.out.txt

; Data specification
Facets = 3
Delements = N
Non-centered = 1
Positive =  1, 2, 3
Labels =
 1,rater ; (elements = 190)
 2,Arabic English Familiarity Level ; (elements = 4)
 3,Speaker ; (elements = 12)
Model = ?,?,?,R5,1

; Output description
Arrange tables in order = MN
Bias/Interaction direction = plus ; ability, easiness, leniency: higher score = positive logit
Fair score = Mean
Pt-biserial = Measure
Heading lines in output data files = Y
Inter-rater coefficients reported for facet = 1
Barchart = Yes
Total score for elements = Yes
T3onscreen show only one line on screen iteration report = Y
T4MAX maximum number of unexpected observations reported in Table 4 = 100
T8NBC show table 8 numbers-barcharts-curves = NBC
Unexpected observations reported if standardized residual >= 3
Usort unexpected observations sort order = u
WHexact - Wilson-Hilferty standardization = Y

; Convergence control
Convergence = .1, .01
Iterations (maximum) = 0 ; unlimited
Xtreme scores adjusted by = .3, .5  ;(estimation, bias)

Rasch Analysis of Expert Ratings 22/12/2014 16:49:51
Table 2. Data Summary Report.

Assigning models to Data= "C:\Users\kevin\Desktop\Main Study Data\Pronunciation Score Data Arabic Familiarity.xlsx"
Total lines in data file = 194
Total data lines = 190
Responses matched to model: ?,?,?,R5,1 = 2280
    Total non-blank responses found = 2280
              Number of blank lines = 3
Valid responses used for estimation = 2280

Rasch Analysis of Expert Ratings 22/12/2014 16:49:51
Table 3. Iteration Report.

+-----------------------------------------------------------+
| Iteration      Max. Score Residual      Max. Logit Change |
|             Elements    %  Categories  Elements     Steps |
|-----------------------------------------------------------|
| PROX   1                                 -1.6094          |
| PROX   2                                   .1590          |
| JMLE   3    401.5981 -18.9  -599.7460      .5635  -2.1163 |
| JMLE   4    -64.0283 -11.3   -38.6908    -.3328    .1414 |
| JMLE   5    -24.2679  -5.3    34.9753    -.1714   -.1439 |
| JMLE   6    -21.4753  -4.3    19.7349    -.1403   -.0859 |
| JMLE   7    -17.6322  -3.3    17.1111    -.1113   -.0742 |
| JMLE   8    -14.8284  -2.7    14.1364    -.0909   -.0615 |
| JMLE   9    -12.5667  -2.2    11.9268    -.0751   -.0520 |
| JMLE  10    -10.7317  -1.8    10.1446    -.0628   -.0443 |
| JMLE  11     -9.2200  -1.6     8.6917    -.0531   -.0380 |
| JMLE  12     -7.9610  -1.3     7.4897    -.0452   -.0327 |
| JMLE  13     -6.9025  -1.1     6.4843    -.0387   -.0284 |
| JMLE  14     -6.0058  -1.0     5.6357    -.0334   -.0247 |
| JMLE  15     -5.2411   -.9     4.9142    -.0289   -.0215 |
| JMLE  16     -4.5856   -.7     4.2969    -.0251   -.0188 |
| JMLE  17     -4.0208   -.7     3.7661    -.0219   -.0165 |
| JMLE  18     -3.5324   -.6     3.3074    -.0192   -.0145 |
| JMLE  19     -3.1084   -.5     2.9097    -.0168   -.0128 |
| JMLE  20     -2.7392   -.4     2.5637    -.0148   -.0112 |
| JMLE  21     -2.4169   -.4     2.2618    -.0130   -.0099 |
| JMLE  22     -2.1349   -.3     1.9977    -.0115   -.0088 |
| JMLE  23     -1.8876   -.3     1.7662    -.0101   -.0078 |
| JMLE  24     -1.6704   -.3     1.5629    -.0089   -.0069 |
| JMLE  25     -1.4792   -.2     1.3840    -.0079   -.0061 |
| JMLE  26     -1.3109   -.2     1.2265    -.0070   -.0054 |
| JMLE  27     -1.1623   -.2     1.0875    -.0062   -.0048 |
| JMLE  28     -1.0311   -.2      .9648    -.0055   -.0042 |
| JMLE  29      -.9152   -.1      .8563    -.0049   -.0038 |
| JMLE  30      -.8126   -.1      .7603    -.0043   -.0033 |
| JMLE  31      -.7218   -.1      .6754    -.0038   -.0030 |
| JMLE  32      -.6413   -.1      .6001    -.0034   -.0026 |
| JMLE  33      -.5699   -.1      .5333    -.0030   -.0023 |
| JMLE  34      -.5067   -.1      .4741    -.0027   -.0021 |
| JMLE  35      -.4505   -.1      .4216    -.0024   -.0019 |
| JMLE  36      -.4006   -.1      .3749    -.0021   -.0016 |
| JMLE  37      -.3564   -.1      .3335    -.0019   -.0015 |
| JMLE  38      -.3170   -.1      .2967    -.0017   -.0013 |
| JMLE  39      -.2821    .0      .2640    -.0015   -.0012 |
| JMLE  40      -.2510    .0      .2349    -.0013   -.0010 |
| JMLE  41      -.2234    .0      .2090    -.0012   -.0009 |
| JMLE  42      -.1988    .0      .1861    -.0011   -.0008 |
| JMLE  43      -.1770    .0      .1656    -.0009   -.0007 |
| JMLE  44      -.1576    .0      .1475    -.0008   -.0006 |
| JMLE  45      -.1403    .0      .1313    -.0007   -.0006 |
| JMLE  46      -.1249    .0      .1169    -.0007   -.0005 |
| JMLE  47      -.1112    .0      .1041    -.0006   -.0005 |
| JMLE  48      -.0990    .0      .0927    -.0005   -.0004 |
+-----------------------------------------------------------+
Warning (6)! There may be 4 disjoint subsets


Rasch Analysis of Expert Ratings 22/12/2014 16:49:51
Table 4. Unexpected Responses - appears after Table 8.
```

```
Rasch Analysis of Expert Ratings 22/12/2014 16:49:51
Table 5. Measurable Data Summary.

+---------------------------------------------------+
| Cat  Score  Exp.  Resd StRes|                     |
|-----------------------------+---------------------|
| 2.83  2.83  2.83   .00   .00 | Mean (Count: 2280) |
|  .97   .97   .63   .74  1.00 | S.D. (Population)  |
|  .97   .97   .63   .74  1.00 | S.D. (Sample)      |
+---------------------------------------------------+
Data log-likelihood chi-square = 5051.5679
Approximate degrees of freedom = 2073
Chi-square significance prob.  = .0000
                                         Count   Mean   S.D.   Params
Responses used for estimation      =      2280   2.83   0.97      207
Count of measurable responses      =      2280
Raw-score variance of observations   =   0.94 100.00%
Variance explained by Rasch measures =   0.39  42.00%
Variance of residuals                =   0.55  58.00%

Rasch Analysis of Expert Ratings 22/12/2014 16:49:51
Table 6.0  All Facet Vertical "Rulers".

Vertical = (1*,2A,3A,S) Yardstick (columns lines low high extreme)= 0,8,-3,3,End
+--------------------------------------------------------------------------------------------------------------------------------------------+
|Measr|+rater    |+Arabic English Familiarity Level         |+Speaker                                                                  |Scale|
|     |       lenient  | lenient                                  | More able                                                                |     |
|-----+----------+------------------------------------------+--------------------------------------------------------------------------+-----|
|   3 +          +                                          +                                                                          + (5) |
|     |          |                                          |                                                                          | --- |
|     |          |                                          |                                                                          |     |
|     |          |                                          |                                                                          |     |
|     |          |                                          |                                                                          |     |
|     |          |                                          |                                                                          |     |
|     |          |                                          |                                                                          |     |
|     | .        |                                          |                                                                          |     |
|   2 +          +                                          +                                                                          +     |
|     | .        |                                          |                                                                          |     |
|     |          |                                          |                                                                          |  4  |
|     | .        |                                          |                                                                          |     |
|     | .        |                                          |                                                                          |     |
|     |          |                                          |                                                                          |     |
|     | .        |                                          |                                                                          |     |
|     | *.       |                                          | Speaker 10 Arabic-English 2          Speaker 11 Spanish-English Female 2 |     |
|   1 + **       +                                          +                                                                          +     |
|     | *.       |                                          |                                                                          | --- |
|     | *        |                                          | Speaker 13 Dhivehi-English Male 2                                        |     |
|     | *.       |                                          | Speaker 12 Arabic-English Female 2                                       |     |
|     | *.       |                                          | Speaker 6 Arabic-English Male 1                                          |     |
|     | ***      |                                          |                                                                          |     |
|     | ****     |                                          |                                                                          |     |
|     | ****     | Some Familiarity     Very Familiar       |                                                                          |     |
*   0 * ***.     *                                          * Speaker 8 Dhivehi-English Male 1                                         *  3  *
|     | *******. | Limited Familiarity  No Familiarity      |                                                                          |     |
|     | **.      |                                          | Speaker 4 Dhivehi-English Female 2                                       |     |
|     | **.      |                                          | Speaker 7 Spanish-English Male 2                                         |     |
|     | ***      |                                          | Speaker 2 Dhivehi-English Female 1                                       |     |
|     | ***      |                                          | Speaker 9 Spanish-English Female 1                                       |     |
|     | ****     |                                          |                                                                          | --- |
|     | ****.    |                                          |                                                                          |     |
|  -1 + .        +                                          + Speaker 3 Arabic-English Female 1                                        +     |
|     | **       |                                          | Speaker 5 Spanish-English Male 1                                         |     |
|     | **       |                                          |                                                                          |     |
|     | **.      |                                          |                                                                          |     |
|     |          |                                          |                                                                          |     |
|     | *        |                                          |                                                                          |     |
|     | .        |                                          |                                                                          |  2  |
|     | .        |                                          |                                                                          |     |
|  -2 +          +                                          +                                                                          +     |
|     | *.       |                                          |                                                                          |     |
|     |          |                                          |                                                                          |     |
|     |          |                                          |                                                                          |     |
|     | .        |                                          |                                                                          |     |
|     |          |                                          |                                                                          |     |
|     | .        |                                          |                                                                          |     |
|     |          |                                          |                                                                          |     |
|  -3 +          +                                          +                                                                          + (1) |
|-----+----------+------------------------------------------+--------------------------------------------------------------------------+-----|
|     |severe    | severe                                   | Less able                                                                |     |
|Measr| * = 3    |+Arabic English Familiarity Level         |+Speaker                                                                  |Scale|
+--------------------------------------------------------------------------------------------------------------------------------------------+

S.1: Model = ?,?,?,R5
There are 4 disconnected subsets identified in Table 7.

Rasch Analysis of Expert Ratings 22/12/2014 16:49:51
Table 6.1  rater Facet Summary.

Logit:
                      11    12 11 1
    1 1   5 1 23 7246132276505112119434136411111  1 1
 +---------+Q-------S+------M--+-----S---+---Q-----+---------+
-3        -2        -1         0         1         2         3

Infit MnSq:
     112111111 1
   4467524553273431 321 1221   1                        1  1
 +--S------M------S--+---Q-----+---------+---------+---------+
 0         1         2         3         4         5         6

Outfit MnSq:
     122111111 1
   4430617090292432 221 2211   1                       1  1
 +--S------M------S--+---Q-----+---------+---------+---------+
 0         1         2         3         4         5         6

Infit ZStd:
               11  1 1111 1
    121  22588854790661016273112 31 32  1           2
 +----Q----+-S--+---M+----+S---+---Q+----+----+----+----+
-4   -3   -2   -1    0    1    2    3    4    5    6    7

Outfit ZStd:
             1 11    11 1 1
    121 113371863889582718253 3 23 231  1           11
 +----Q----+-S--+---M+----+S---+--Q-+----+----+----+----+
-4   -3   -2   -1    0    1    2    3    4    5    6    7


Rasch Analysis of Expert Ratings 22/12/2014 16:49:51
Table 6.2  Arabic English Familiarity Level Facet Summary.

Logit:
                              2 2
 +---------+---------+-------QSMSQ-------+---------+---------+
-3        -2        -1         0         1         2         3

Infit MnSq:
```

```
          12 1
 +-------QSMS-Q------+---------+---------+---------+---------+
 0         1         2         3         4         5         6
```

Outfit MnSq:

```
          12 1
 +-------QSMSQ-------+---------+---------+---------+---------+
 0         1         2         3         4         5         6
```

Infit ZStd:

```
     1              1   1           1
 +----+---S+----+----M----+----+S---+----+-Q--+----+----+
-4   -3   -2   -1    0    1    2    3    4    5    6    7
```

Outfit ZStd:

```
     1              1   1          1
 +----+---S+----+----M----+----+S---+----+-Q--+----+----+
-4   -3   -2   -1    0    1    2    3    4    5    6    7
```

Rasch Analysis of Expert Ratings 22/12/2014 16:49:51
Table 6.3 Speaker Facet Summary.

Logit:

```
                    11   1111 1     11 1  2
 +---------+----Q----+--S------M------S--+----Q----+---------+
-3        -2        -1         0         1         2         3
```

Infit MnSq:

```
        1512  11 1
 +----Q-S--M--S-Q----+---------+---------+---------+---------+
 0         1         2         3         4         5         6
```

Outfit MnSq:

```
        1512  11 1
 +----Q-S--M--S-Q----+---------+---------+---------+---------+
 0         1         2         3         4         5         6
```

Infit ZStd:

```
       1 2  21   1   11            1 1       1
 +----+-S--+----+---M+----+----+S---+----+--Q-+----+----+
-4   -3   -2   -1    0    1    2    3    4    5    6    7
```

Outfit ZStd:

```
       11 1 21   1   11            1  1      1
 +----+-S--+----+---M+----+----+S---+----+--Q-+----+----+
-4   -3   -2   -1    0    1    2    3    4    5    6    7
```

Rasch Analysis of Expert Ratings 22/12/2014 16:49:51
Table 7.1.1 rater Measurement Report (arranged by MN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	Exact Agree. Obs %	Exact Agree. Exp %	Num	rater	
20	12	1.67	1.61	-2.71	.47	.57	-1.1	.59	-1.0	1.51	.71	.46	16.9	18.4	24	24	in subset: 3
20	12	1.67	1.67	-2.54	.47	.37	-1.9	.43	-1.7	1.70	.63	.46	23.1	22.6	138	138	in subset: 1
22	12	1.83	1.85	-2.13	.44	.31	-2.3	.35	-2.1	1.79	.49	.48	25.4	26.0	145	145	in subset: 1
22	12	1.83	1.85	-2.13	.44	1.10	.3	1.18	.5	.80	.23	.48	29.4	26.0	183	183	in subset: 1
22	12	1.83	1.85	-2.12	.44	2.08	2.2	2.10	2.2	-.34	.08	.48	21.8	25.8	72	72	in subset: 2
22	12	1.83	1.85	-2.12	.44	1.00	.1	1.03	.1	.97	.35	.48	24.8	25.8	162	162	in subset: 2
23	12	1.92	1.86	-2.10	.43	2.25	2.4	2.30	2.6	-.68	-.60	.48	24.5	23.6	75	75	in subset: 3
24	12	2.00	1.94	-1.92	.43	.44	-1.6	.45	-1.6	1.65	.63	.49	25.0	25.1	65	65	in subset: 3
24	12	2.00	2.02	-1.75	.43	.80	-.4	.77	-.5	1.21	.52	.49	28.3	28.8	147	147	in subset: 1
24	12	2.00	2.02	-1.74	.43	1.08	.3	1.04	.2	.97	.70	.49	29.3	28.8	40	40	in subset: 2
25	12	2.08	2.10	-1.57	.42	1.21	.6	1.21	.6	.67	-.09	.50	30.9	30.1	85	85	in subset: 1
25	12	2.08	2.10	-1.57	.42	.87	-.2	.94	.0	1.11	.53	.50	29.5	30.1	111	111	in subset: 1
26	12	2.17	2.11	-1.57	.41	1.04	.2	1.03	.1	.97	.64	.50	29.1	28.0	151	151	in subset: 3
27	12	2.25	2.19	-1.40	.41	.18	-3.1	.17	-3.2	1.94	.72	.51	34.1	29.3	33	33	in subset: 3
27	12	2.25	2.19	-1.40	.41	.94	.0	.98	.0	1.01	.04	.51	29.4	29.3	105	105	in subset: 3
26	12	2.17	2.19	-1.40	.41	1.56	1.3	1.65	1.5	.21	-.01	.50	29.1	31.2	5	5	in subset: 1
26	12	2.17	2.19	-1.40	.41	1.19	.5	1.20	.6	.71	.33	.50	34.4	31.2	69	69	in subset: 1
26	12	2.17	2.19	-1.40	.41	.93	.0	.94	.0	1.03	.30	.50	30.3	31.2	53	53	in subset: 2
26	12	2.17	2.19	-1.40	.41	1.25	.7	1.24	.6	.72	.67	.50	31.2	31.2	108	108	in subset: 2
26	12	2.17	2.19	-1.40	.41	.47	-1.5	.47	-1.5	1.61	.16	.50	29.5	31.2	180	180	in subset: 2
28	12	2.33	2.25	-1.27	.40	.63	-.9	.63	-.9	1.40	.46	.51	34.2	30.2	81	81	in subset: 4
28	12	2.33	2.25	-1.27	.40	1.07	.2	1.02	.1	1.03	.49	.51	26.8	30.2	181	181	in subset: 4
28	12	2.33	2.27	-1.24	.40	.66	-.8	.64	-.9	1.41	.62	.51	30.8	30.4	99	99	in subset: 3
28	12	2.33	2.27	-1.24	.40	.42	-1.7	.45	-1.6	1.63	.40	.51	35.7	30.4	125	125	in subset: 3
27	12	2.25	2.27	-1.23	.41	1.39	1.0	1.35	.9	.60	.75	.51	27.3	32.1	167	167	in subset: 1
27	12	2.25	2.28	-1.23	.41	1.07	.3	1.13	.4	.91	.45	.51	29.1	32.2	165	165	in subset: 2
29	12	2.42	2.35	-1.08	.40	.65	-.8	.68	-.7	1.34	.45	.52	34.3	31.4	43	43	in subset: 3
29	12	2.42	2.35	-1.08	.40	.86	-.2	.85	-.2	1.19	.80	.52	30.7	31.4	135	135	in subset: 3
28	12	2.33	2.36	-1.07	.40	1.13	.4	1.13	.4	.82	.42	.51	33.9	32.8	13	13	in subset: 1
28	12	2.33	2.36	-1.07	.40	.96	.0	1.00	.1	.95	.35	.51	32.5	32.8	136	136	in subset: 1
28	12	2.33	2.36	-1.07	.40	.92	.0	.97	.0	.98	.60	.51	37.0	32.8	142	142	in subset: 1
28	12	2.33	2.36	-1.07	.40	.50	-1.4	.49	-1.4	1.52	.57	.51	39.5	33.0	14	14	in subset: 2
30	12	2.50	2.42	-.96	.39	1.61	1.4	1.64	1.4	.33	.46	.52	29.4	32.2	118	118	in subset: 4
30	12	2.50	2.44	-.92	.39	1.21	.6	1.27	.7	.69	.36	.52	30.6	32.2	44	44	in subset: 3
30	12	2.50	2.44	-.92	.39	.19	-3.0	.20	-3.0	1.87	.76	.52	38.7	32.2	82	82	in subset: 3
30	12	2.50	2.44	-.92	.39	.43	-1.7	.46	-1.5	1.59	.43	.52	37.4	32.2	83	83	in subset: 3
29	12	2.42	2.45	-.91	.40	.66	-.8	.69	-.7	1.35	.80	.52	38.4	33.4	1	1	in subset: 1
29	12	2.42	2.45	-.91	.40	.28	-2.4	.26	-2.6	1.85	.81	.52	38.0	33.4	39	39	in subset: 1
29	12	2.42	2.45	-.91	.40	1.32	.8	1.25	.7	.76	.08	.52	32.2	33.4	45	45	in subset: 1
29	12	2.42	2.45	-.91	.40	1.72	1.6	1.71	1.6	.18	.40	.52	29.5	33.4	64	64	in subset: 1
29	12	2.42	2.45	-.91	.40	.33	-2.2	.30	-2.3	1.80	.76	.52	38.7	33.4	88	88	in subset: 1
29	12	2.42	2.45	-.91	.40	1.07	.3	1.14	.4	.88	.48	.52	33.2	33.4	93	93	in subset: 1
29	12	2.42	2.45	-.91	.40	.87	-.2	.85	-.2	1.14	.44	.52	34.6	33.4	184	184	in subset: 1
29	12	2.42	2.45	-.91	.40	.82	-.3	.83	-.3	1.13	.25	.52	37.8	33.6	76	76	in subset: 2
29	12	2.42	2.45	-.91	.40	1.48	1.1	1.47	1.1	.34	-.08	.52	34.8	33.6	121	121	in subset: 2
29	12	2.42	2.45	-.91	.40	.73	-.6	.77	-.5	1.24	.35	.52	35.9	33.6	164	164	in subset: 2
31	12	2.58	2.52	-.77	.39	.83	-.3	.82	-.3	1.21	.65	.52	36.2	32.9	15	15	in subset: 3
31	12	2.58	2.52	-.77	.39	1.23	.6	1.24	.6	.82	.80	.52	34.2	32.9	56	56	in subset: 3
31	12	2.58	2.52	-.77	.39	.88	-.1	.95	.0	1.13	.18	.52	32.3	32.9	63	63	in subset: 3
30	12	2.50	2.53	-.76	.39	1.39	1.0	1.34	.9	.55	.42	.52	31.7	33.9	3	3	in subset: 1
30	12	2.50	2.53	-.76	.39	.18	-3.1	.19	-3.1	1.89	.78	.52	41.9	33.9	28	28	in subset: 1
30	12	2.50	2.53	-.76	.39	.54	-1.2	.56	-1.2	1.48	.30	.52	38.0	33.9	35	35	in subset: 1
30	12	2.50	2.53	-.76	.39	.56	-1.2	.53	-1.3	1.52	.53	.52	37.7	33.9	68	68	in subset: 1
30	12	2.50	2.53	-.76	.39	.79	-.4	.78	-.4	1.27	.68	.52	37.4	33.9	94	94	in subset: 1
30	12	2.50	2.53	-.76	.39	1.35	.9	1.36	.9	.62	.76	.52	34.0	33.9	155	155	in subset: 1
30	12	2.50	2.53	-.75	.39	1.42	1.0	1.42	1.0	.51	-.02	.52	33.8	34.1	4	4	in subset: 2
30	12	2.50	2.53	-.75	.39	1.01	.1	.96	.0	1.02	.52	.52	37.2	34.1	168	168	in subset: 2
30	12	2.50	2.53	-.75	.39	.59	-1.0	.62	-.9	1.44	.69	.52	36.3	34.1	185	185	in subset: 2
32	12	2.67	2.59	-.65	.39	.54	-1.2	.55	-1.2	1.46	.25	.53	35.1	33.6	66	66	in subset: 4
32	12	2.67	2.59	-.65	.39	.44	-1.6	.43	-1.7	1.61	.61	.53	38.2	33.6	87	87	in subset: 4
32	12	2.67	2.61	-.62	.39	.95	.0	.92	.0	1.11	.66	.53	36.6	33.4	52	52	in subset: 3
32	12	2.67	2.61	-.62	.39	.55	-1.2	.55	-1.2	1.52	.67	.53	38.3	33.4	57	57	in subset: 3
32	12	2.67	2.61	-.62	.39	.75	-.5	.74	-.5	1.27	.29	.53	37.8	33.4	154	154	in subset: 3
31	12	2.58	2.62	-.60	.39	.71	-.6	.71	-.7	1.29	.07	.52	35.4	34.1	26	26	in subset: 1
31	12	2.58	2.62	-.60	.39	1.16	.5	1.19	.5	.82	.70	.52	29.5	34.1	144	144	in subset: 1
31	12	2.58	2.62	-.60	.39	1.47	1.1	1.45	1.1	.45	.36	.52	32.1	34.3	112	112	in subset: 2
31	12	2.58	2.62	-.60	.39	.62	-.9	.63	-.9	1.38	.65	.52	37.4	34.3	146	146	in subset: 2
33	12	2.75	2.67	-.51	.38	1.59	1.4	1.58	1.3	.36	.18	.53	28.1	34.0	36	36	in subset: 4
33	12	2.75	2.67	-.51	.38	.93	.0	.94	.0	1.09	.29	.53	35.1	34.0	48	48	in subset: 4
33	12	2.75	2.69	-.47	.38	1.08	.3	1.08	.3	.91	.54	.53	35.4	33.8	37	37	in subset: 3
33	12	2.75	2.69	-.47	.38	.95	.0	.91	-.1	1.16	.64	.53	40.2	33.8	127	127	in subset: 3
33	12	2.75	2.69	-.47	.38	.71	-.6	.70	-.7	1.30	.27	.53	39.3	33.8	176	176	in subset: 3
32	12	2.67	2.70	-.45	.39	.61	-1.0	.59	-1.1	1.44	.44	.53	40.2	34.2	2	2	in subset: 1
32	12	2.67	2.71	-.45	.39	.94	.0	.98	.0	1.02	.69	.53	34.2	34.4	55	55	in subset: 2

32	12	2.67	2.71	-.45	.39	1.10	.3	1.09	.3	.91	.84	.53	33.8	34.4	140	140	in subset: 2
32	12	2.67	2.71	-.45	.39	2.39	2.7	2.31	2.5	-.57	.08	.53	27.1	34.4	173	173	in subset: 2
34	12	2.83	2.76	-.36	.38	1.07	.2	1.10	.3	.96	.31	.53	33.8	34.3	23	23	in subset: 4
34	12	2.83	2.76	-.36	.38	1.16	.5	1.19	.5	.81	-.01	.53	33.8	34.3	86	86	in subset: 4
34	12	2.83	2.78	-.32	.38	1.39	1.0	1.40	1.0	.57	.58	.53	31.1	34.0	6	6	in subset: 3
34	12	2.83	2.78	-.32	.38	.63	-.9	.63	-.9	1.44	.66	.53	40.2	34.0	32	32	in subset: 3
34	12	2.83	2.78	-.32	.38	1.29	.8	1.25	.7	.75	.50	.53	34.3	34.0	47	47	in subset: 3
34	12	2.83	2.78	-.32	.38	.68	-.7	.71	-.7	1.33	.78	.53	39.6	34.0	106	106	in subset: 3
34	12	2.83	2.78	-.32	.38	1.02	.1	1.02	.1	.98	.13	.53	31.7	34.0	131	131	in subset: 3
34	12	2.83	2.78	-.32	.38	.84	-.3	.83	-.3	1.19	.31	.53	34.6	34.0	137	137	in subset: 3
33	12	2.75	2.79	-.30	.38	1.34	.9	1.37	.9	.65	.51	.53	36.6	34.1	91	91	in subset: 1
33	12	2.75	2.79	-.30	.38	2.57	2.9	2.65	3.0	-.75	-.08	.53	21.4	34.1	186	186	in subset: 1
33	12	2.75	2.79	-.30	.38	.97	.0	.96	.0	1.03	.76	.53	34.6	34.3	126	126	in subset: 2
33	12	2.75	2.79	-.30	.38	1.37	.9	1.36	.9	.55	.14	.53	33.8	34.3	182	182	in subset: 2
35	12	2.92	2.84	-.21	.38	.56	-1.2	.56	-1.2	1.51	.67	.53	41.7	34.4	30	30	in subset: 4
35	12	2.92	2.84	-.21	.38	.73	-.6	.73	-.6	1.31	.52	.53	37.3	34.4	96	96	in subset: 4
35	12	2.92	2.84	-.21	.38	.52	-1.3	.52	-1.3	1.52	.54	.53	40.4	34.4	119	119	in subset: 4
35	12	2.92	2.84	-.21	.38	.54	-1.2	.54	-1.2	1.49	.52	.53	39.0	34.4	122	122	in subset: 4
35	12	2.92	2.86	-.18	.38	.45	-1.6	.45	-1.6	1.60	.61	.53	40.8	34.0	12	12	in subset: 3
35	12	2.92	2.86	-.18	.38	.54	-1.3	.53	-1.3	1.57	.83	.53	42.3	34.0	18	18	in subset: 3
35	12	2.92	2.86	-.18	.38	.51	-1.4	.51	-1.4	1.57	.71	.53	42.7	34.0	19	19	in subset: 3
35	12	2.92	2.86	-.18	.38	.55	-1.2	.55	-1.2	1.52	.68	.53	38.9	34.0	98	98	in subset: 3
35	12	2.92	2.86	-.18	.38	.77	-.5	.77	-.5	1.23	.28	.53	38.2	34.0	116	116	in subset: 3
35	12	2.92	2.86	-.18	.38	.51	-1.4	.53	-1.3	1.47	.72	.53	37.3	34.0	143	143	in subset: 3
35	12	2.92	2.86	-.18	.38	.79	-.4	.82	-.3	1.25	.47	.53	38.1	34.0	156	156	in subset: 3
34	12	2.83	2.88	-.16	.38	.59	-1.1	.57	-1.1	1.46	.55	.53	38.1	33.8	8	8	in subset: 1
34	12	2.83	2.88	-.16	.38	2.43	2.7	2.50	2.8	-.63	-.06	.53	23.6	33.8	10	10	in subset: 1
34	12	2.83	2.88	-.16	.38	.41	-1.8	.40	-1.8	1.66	.72	.53	39.9	33.8	25	25	in subset: 1
34	12	2.83	2.88	-.16	.38	1.50	1.2	1.55	1.3	.42	.18	.53	31.7	33.8	50	50	in subset: 1
34	12	2.83	2.88	-.16	.38	.57	-1.1	.57	-1.1	1.47	.56	.53	36.6	33.8	59	59	in subset: 1
34	12	2.83	2.88	-.16	.38	1.39	1.0	1.40	1.0	.52	.58	.53	33.7	33.8	61	61	in subset: 1
34	12	2.83	2.88	-.16	.38	.80	-.4	.79	-.4	1.19	.69	.53	37.2	33.8	104	104	in subset: 1
34	12	2.83	2.88	-.16	.38	.45	-1.6	.46	-1.6	1.57	.49	.53	37.4	33.8	128	128	in subset: 1
34	12	2.83	2.88	-.16	.38	.61	-1.0	.60	-1.0	1.43	.52	.53	36.6	33.8	139	139	in subset: 1
34	12	2.83	2.88	-.16	.38	.64	-.9	.63	-.9	1.38	.66	.53	37.0	33.8	178	178	in subset: 1
34	12	2.83	2.88	-.16	.38	5.62	6.1	5.55	6.0	-4.13	-.01	.53	14.8	33.8	189	189	in subset: 1
34	12	2.83	2.88	-.15	.38	.49	-1.4	.48	-1.5	1.57	.64	.53	35.3	34.1	16	16	in subset: 2
34	12	2.83	2.88	-.15	.38	1.23	.6	1.28	.7	.71	.37	.53	33.1	34.1	117	117	in subset: 2
34	12	2.83	2.88	-.15	.38	5.85	6.2	5.81	6.2	-4.42	.22	.53	7.9	34.1	188	188	in subset: 2
36	12	3.00	2.93	-.07	.38	1.05	.2	1.04	.2	.97	.53	.53	35.5	34.2	78	78	in subset: 4
36	12	3.00	2.95	-.03	.38	1.21	.6	1.22	.6	.76	.70	.53	33.7	33.9	17	17	in subset: 3
36	12	3.00	2.95	-.03	.38	.70	-.7	.69	-.7	1.34	.47	.53	35.2	33.9	27	27	in subset: 3
36	12	3.00	2.95	-.03	.38	.57	-1.1	.56	-1.2	1.52	.74	.53	40.8	33.9	54	54	in subset: 3
36	12	3.00	2.95	-.03	.38	.33	-2.2	.33	-2.2	1.76	.81	.53	42.9	33.9	124	124	in subset: 3
36	12	3.00	2.95	-.03	.38	.38	-1.9	.38	-1.9	1.70	.76	.53	41.7	33.9	129	129	in subset: 3
36	12	3.00	2.95	-.03	.38	.47	-1.5	.47	-1.5	1.63	.82	.53	41.3	33.9	133	133	in subset: 3
35	12	2.92	2.96	-.01	.38	1.26	.7	1.26	.7	.73	.73	.53	29.4	33.4	89	89	in subset: 1
35	12	2.92	2.96	-.01	.38	.41	-1.8	.42	-1.8	1.60	.46	.53	33.2	33.4	110	110	in subset: 1
35	12	2.92	2.97	-.01	.38	1.43	1.1	1.48	1.2	.44	.49	.53	30.8	33.7	11	11	in subset: 2
35	12	2.92	2.97	-.01	.38	1.41	1.0	1.41	1.0	.47	.35	.53	31.0	33.7	101	101	in subset: 2
35	12	2.92	2.97	-.01	.38	1.35	.9	1.37	.9	.61	.39	.53	28.0	33.7	102	102	in subset: 2
37	12	3.08	3.04	.11	.38	.50	-1.4	.50	-1.4	1.54	.55	.53	39.7	33.6	49	49	in subset: 3
37	12	3.08	3.04	.11	.38	.63	-.9	.62	-1.0	1.47	.75	.53	39.8	33.6	77	77	in subset: 3
37	12	3.08	3.04	.11	.38	1.01	.1	1.02	.1	.97	.62	.53	36.7	33.6	79	79	in subset: 3
37	12	3.08	3.04	.11	.38	.48	-1.5	.48	-1.5	1.54	.37	.53	36.9	33.6	103	103	in subset: 3
37	12	3.08	3.04	.11	.38	1.44	1.1	1.43	1.1	.42	.48	.53	33.2	33.6	115	115	in subset: 3
36	12	3.00	3.05	.13	.38	.96	.0	.96	.0	.98	.43	.53	31.3	32.8	74	74	in subset: 1
36	12	3.00	3.05	.13	.38	.61	-1.0	.61	-1.0	1.44	.55	.53	36.1	32.8	92	92	in subset: 1
36	12	3.00	3.05	.13	.38	.66	-.8	.65	-.8	1.38	.50	.53	35.7	32.8	150	150	in subset: 1
36	12	3.00	3.05	.13	.38	1.06	.2	1.06	.2	.94	.14	.53	32.2	32.8	170	170	in subset: 1
36	12	3.00	3.05	.14	.38	.92	.0	.95	.0	1.11	.45	.53	29.5	33.1	134	134	in subset: 2
36	12	3.00	3.05	.14	.38	.60	-1.0	.60	-1.0	1.49	.72	.53	33.8	33.1	141	141	in subset: 2
36	12	3.00	3.05	.14	.38	1.14	.4	1.13	.4	.88	.74	.53	30.3	33.1	166	166	in subset: 2
38	12	3.17	3.10	.22	.38	.38	-1.9	.39	-1.9	1.66	.57	.53	37.3	33.5	107	107	in subset: 4
38	12	3.17	3.12	.25	.38	.73	-.6	.73	-.6	1.27	.73	.53	37.8	33.2	34	34	in subset: 3
38	12	3.17	3.12	.25	.38	1.23	.6	1.23	.6	.72	.52	.53	33.1	33.2	38	38	in subset: 3
38	12	3.17	3.12	.25	.38	1.42	1.0	1.39	1.0	.53	.67	.53	32.8	33.2	46	46	in subset: 3
38	12	3.17	3.12	.25	.38	.64	-.9	.63	-.9	1.41	.49	.53	35.6	33.2	62	62	in subset: 3
38	12	3.17	3.12	.25	.38	2.47	2.8	2.45	2.8	-.71	.36	.53	23.5	33.2	113	113	in subset: 3
38	12	3.17	3.12	.25	.38	.48	-1.5	.47	-1.5	1.62	.79	.53	40.0	33.2	120	120	in subset: 3
38	12	3.17	3.12	.25	.38	.76	-.5	.75	-.5	1.31	.56	.53	36.3	33.2	160	160	in subset: 3
37	12	3.08	3.14	.28	.38	.96	.0	.96	.0	1.13	.65	.53	34.8	32.1	58	58	in subset: 1
37	12	3.08	3.14	.28	.38	.90	-.1	.88	-.1	1.15	.70	.53	34.8	32.4	20	20	in subset: 2
37	12	3.08	3.14	.28	.38	.57	-1.2	.56	-1.2	1.47	.49	.53	34.8	32.4	80	80	in subset: 2
37	12	3.08	3.14	.28	.38	2.02	2.1	2.01	2.1	-.16	.28	.53	26.3	32.4	190	190	in subset: 2
39	12	3.25	3.19	.37	.38	.52	-1.4	.50	-1.4	1.60	.80	.53	31.1	32.9	42	42	in subset: 4
39	12	3.25	3.21	.40	.38	.60	-1.0	.59	-1.1	1.47	.58	.53	40.7	32.6	29	29	in subset: 3
39	12	3.25	3.21	.40	.38	.59	-1.1	.58	-1.1	1.48	.59	.53	38.6	32.6	71	71	in subset: 3
39	12	3.25	3.21	.40	.38	.39	-1.9	.40	-1.9	1.67	.61	.53	40.0	32.6	130	130	in subset: 3
38	12	3.17	3.22	.42	.38	1.50	1.2	1.49	1.2	.43	.62	.53	28.4	31.3	84	84	in subset: 1
38	12	3.17	3.22	.42	.38	1.95	2.0	1.93	2.0	-.15	.63	.53	25.4	31.3	90	90	in subset: 1
38	12	3.17	3.22	.42	.38	.77	-.5	.75	-.5	1.33	.70	.53	37.6	31.3	161	161	in subset: 1
38	12	3.17	3.23	.43	.38	1.17	.5	1.16	.5	.77	.22	.53	37.2	31.5	169	169	in subset: 2
38	12	3.17	3.23	.43	.38	.57	-1.1	.57	-1.1	1.48	.55	.53	36.1	31.5	174	174	in subset: 2
40	12	3.33	3.28	.51	.38	1.00	.1	1.00	.1	.98	.01	.53	34.6	32.1	153	153	in subset: 4
40	12	3.33	3.30	.54	.38	1.23	.6	1.20	.6	.71	.61	.53	33.3	31.9	9	9	in subset: 3
40	12	3.33	3.30	.54	.38	.94	.0	.94	.0	1.09	.79	.53	37.2	31.9	114	114	in subset: 3
40	12	3.33	3.30	.54	.38	.41	-1.8	.42	-1.8	1.67	.63	.53	38.8	31.9	172	172	in subset: 3
39	12	3.25	3.31	.57	.38	.56	-1.2	.56	-1.2	1.44	.15	.53	31.1	30.4	51	51	in subset: 1
39	12	3.25	3.31	.57	.38	.60	-1.0	.59	-1.1	1.47	.58	.53	31.3	30.4	100	100	in subset: 1
39	12	3.25	3.31	.57	.38	.79	-.4	.79	-.4	1.28	.58	.53	31.8	30.4	159	159	in subset: 1
41	12	3.42	3.36	.66	.38	1.89	1.9	1.89	1.9	.03	.52	.53	26.3	31.2	7	7	in subset: 4
41	12	3.42	3.38	.69	.38	1.31	.8	1.28	.7	.62	.58	.53	27.9	31.0	132	132	in subset: 3
40	12	3.33	3.40	.71	.38	.45	-1.7	.45	-1.6	1.63	.59	.53	30.9	29.3	67	67	in subset: 1
40	12	3.33	3.40	.71	.38	.67	-.8	.68	-.8	1.40	.72	.53	33.6	29.3	163	163	in subset: 1
42	12	3.50	3.47	.84	.38	.37	-2.0	.35	-2.1	1.76	.71	.53	38.9	30.0	70	70	in subset: 3
41	12	3.42	3.48	.86	.38	1.09	.3	1.06	.2	.88	.71	.53	25.5	28.1	73	73	in subset: 1
41	12	3.42	3.48	.86	.38	.64	-.9	.64	-.9	1.45	.76	.53	28.8	28.1	148	148	in subset: 1
41	12	3.42	3.48	.86	.38	.94	.0	.93	.0	1.07	.52	.53	32.5	28.3	41	41	in subset: 2
43	12	3.58	3.56	.99	.39	.17	-3.3	.16	-3.4	1.96	.78	.52	37.9	28.9	157	157	in subset: 3
42	12	3.50	3.57	1.01	.38	.78	-.4	.78	-.4	1.25	.50	.53	25.4	26.9	60	60	in subset: 1
42	12	3.50	3.57	1.01	.38	.85	-.2	.85	-.2	1.16	.61	.53	27.5	26.9	123	123	in subset: 1
42	12	3.50	3.57	1.01	.38	.41	-1.8	.42	-1.8	1.70	.82	.53	28.0	27.0	95	95	in subset: 2
42	12	3.50	3.57	1.01	.38	.96	.0	.95	.0	1.07	.68	.53	29.9	27.0	149	149	in subset: 2
42	12	3.50	3.57	1.01	.38	1.27	.7	1.24	.7	.74	.62	.53	22.6	27.0	177	177	in subset: 2
44	12	3.67	3.62	1.10	.39	1.10	.3	1.05	.2	.92	.58	.52	31.6	27.6	175	175	in subset: 4
44	12	3.67	3.64	1.14	.39	1.00	.1	.98	.0	1.01	.66	.52	33.1	27.7	21	21	in subset: 3
44	12	3.67	3.64	1.14	.39	1.39	1.0	1.31	.8	.55	.68	.52	25.7	27.7	31	31	in subset: 3
44	12	3.67	3.64	1.14	.39	2.04	2.2	1.98	2.1	-.23	.44	.52	29.5	27.7	187	187	in subset: 3
43	12	3.58	3.65	1.16	.39	.97	.0	.96	.0	1.04	.34	.52	23.5	25.6	179	179	in subset: 2
45	12	3.75	3.73	1.29	.39	.43	-1.7	.42	-1.8	1.66	.59	.52	30.6	26.3	109	109	in subset: 3
46	12	3.83	3.82	1.45	.40	3.03	3.5	2.98	3.5	-1.36	.49	.51	11.8	24.9	158	158	in subset: 3
45	12	3.75	3.82	1.46	.39	.78	-.4	.82	-.3	1.19	.19	.52	24.3	22.7	152	152	in subset: 1
47	12	3.92	3.90	1.61	.41	1.45	1.1	1.40	1.0	.62	.77	.51	19.8	23.5	97	97	in subset: 3
49	12	4.08	4.06	1.92	.42	2.54	2.9	2.42	2.7	-.67	.40	.49	14.9	19.9	171	171	in subset: 4
49	12	4.08	4.16	2.12	.42	1.86	1.8	1.71	1.6	.21	.63	.49	13.1	16.7	22	22	in subset: 1
34.0	12.0	2.83	2.83	-.26	.39	1.00	-.2	1.00	-.2		.50				Mean (Count: 190)		
5.6	.0	.47	.47	.84	.02	.70	1.4	.69	1.4		.24				S.D. (Population)		
5.6	.0	.47	.48	.84	.02	.70	1.4	.70	1.4		.24				S.D. (Sample)		

Model, Populn: RMSE .39 Adj (True) S.D. .75 Separation 1.91 Strata 2.88 Reliability (not inter-rater) .78
Model, Sample: RMSE .39 Adj (True) S.D. .75 Separation 1.92 Strata 2.89 Reliability (not inter-rater) .79
Model, Fixed (all same) chi-square: 810.7 d.f.: 189 significance (probability): .00
Model, Random (normal) chi-square: 157.1 d.f.: 188 significance (probability): .95
Inter-Rater agreement opportunities: 61668 Exact agreements: 20436 = 33.1% Expected: 19591.2 = 31.8%

Rasch Analysis of Expert Ratings 22/12/2014 16:49:51
Table 7.2.1 Arabic English Familiarity Level Measurement Report (arranged by MN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	N Arabic English Familiarity Level	
1314	480	2.74	2.76	-.10	.06	1.19	2.9	1.19	2.8	.77	.60	.64	1 No Familiarity	in subset: 2
2040	744	2.74	2.76	-.09	.05	1.03	.5	1.03	.5	.96	.64	.65	2 Limited Familiarity	in subset: 1
2388	816	2.93	2.86	.08	.05	.86	-3.1	.86	-3.1	1.16	.68	.64	3 Some Familiarity	in subset: 3
711	240	2.96	2.88	.11	.09	.98	-.2	.99	-.1	1.04	.60	.63	4 Very Familiar	in subset: 4
1613.3	570.0	2.84	2.81	.00	.06	1.01	.0	1.01	.0		.63		Mean (Count: 4)	
649.2	227.9	.10	.06	.09	.02	.12	2.2	.12	2.1		.04		S.D. (Population)	
749.7	263.2	.12	.06	.11	.02	.14	2.5	.14	2.5		.04		S.D. (Sample)	

Model, Populn: RMSE .06 Adj (True) S.D. .07 Separation 1.11 Strata 1.81 Reliability .55
Model, Sample: RMSE .06 Adj (True) S.D. .09 Separation 1.40 Strata 2.21 Reliability .66
Model, Fixed (all same) chi-square: 10.0 d.f.: 3 significance (probability): .02
Model, Random (normal) chi-square: 2.2 d.f.: 2 significance (probability): .33

Rasch Analysis of Expert Ratings 22/12/2014 16:49:51
Table 7.3.1 Speaker Measurement Report (arranged by MN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	Nu Speaker
424	190	2.23	2.21	-1.10	.10	.74	-2.7	.74	-2.8	1.30	.52	.51	4 Speaker 5 Spanish-English Male 1
436	190	2.29	2.27	-.98	.10	1.35	3.2	1.37	3.3	.58	.35	.52	2 Speaker 3 Arabic-English Female 1
471	190	2.48	2.46	-.62	.10	1.00	.0	1.00	.0	.97	.44	.53	8 Speaker 9 Spanish-English Female 1
479	190	2.52	2.50	-.54	.10	1.30	2.7	1.29	2.7	.68	.40	.53	1 Speaker 2 Dhivehi-English Female 1
490	190	2.58	2.56	-.44	.10	.82	-1.8	.83	-1.8	1.18	.53	.53	6 Speaker 7 Spanish-English Male 2
505	190	2.66	2.64	-.29	.10	.85	-1.6	.84	-1.6	1.19	.51	.54	3 Speaker 4 Dhivehi-English Female 2
530	190	2.79	2.78	-.06	.10	1.56	4.8	1.55	4.8	.38	.59	.54	7 Speaker 8 Dhivehi-English Male 1
590	190	3.11	3.11	.50	.10	.84	-1.7	.83	-1.8	1.19	.65	.54	5 Speaker 6 Arabic-English Male 1
602	190	3.17	3.18	.61	.10	1.02	.2	1.02	.2	.99	.63	.54	11 Speaker 12 Arabic-English Female 2
618	190	3.25	3.26	.75	.10	.79	-2.3	.79	-2.2	1.24	.64	.54	12 Speaker 13 Dhivehi-English Male 2
654	190	3.44	3.46	1.09	.10	.78	-2.4	.77	-2.5	1.25	.60	.54	9 Speaker 10 Arabic-English 2
654	190	3.44	3.46	1.09	.10	.92	-.7	.91	-.8	1.10	.56	.54	10 Speaker 11 Spanish-English Female 2
537.8	190.0	2.83	2.82	.00	.10	1.00	-.2	1.00	-.2		.53		Mean (Count: 12)
79.0	.0	.42	.43	.75	.00	.25	2.4	.25	2.4		.09		S.D. (Population)
82.5	.0	.43	.45	.78	.00	.27	2.5	.27	2.6		.10		S.D. (Sample)

Model, Populn: RMSE .10 Adj (True) S.D. .74 Separation 7.54 Strata 10.39 Reliability .98
Model, Sample: RMSE .10 Adj (True) S.D. .77 Separation 7.88 Strata 10.84 Reliability .98
Model, Fixed (all same) chi-square: 685.6 d.f.: 11 significance (probability): .00
Model, Random (normal) chi-square: 10.8 d.f.: 10 significance (probability): .37

Rasch Analysis of Expert Ratings 22/12/2014 16:49:51
Table 8.1 Category Statistics.

Model = ?,?,?,R5

Score	DATA Category Counts Total	Used	%	Cum. %	QUALITY CONTROL Avge Meas	Exp. Meas	OUTFIT MnSq	RASCH-ANDRICH Thresholds Measure	S.E.	EXPECTATION Measure at Category	-0.5	MOST PROBABLE from	RASCH-THURSTONE Thresholds	Cat PEAK Prob
1	166	166	7%	7%	-1.48	-1.67	1.2			(-3.90)		low	low	100%
2	709	709	31%	38%	-.94	-.90	1.0	-2.74	.09	-1.78	-2.99	-2.74	-2.85	57%
3	842	842	37%	75%	-.19	-.14	1.0	-.69	.05	.05	-.81	-.69	-.76	50%
4	472	472	21%	96%	.69	.61	.9	.82	.06	1.80	.89	.82	.85	54%
5	91	91	4%	100%	1.33	1.31	1.0	2.61	.12	(3.80)	2.93	2.61	2.74	100%
										(Mean)		(Modal)	(Median)	

Scale structure

```
Measr:-4.0             -2.0              0.0              2.0              4.0
          +               +                 +                +                +
  Mode:<(^)-------12-------^-------23-----^-----34-------^-----45--------(^)5>

Median:<(^)------12--------^-------23-----^-----34-------^------45-------(^)5>

  Mean:<(^)-----12---------^------23------^------34------^--------45-----(^)5>
          +               +                 +                +                +
Measr:-4.0             -2.0              0.0              2.0              4.0
```

Probability Curves

```
       -4.0             -2.0              0.0              2.0              4.0
        ++----------------+----------------+----------------+-----------------++
      1 |                                                                      |
        |                                                                      |
        |                                                                      |
        |                                                                      |
        |                                                                     5|
      P |111                                                               555 |
      r |   11                                                            5    |
      o |     1                                                         55     |
      b |      11                                                     55       |
      a |        1     2222222222                        444444      5         |
      b |         11222          222    3333333       444      444 55          |
      i |         2211              2233       333 44             *4           |
      l |       22    1             332          4*3            55  44         |
      i |     22       11         33   22      44    33        5      444      |
      t |   22           1      33       22  44        33    55          44    |
      y | 22              11  33           24            3355              44  |
        |2                  **            4422           5533                44|
        |                333  111      444    222      55    333               |
        |             333        111444          22*55          333            |
        |      3333333         44444111111   555555 22222         333333       |
      0 |**********************55555555555***111111111111**********************|
        ++----------------+----------------+----------------+-----------------++
       -4.0             -2.0              0.0              2.0              4.0
```

Expected Score Ogive (Model ICC)

```
       -4.0             -2.0              0.0              2.0              4.0
        ++----------------+----------------+----------------+-----------------++
      5 |                                                                      |
        |                                                                555555|
        |                                                          445555      |
        |                                                     44444            |
      4 |                                                4444                  |
        |                                            4444                      |
        |                                        3344                          |
        |                                     333                              |
      3 |                                 3333                                 |
        |                             3333                                     |
        |                          223                                         |
        |                      2222                                            |
      2 |                 22222                                                |
        |            22222                                                     |
        |     1111222                                                          |
        |11111                                                                 |
      1 |                                                                      |
        ++----------------+----------------+----------------+-----------------++
       -4.0             -2.0              0.0              2.0              4.0
```

Rasch Analysis of Expert Ratings 22/12/2014 16:49:51

Table 4.1 Unexpected Responses (9 residuals sorted by u).

Cat	Score	Exp.	Resd	StRes	Num	rat	N	Arabic English Fami	Nu	Speaker
5	5	2.4	2.6	3.5	118	118	4	Very Familiar	7	Speaker 8 Dhivehi-English Male 1
1	1	3.6	-2.6	-3.4	158	158	3	Some Familiarity	1	Speaker 2 Dhivehi-English Female 1
5	5	2.5	2.5	3.3	188	188	1	No Familiarity	1	Speaker 2 Dhivehi-English Female 1
5	5	2.5	2.5	3.3	189	189	2	Limited Familiarity	1	Speaker 2 Dhivehi-English Female 1
2	2	4.2	-2.2	-3.2	158	158	3	Some Familiarity	11	Speaker 12 Arabic-English Female 2
1	1	3.5	-2.5	-3.2	189	189	2	Limited Familiarity	9	Speaker 10 Arabic-English 2
1	1	3.5	-2.5	-3.2	189	189	2	Limited Familiarity	10	Speaker 11 Spanish-English Female 2
5	5	2.7	2.3	3.1	188	188	1	No Familiarity	3	Speaker 4 Dhivehi-English Female 2
5	5	2.7	2.3	3.1	189	189	2	Limited Familiarity	3	Speaker 4 Dhivehi-English Female 2
Cat	Score	Exp.	Resd	StRes	Num	rat	N	Arabic English Fami	Nu	Speaker

The Facets analyses output of the Dhivehi-English pronunciation score data

```
Facets (Many-Facet Rasch Measurement) Version No. 3.71.1  Copyright �(c) 1987-2013, John M. Linacre. All rights reserved.
22/12/2014 16:45:26

Rasch Analysis of Expert Ratings 22/12/2014 16:45:26
Table 1. Specifications from file "C:\Users\kevin\Desktop\Main Study Data\Dhivehi English Pronunciaton Operating File.txt.txt".

Title = Rasch Analysis of Expert Ratings 22/12/2014 16:45:26
Data file = "C:\Users\kevin\Desktop\Main Study Data\Pronunciation Score Data Dhivehi Familiarity.xlsx"
Output file = C:\Users\kevin\Desktop\Main Study Data\Dhivehi English Pronunciaton Operating File.txt.out.txt

; Data specification
Facets = 3
Delements = N
Non-centered = 1
Positive =  1, 2, 3
Labels =
 1,rater ; (elements = 190)
 2,Dhivehi English Familiarity Level ; (elements = 4)
 3,Speaker ; (elements = 12)
Model = ?,?,?,R5,1

; Output description
Arrange tables in order = MN
Bias/Interaction direction = plus ; ability, easiness, leniency: higher score = positive logit
Fair score = Mean
Pt-biserial = Measure
Heading lines in output data files = Y
Inter-rater coefficients reported for facet = 1
Barchart = Yes
Total score for elements = Yes
T3onscreen show only one line on screen iteration report = Y
T4MAX maximum number of unexpected observations reported in Table 4 = 100
T8NBC show table 8 numbers-barcharts-curves = NBC
Unexpected observations reported if standardized residual >= 3
Usort unexpected observations sort order = u
WHexact - Wilson-Hilferty standardization = Y

; Convergence control
Convergence = .1, .01
Iterations (maximum) = 0 ; unlimited
Xtreme scores adjusted by = .3, .5  ;(estimation, bias)

Rasch Analysis of Expert Ratings 22/12/2014 16:45:26
Table 2. Data Summary Report.

Assigning models to Data= "C:\Users\kevin\Desktop\Main Study Data\Pronunciation Score Data Dhivehi Familiarity.xlsx"
Total lines in data file = 194
Total data lines = 190
Responses matched to model: ?,?,?,R5,1 = 2280
    Total non-blank responses found = 2280
              Number of blank lines = 3
Valid responses used for estimation = 2280

Rasch Analysis of Expert Ratings 22/12/2014 16:45:26
Table 3. Iteration Report.
```

Iteration		Max. Score Residual Elements	%	Categories	Max. Logit Change Elements	Steps
PROX	1				-1.6094	
PROX	2				-.5005	
JMLE	3	-106.4094	-33.4	516.4205	.9815	-2.0589
JMLE	4	-40.6022	-10.9	-22.6258	-.2909	-.0921
JMLE	5	-57.7551	-5.4	33.2032	-.1741	-.1391
JMLE	6	-60.2012	-4.4	21.2010	-.1445	-.0878
JMLE	7	-50.6624	-3.4	18.0228	-.1141	-.0750
JMLE	8	-42.0913	-2.7	15.0052	-.0925	-.0623
JMLE	9	-35.1638	-2.2	12.5693	-.0760	-.0527
JMLE	10	-29.6724	-1.9	10.6248	-.0633	-.0449
JMLE	11	-25.2715	-1.6	9.0594	-.0533	-.0385
JMLE	12	-21.6917	-1.4	7.7817	-.0452	-.0332
JMLE	13	-18.7373	-1.2	6.7246	-.0387	-.0288
JMLE	14	-16.2676	-1.0	5.8397	-.0333	-.0250
JMLE	15	-14.1819	-.9	5.0914	-.0288	-.0218
JMLE	16	-12.4053	-.8	4.4537	-.0250	-.0191
JMLE	17	-10.8813	-.7	3.9063	-.0218	-.0168
JMLE	18	-9.5659	-.6	3.4340	-.0191	-.0147
JMLE	19	-8.4255	-.5	3.0243	-.0167	-.0130
JMLE	20	-7.4331	-.5	-2.6685	-.0147	-.0114
JMLE	21	-6.5662	-.4	-2.3584	-.0129	-.0101
JMLE	22	-5.8069	-.4	-2.0865	-.0114	-.0089
JMLE	23	-5.1404	-.3	-1.8476	-.0101	-.0079
JMLE	24	-4.5542	-.3	-1.6374	-.0089	-.0070
JMLE	25	-4.0378	-.2	-1.4520	-.0079	-.0062
JMLE	26	-3.5820	-.2	-1.2884	-.0070	-.0055
JMLE	27	-3.1794	-.2	-1.1438	-.0062	-.0049
JMLE	28	-2.8235	-.2	-1.0158	-.0055	-.0043
JMLE	29	-2.5083	-.2	-.9026	-.0049	-.0038
JMLE	30	-2.2292	-.1	-.8022	-.0043	-.0034
JMLE	31	-1.9817	-.1	-.7132	-.0038	-.0030
JMLE	32	-1.7622	-.1	-.6342	-.0034	-.0027
JMLE	33	-1.5674	-.1	-.5642	-.0030	-.0024
JMLE	34	-1.3943	-.1	-.5019	-.0027	-.0021
JMLE	35	-1.2407	-.1	-.4467	-.0024	-.0019
JMLE	36	-1.1040	-.1	-.3974	-.0021	-.0017
JMLE	37	-.9827	-.1	-.3538	-.0019	-.0015
JMLE	38	-.8748	-.1	-.3150	-.0017	-.0013
JMLE	39	-.7788	.0	-.2804	-.0015	-.0012
JMLE	40	-.6934	.0	-.2497	-.0013	-.0011
JMLE	41	-.6174	.0	-.2223	-.0012	-.0009
JMLE	42	-.5498	.0	-.1980	-.0011	-.0008
JMLE	43	-.4897	.0	-.1763	-.0009	-.0007
JMLE	44	-.4360	.0	-.1570	-.0008	-.0007
JMLE	45	-.3884	.0	-.1399	-.0007	-.0006

```
| JMLE  46      -.3459    .0      -.1246     -.0007   -.0005 |
| JMLE  47      -.3083    .0      -.1110     -.0006   -.0005 |
| JMLE  48      -.2747    .0      -.0989     -.0005   -.0004 |
| JMLE  49      -.2446    .0      -.0881     -.0005   -.0004 |
| JMLE  50      -.2180    .0      -.0785     -.0004   -.0003 |
| JMLE  51      -.1941    .0      -.0699     -.0004   -.0003 |
| JMLE  52      -.1729    .0      -.0623     -.0003   -.0003 |
| JMLE  53      -.1541    .0      -.0555     -.0003   -.0002 |
| JMLE  54      -.1372    .0      -.0494     -.0003   -.0002 |
| JMLE  55      -.1223    .0      -.0440     -.0002   -.0002 |
| JMLE  56      -.1089    .0      -.0392     -.0002   -.0002 |
| JMLE  57      -.0972    .0      -.0349     -.0002   -.0001 |
+------------------------------------------------------------+
Warning (6)! There may be 4 disjoint subsets


Rasch Analysis of Expert Ratings 22/12/2014 16:45:26
Table 4. Unexpected Responses - appears after Table 8.


Rasch Analysis of Expert Ratings 22/12/2014 16:45:26
Table 5. Measurable Data Summary.

+---------------------------------------------------+
| Cat  Score  Exp.  Resd StRes|                     |
|-----------------------------+---------------------|
| 2.83  2.83  2.83   .00   .00 | Mean (Count: 2280) |
|  .97   .97   .63   .74 1.00 | S.D. (Population)  |
|  .97   .97   .63   .74 1.00 | S.D. (Sample)      |
+---------------------------------------------------+
Data log-likelihood chi-square = 5051.5640
Approximate degrees of freedom = 2073
Chi-square significance prob.  = .0000
                                        Count   Mean   S.D.   Params
Responses used for estimation      =     2280   2.83   0.97      207
Count of measurable responses      =     2280
Raw-score variance of observations   =   0.94 100.00%
Variance explained by Rasch measures =   0.39  42.00%
Variance of residuals                =   0.55  58.00%

Rasch Analysis of Expert Ratings 22/12/2014 16:45:26
Table 6.0  All Facet Vertical "Rulers".

Vertical = (1*,2A,3A,S) Yardstick (columns lines low high extreme)= 0,8,-3,3,End
+---------------------------------------------------------------------------------------------------------------------------------------+
|Measr|+rater      |+Dhivehi English Familiarity Level|+Speaker                                                                  |Scale|
|-----+------------+----------------------------------+--------------------------------------------------------------------------+-----|
|   3 +            +                                  +                                                                          + (5) |
|     |            |                                  |                                                                          | --- |
|     |            |                                  |                                                                          |     |
|     |            |                                  |                                                                          |     |
|     |            |                                  |                                                                          |     |
|     |            |                                  |                                                                          |     |
|     |            |                                  |                                                                          |     |
|     |            |                                  |                                                                          |     |
|   2 + *          +                                  +                                                                          +     |
|     |            |                                  |                                                                          |     |
|     |            |                                  |                                                                          |  4  |
|     | .          |                                  |                                                                          |     |
|     | .          |                                  |                                                                          |     |
|     | *          |                                  |                                                                          |     |
|     | *          |                                  |                                                                          |     |
|     |            |                                  | Speaker 10 Arabic-English 2          Speaker 11 Spanish-English Female 2 |     |
|   1 + *          +                                  +                                                                          +     |
|     | ****       |                                  |                                                                          | --- |
|     | **         |                                  | Speaker 13 Dhivehi-English Male 2                                        |     |
|     | ***        |                                  | Speaker 12 Arabic-English Female 2                                       |     |
|     | **.        |                                  | Speaker 6 Arabic-English Male 1                                          |     |
|     | *******    |                                  |                                                                          |     |
|     | *          | Very Familiar                    |                                                                          |     |
|     | ****       | Limited Familiarity              |                                                                          |     |
*   0 * ******     * No Familiarity                   * Speaker 8 Dhivehi-English Male 1                                         *  3  *
|     | *********  |                                  |                                                                          |     |
|     | *********. |                                  | Speaker 4 Dhivehi-English Female 2                                       |     |
|     | *****      | Some Familiarity                 |                                                                          |     |
|     | *****      |                                  | Speaker 2 Dhivehi-English Female 1   Speaker 7 Spanish-English Male 2    |     |
|     | *          |                                  | Speaker 9 Spanish-English Female 1                                       |     |
|     | ****       |                                  |                                                                          | --- |
|     | ******.    |                                  |                                                                          |     |
|  -1 + *****.     +                                  + Speaker 3 Arabic-English Female 1                                        +     |
|     | ***        |                                  | Speaker 5 Spanish-English Male 1                                         |     |
|     | *          |                                  |                                                                          |     |
|     | **         |                                  |                                                                          |     |
|     | ***        |                                  |                                                                          |     |
|     | *          |                                  |                                                                          |     |
|     |            |                                  |                                                                          |  2  |
|     | *          |                                  |                                                                          |     |
|  -2 + *          +                                  +                                                                          +     |
|     |            |                                  |                                                                          |     |
|     | **         |                                  |                                                                          |     |
|     |            |                                  |                                                                          |     |
|     |            |                                  |                                                                          |     |
|     | *          |                                  |                                                                          |     |
|     |            |                                  |                                                                          |     |
|     |            |                                  |                                                                          |     |
|  -3 +            +                                  +                                                                          + (1) |
|-----+------------+----------------------------------+--------------------------------------------------------------------------+-----|
|Measr| * = 2      |+Dhivehi English Familiarity Level|+Speaker                                                                  |Scale|
+---------------------------------------------------------------------------------------------------------------------------------------+

S.1: Model = ?,?,?,R5
There are 4 disconnected subsets identified in Table 7.

Rasch Analysis of Expert Ratings 22/12/2014 16:45:26
Table 6.1  rater Facet Summary.

Logit:
                    11 1  11 11  1
    2    4 22 2 6 65113 08209 822813551482 2 21 1  2
 +---------Q-------S-+------M--+----S----+---Q-----+---------+
-3        -2        -1         0         1         2         3

Infit MnSq:
     112111111 1
   4467524553273431 321 1221   1                        1  1
 +--S------M------S--+---Q-----+---------+---------+---------+
 0         1         2         3         4         5         6

Outfit MnSq:
     122111111 1
   4430617090292432 221 221 1  1                        1 1
 +--S------M------S--+---Q-----+---------+---------+---------+
 0         1         2         3         4         5         6

Infit ZStd:
               11  1 1 11 1
    121  22588854790669216273112 31 32  1            2
 +----Q----+-S--+---M+----+S---+---Q+----+----+----+----+----+
-4   -3   -2   -1    0    1    2    3    4    5    6    7    8    9
```

```
Outfit ZStd:
             1 11    11 1 1
    121 1133718547995827182533 221231  1           11
 +----Q----+-S--+---M+----+S---+--Q-+----+----+----+----+----+----+
-4   -3   -2   -1    0    1    2    3    4    5    6    7    8    9
```

Rasch Analysis of Expert Ratings 22/12/2014 16:45:26
Table 6.2 Dhivehi English Familiarity Level Facet Summary.

```
Logit:
                          1   11 1
 +---------+---------+----Q-S--M--S-Q----+---------+---------+
-3        -2        -1         0         1         2         3

Infit MnSq:
     1   11                 1
 +--S------+-M-------+-S-------+Q--------+---------+---------+
 0         1         2         3         4         5         6

Outfit MnSq:
     1   11                 1
 +--S------+-M-------+-S-------+Q--------+---------+---------+
 0         1         2         3         4         5         6

Infit ZStd:
        1  11                                                  1
 S----+----+----+----+--M-+----+----+----+----+S---+----+----+----+
-4   -3   -2   -1    0    1    2    3    4    5    6    7    8    9

Outfit ZStd:
        1   11                                                1
 S----+----+----+----+--M-+----+----+----+----+S---+----+----+----+
-4   -3   -2   -1    0    1    2    3    4    5    6    7    8    9
```

Rasch Analysis of Expert Ratings 22/12/2014 16:45:26
Table 6.3 Speaker Facet Summary.

```
Logit:
                    11   1111 1     11 1  2
 +---------+----Q----+--S------M------S--+----Q----+---------+
-3        -2        -1         0         1         2         3

Infit MnSq:
        1512  11 1
 +----Q-S--M--S-Q----+---------+---------+---------+---------+
 0         1         2         3         4         5         6

Outfit MnSq:
        1512  11 1
 +----Q-S--M--S-Q----+---------+---------+---------+---------+
 0         1         2         3         4         5         6

Infit ZStd:
       1 2  21   1   11             1 1       1
 +----+-S--+----+---M+----+----+S---+----+--Q-+----+----+----+----+
-4   -3   -2   -1    0    1    2    3    4    5    6    7    8    9

Outfit ZStd:
       11 1 21   1   11             1  1      1
 +----+-S--+----+---M+----+----+S---+----+--Q-+----+----+----+----+
-4   -3   -2   -1    0    1    2    3    4    5    6    7    8    9
```

Rasch Analysis of Expert Ratings 22/12/2014 16:45:26
Table 7.1.1 rater Measurement Report (arranged by MN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	Exact Agree. Obs %	Exact Agree. Exp %	Num rater	
20	12	1.67	1.63	-2.66	.47	.57	-1.1	.59	-1.0	1.51	.71	.46	19.4	21.3	24 24	in subset: 1
20	12	1.67	1.63	-2.66	.47	.37	-1.9	.43	-1.7	1.70	.63	.46	22.1	21.3	138 138	in subset: 1
22	12	1.83	1.80	-2.24	.44	2.08	2.2	2.10	2.3	-.34	.08	.48	19.8	24.6	72 72	in subset: 1
22	12	1.83	1.80	-2.24	.44	.31	-2.3	.35	-2.1	1.79	.49	.48	24.6	24.6	145 145	in subset: 1
22	12	1.83	1.80	-2.24	.44	1.00	.1	1.03	.1	.97	.35	.48	26.1	24.6	162 162	in subset: 1
22	12	1.83	1.80	-2.24	.44	1.10	.3	1.18	.5	.80	.23	.48	27.3	24.6	183 183	in subset: 1
23	12	1.92	1.88	-2.04	.43	2.26	2.4	2.30	2.6	-.69	-.60	.48	26.0	26.2	75 75	in subset: 1
24	12	2.00	1.92	-1.95	.43	.80	-.4	.77	-.5	1.21	.52	.49	25.5	24.5	147 147	in subset: 2
24	12	2.00	1.97	-1.86	.43	1.08	.3	1.04	.2	.97	.70	.49	29.0	27.6	40 40	in subset: 1
24	12	2.00	1.97	-1.86	.43	.44	-1.6	.45	-1.6	1.65	.63	.49	28.6	27.6	65 65	in subset: 1
25	12	2.08	2.05	-1.68	.42	1.21	.6	1.21	.6	.67	-.09	.50	30.7	28.9	85 85	in subset: 1
25	12	2.08	2.05	-1.68	.42	.87	-.2	.94	.0	1.10	.53	.50	28.3	28.9	111 111	in subset: 1
26	12	2.17	2.13	-1.51	.41	1.56	1.3	1.65	1.5	.21	-.01	.50	27.8	30.1	5 5	in subset: 1
26	12	2.17	2.13	-1.51	.41	.93	.0	.94	.0	1.03	.30	.50	29.3	30.1	53 53	in subset: 1
26	12	2.17	2.13	-1.51	.41	1.19	.5	1.20	.6	.71	.33	.50	31.8	30.1	69 69	in subset: 1
26	12	2.17	2.13	-1.51	.41	1.25	.7	1.24	.6	.72	.67	.50	30.6	30.1	108 108	in subset: 1
26	12	2.17	2.13	-1.51	.41	1.04	.2	1.03	.1	.97	.64	.50	31.5	30.1	151 151	in subset: 1
26	12	2.17	2.13	-1.51	.41	.47	-1.5	.47	-1.5	1.61	.16	.50	32.1	30.1	180 180	in subset: 1
27	12	2.25	2.22	-1.34	.41	.18	-3.1	.17	-3.2	1.94	.72	.51	36.8	31.2	33 33	in subset: 1
27	12	2.25	2.22	-1.34	.41	.94	.0	.98	.0	1.01	.04	.51	31.3	31.2	105 105	in subset: 1
27	12	2.25	2.22	-1.34	.41	1.07	.3	1.13	.4	.91	.45	.51	30.0	31.2	165 165	in subset: 1
27	12	2.25	2.22	-1.34	.41	1.39	1.0	1.35	.9	.60	.75	.51	28.9	31.2	167 167	in subset: 1
28	12	2.33	2.25	-1.27	.40	.66	-.8	.64	-.9	1.41	.62	.51	32.8	30.2	99 99	in subset: 2
28	12	2.33	2.25	-1.27	.40	.92	.0	.97	.0	.98	.60	.51	31.3	30.2	142 142	in subset: 2
28	12	2.33	2.30	-1.18	.40	1.13	.4	1.13	.4	.82	.42	.51	33.5	32.1	13 13	in subset: 1
28	12	2.33	2.30	-1.18	.40	.50	-1.4	.49	-1.4	1.52	.57	.51	37.9	32.1	14 14	in subset: 1
28	12	2.33	2.30	-1.18	.40	.63	-.9	.63	-.9	1.40	.46	.51	35.0	32.1	81 81	in subset: 1
28	12	2.33	2.30	-1.18	.40	.96	.0	1.00	.1	.95	.35	.51	33.6	32.1	136 136	in subset: 1
28	12	2.33	2.30	-1.18	.40	1.07	.3	1.02	.1	1.03	.49	.51	30.2	32.1	181 181	in subset: 1
29	12	2.42	2.34	-1.12	.40	.65	-.8	.68	-.7	1.34	.45	.52	33.3	31.3	43 43	in subset: 2
29	12	2.42	2.39	-1.02	.40	.66	-.8	.69	-.7	1.35	.80	.52	36.7	32.8	1 1	in subset: 1
29	12	2.42	2.39	-1.02	.40	.28	-2.4	.26	-2.6	1.85	.81	.52	39.6	32.8	39 39	in subset: 1
29	12	2.42	2.39	-1.02	.40	1.32	.8	1.25	.7	.76	.08	.52	30.7	32.8	45 45	in subset: 1
29	12	2.42	2.39	-1.02	.40	1.73	1.6	1.72	1.6	.18	.40	.52	29.3	32.8	64 64	in subset: 1
29	12	2.42	2.39	-1.02	.40	.82	-.3	.83	-.3	1.13	.25	.52	37.8	32.8	76 76	in subset: 1
29	12	2.42	2.39	-1.02	.40	.33	-2.2	.30	-2.3	1.80	.76	.52	38.5	32.8	88 88	in subset: 1
29	12	2.42	2.39	-1.02	.40	1.07	.3	1.14	.4	.88	.48	.52	32.3	32.8	93 93	in subset: 1
29	12	2.42	2.39	-1.02	.40	1.48	1.1	1.47	1.1	.34	-.08	.52	34.7	32.8	121 121	in subset: 1
29	12	2.42	2.39	-1.02	.40	.86	-.2	.85	-.2	1.19	.80	.52	31.7	32.8	135 135	in subset: 1
29	12	2.42	2.39	-1.02	.40	.73	-.6	.77	-.5	1.23	.35	.52	35.6	32.8	164 164	in subset: 1
29	12	2.42	2.39	-1.02	.40	.87	-.1	.85	-.2	1.14	.44	.52	34.7	32.8	184 184	in subset: 1
30	12	2.50	2.47	-.87	.39	1.39	1.0	1.35	.9	.55	.42	.52	33.9	33.4	3 3	in subset: 1
30	12	2.50	2.47	-.87	.39	1.42	1.0	1.42	1.0	.51	-.02	.52	32.5	33.4	4 4	in subset: 1
30	12	2.50	2.47	-.87	.39	.18	-3.1	.19	-3.1	1.89	.78	.52	40.2	33.4	28 28	in subset: 1
30	12	2.50	2.47	-.87	.39	.54	-1.2	.56	-1.2	1.48	.30	.52	39.1	33.4	35 35	in subset: 1
30	12	2.50	2.47	-.87	.39	1.21	.6	1.27	.7	.69	.36	.52	30.6	33.4	44 44	in subset: 1
30	12	2.50	2.47	-.87	.39	.56	-1.2	.53	-1.3	1.52	.53	.52	37.6	33.4	68 68	in subset: 1
30	12	2.50	2.47	-.87	.39	.19	-3.0	.20	-3.0	1.87	.76	.52	40.8	33.4	82 82	in subset: 1
30	12	2.50	2.47	-.87	.39	.43	-1.7	.46	-1.5	1.59	.43	.52	39.2	33.4	83 83	in subset: 1
30	12	2.50	2.47	-.87	.39	.79	-.4	.78	-.4	1.27	.68	.52	36.5	33.4	94 94	in subset: 1
30	12	2.50	2.47	-.87	.39	1.61	1.4	1.64	1.4	.33	.46	.52	33.2	33.4	118 118	in subset: 1
30	12	2.50	2.47	-.87	.39	1.35	.9	1.36	.9	.62	.76	.52	33.4	33.4	155 155	in subset: 1
30	12	2.50	2.47	-.87	.39	1.02	.1	.96	.0	1.02	.52	.52	37.4	33.4	168 168	in subset: 1
30	12	2.50	2.47	-.87	.39	.59	-1.0	.62	-.9	1.44	.69	.52	35.7	33.4	185 185	in subset: 1
28	12	2.33	2.54	-.75	.40	.42	-1.7	.45	-1.6	1.63	.40	.51	.0	.0	125 125	in subset: 3
31	12	2.58	2.56	-.71	.39	.83	-.3	.82	-.3	1.21	.65	.52	36.7	33.8	15 15	in subset: 1
31	12	2.58	2.56	-.71	.39	.71	-.6	.71	-.6	1.29	.07	.52	35.8	33.8	26 26	in subset: 1
31	12	2.58	2.56	-.71	.39	1.23	.6	1.24	.6	.82	.80	.52	33.1	33.8	56 56	in subset: 1

31	12	2.58	2.56	-.71	.39	.88	-.1	.95	.0	1.13	.18	.52	35.4	33.8	63	63	in subset: 1
31	12	2.58	2.56	-.71	.39	1.47	1.1	1.45	1.1	.45	.36	.52	30.7	33.8	112	112	in subset: 1
31	12	2.58	2.56	-.71	.39	1.16	.5	1.19	.5	.82	.70	.52	32.5	33.8	144	144	in subset: 1
31	12	2.58	2.56	-.71	.39	.62	-.9	.63	-.9	1.38	.65	.52	37.5	33.8	146	146	in subset: 1
33	12	2.75	2.57	-.68	.38	2.58	2.9	2.65	3.0	-.75	-.08	.53	25.0	32.8	186	186	in subset: 4
32	12	2.67	2.59	-.66	.39	.44	-1.6	.43	-1.7	1.61	.61	.53	39.1	33.6	87	87	in subset: 2
32	12	2.67	2.64	-.56	.39	.61	-1.0	.59	-1.0	1.44	.44	.53	37.3	34.0	2	2	in subset: 1
32	12	2.67	2.64	-.56	.39	.95	.0	.92	.0	1.11	.66	.53	36.4	34.0	52	52	in subset: 1
32	12	2.67	2.64	-.56	.39	.94	.0	.98	.0	1.02	.69	.53	34.0	34.0	55	55	in subset: 1
32	12	2.67	2.64	-.56	.39	.55	-1.2	.55	-1.2	1.52	.67	.53	39.0	34.0	57	57	in subset: 1
32	12	2.67	2.64	-.56	.39	.54	-1.2	.55	-1.2	1.45	.25	.53	35.8	34.0	66	66	in subset: 1
32	12	2.67	2.64	-.56	.39	1.10	.3	1.09	.3	.91	.84	.53	33.5	34.0	140	140	in subset: 1
32	12	2.67	2.64	-.56	.39	.75	-.5	.74	-.5	1.27	.29	.53	38.5	34.0	154	154	in subset: 1
32	12	2.67	2.64	-.56	.39	2.40	2.7	2.31	2.5	-.57	.08	.53	28.8	34.0	173	173	in subset: 1
34	12	2.83	2.66	-.54	.38	5.86	6.2	5.81	6.2	-4.42	.22	.53	26.4	33.5	188	188	in subset: 4
34	12	2.83	2.66	-.54	.38	5.62	6.1	5.55	6.0	-4.13	-.01	.53	33.3	33.5	189	189	in subset: 4
33	12	2.75	2.73	-.41	.38	1.59	1.4	1.58	1.3	.36	.18	.53	31.0	34.1	36	36	in subset: 1
33	12	2.75	2.73	-.41	.38	1.09	.3	1.08	.3	.91	.54	.53	34.9	34.1	37	37	in subset: 1
33	12	2.75	2.73	-.41	.38	.93	.0	.94	.0	1.09	.29	.53	36.4	34.1	48	48	in subset: 1
33	12	2.75	2.73	-.41	.38	1.34	.9	1.37	.9	.65	.51	.53	36.5	34.1	91	91	in subset: 1
33	12	2.75	2.73	-.41	.38	.97	.0	.97	.0	1.03	.76	.53	35.4	34.1	126	126	in subset: 1
33	12	2.75	2.73	-.41	.38	.95	.0	.91	-.1	1.16	.64	.53	36.7	34.1	127	127	in subset: 1
33	12	2.75	2.73	-.41	.38	.71	-.6	.70	-.7	1.30	.27	.53	38.2	34.1	176	176	in subset: 1
33	12	2.75	2.73	-.41	.38	1.37	.9	1.36	.9	.55	.14	.53	33.1	34.1	182	182	in subset: 1
35	12	2.92	2.74	-.39	.38	.51	-1.4	.53	-1.3	1.47	.72	.53	25.0	34.0	143	143	in subset: 4
34	12	2.83	2.76	-.36	.38	.84	-.3	.83	-.3	1.19	.31	.53	36.5	34.3	137	137	in subset: 2
34	12	2.83	2.81	-.27	.38	1.39	1.0	1.40	1.0	.57	.58	.53	31.8	34.0	6	6	in subset: 1
34	12	2.83	2.81	-.27	.38	.59	-1.1	.57	-1.1	1.46	.55	.53	36.9	34.0	8	8	in subset: 1
34	12	2.83	2.81	-.27	.38	2.43	2.7	2.50	2.8	-.63	-.06	.53	21.7	34.0	10	10	in subset: 1
34	12	2.83	2.81	-.27	.38	.49	-1.4	.48	-1.5	1.57	.64	.53	39.7	34.0	16	16	in subset: 1
34	12	2.83	2.81	-.27	.38	1.07	.3	1.10	.3	.96	.31	.53	34.2	34.0	23	23	in subset: 1
34	12	2.83	2.81	-.27	.38	.41	-1.8	.41	-1.8	1.66	.72	.53	38.8	34.0	25	25	in subset: 1
34	12	2.83	2.81	-.27	.38	.63	-.9	.63	-.9	1.44	.66	.53	37.6	34.0	32	32	in subset: 1
34	12	2.83	2.81	-.27	.38	1.29	.8	1.25	.7	.75	.50	.53	32.0	34.0	47	47	in subset: 1
34	12	2.83	2.81	-.27	.38	1.50	1.2	1.55	1.3	.42	.18	.53	30.2	34.0	50	50	in subset: 1
34	12	2.83	2.81	-.27	.38	.57	-1.1	.57	-1.1	1.47	.56	.53	38.7	34.0	59	59	in subset: 1
34	12	2.83	2.81	-.27	.38	1.39	1.0	1.41	1.0	.52	.58	.53	33.3	34.0	61	61	in subset: 1
34	12	2.83	2.81	-.27	.38	1.16	.5	1.19	.5	.81	-.01	.53	32.4	34.0	86	86	in subset: 1
34	12	2.83	2.81	-.27	.38	.80	-.4	.79	-.4	1.19	.69	.53	39.5	34.0	104	104	in subset: 1
34	12	2.83	2.81	-.27	.38	.68	-.7	.71	-.7	1.33	.78	.53	38.0	34.0	106	106	in subset: 1
34	12	2.83	2.81	-.27	.38	1.23	.6	1.28	.7	.71	.37	.53	32.8	34.0	117	117	in subset: 1
34	12	2.83	2.81	-.27	.38	.45	-1.6	.46	-1.6	1.57	.49	.53	38.5	34.0	128	128	in subset: 1
34	12	2.83	2.81	-.27	.38	1.02	.1	1.02	.1	.98	.13	.53	32.9	34.0	131	131	in subset: 1
34	12	2.83	2.81	-.27	.38	.61	-1.0	.60	-1.0	1.43	.52	.53	39.3	34.0	139	139	in subset: 1
34	12	2.83	2.81	-.27	.38	.64	-.9	.63	-.9	1.38	.66	.53	36.7	34.0	178	178	in subset: 1
35	12	2.92	2.90	-.12	.38	1.43	1.1	1.48	1.2	.44	.49	.53	30.9	33.8	11	11	in subset: 1
35	12	2.92	2.90	-.12	.38	.45	-1.6	.45	-1.6	1.60	.61	.53	39.1	33.8	12	12	in subset: 1
35	12	2.92	2.90	-.12	.38	.54	-1.3	.53	-1.3	1.57	.83	.53	39.7	33.8	18	18	in subset: 1
35	12	2.92	2.90	-.12	.38	.51	-1.4	.51	-1.4	1.57	.71	.53	40.8	33.8	19	19	in subset: 1
35	12	2.92	2.90	-.12	.38	.56	-1.2	.56	-1.2	1.51	.67	.53	39.2	33.8	30	30	in subset: 1
35	12	2.92	2.90	-.12	.38	1.26	.7	1.27	.7	.73	.73	.53	31.2	33.8	89	89	in subset: 1
35	12	2.92	2.90	-.12	.38	.73	-.6	.73	-.6	1.31	.52	.53	35.3	33.8	96	96	in subset: 1
35	12	2.92	2.90	-.12	.38	.55	-1.2	.55	-1.2	1.52	.68	.53	37.1	33.8	98	98	in subset: 1
35	12	2.92	2.90	-.12	.38	1.42	1.0	1.41	1.0	.47	.35	.53	33.9	33.8	101	101	in subset: 1
35	12	2.92	2.90	-.12	.38	1.35	.9	1.37	.9	.61	.39	.53	29.4	33.8	102	102	in subset: 1
35	12	2.92	2.90	-.12	.38	.41	-1.8	.42	-1.8	1.60	.46	.53	36.8	33.8	110	110	in subset: 1
35	12	2.92	2.90	-.12	.38	.77	-.5	.77	-.5	1.23	.28	.53	36.1	33.8	116	116	in subset: 1
35	12	2.92	2.90	-.12	.38	.52	-1.3	.52	-1.3	1.52	.54	.53	40.0	33.8	119	119	in subset: 1
35	12	2.92	2.90	-.12	.38	.54	-1.2	.54	-1.2	1.49	.52	.53	40.5	33.8	122	122	in subset: 1
35	12	2.92	2.90	-.12	.38	.79	-.4	.82	-.3	1.24	.47	.53	38.1	33.8	156	156	in subset: 1
37	12	3.08	2.91	-.10	.38	2.03	2.1	2.01	2.1	-.16	.28	.53	29.2	34.5	190	190	in subset: 4
36	12	3.00	2.93	-.07	.38	1.05	.2	1.04	.2	.96	.53	.53	36.5	34.3	78	78	in subset: 2
36	12	3.00	2.93	-.07	.38	.66	-.8	.65	-.8	1.38	.50	.53	37.0	34.3	150	150	in subset: 2
36	12	3.00	2.99	.03	.38	1.21	.6	1.22	.6	.76	.70	.53	30.7	33.4	17	17	in subset: 1
36	12	3.00	2.99	.03	.38	.70	-.7	.69	-.7	1.34	.47	.53	36.3	33.4	27	27	in subset: 1
36	12	3.00	2.99	.03	.38	.57	-1.1	.56	-1.2	1.52	.74	.53	37.3	33.4	54	54	in subset: 1
36	12	3.00	2.99	.03	.38	.96	.0	.96	.0	.98	.43	.53	33.2	33.4	74	74	in subset: 1
36	12	3.00	2.99	.03	.38	.61	-1.0	.61	-1.0	1.44	.55	.53	39.0	33.4	92	92	in subset: 1
36	12	3.00	2.99	.03	.38	.33	-2.2	.33	-2.2	1.76	.81	.53	40.9	33.4	124	124	in subset: 1
36	12	3.00	2.99	.03	.38	.38	-1.9	.38	-1.9	1.69	.76	.53	39.6	33.4	129	129	in subset: 1
36	12	3.00	2.99	.03	.38	.47	-1.5	.47	-1.5	1.63	.82	.53	37.7	33.4	133	133	in subset: 1
36	12	3.00	2.99	.03	.38	.92	.0	.95	.0	1.11	.45	.53	35.1	33.4	134	134	in subset: 1
36	12	3.00	2.99	.03	.38	.60	-1.0	.60	-1.0	1.49	.72	.53	36.2	33.4	141	141	in subset: 1
36	12	3.00	2.99	.03	.38	1.14	.4	1.14	.4	.88	.74	.53	32.5	33.4	166	166	in subset: 1
36	12	3.00	2.99	.03	.38	1.06	.2	1.06	.2	.94	.14	.53	32.9	33.4	170	170	in subset: 1
37	12	3.08	3.02	.08	.38	.51	-1.4	.50	-1.4	1.54	.55	.53	39.1	34.0	49	49	in subset: 2
37	12	3.08	3.02	.08	.38	.96	.0	.96	.0	1.13	.65	.53	38.0	34.0	58	58	in subset: 2
37	12	3.08	3.07	.17	.38	.90	-.1	.88	-.1	1.15	.70	.53	36.8	32.8	20	20	in subset: 1
37	12	3.08	3.07	.17	.38	.63	-.9	.62	-1.0	1.47	.75	.53	37.7	32.8	77	77	in subset: 1
37	12	3.08	3.07	.17	.38	1.01	.1	1.02	.1	.97	.62	.53	32.6	32.8	79	79	in subset: 1
37	12	3.08	3.07	.17	.38	.57	-1.2	.56	-1.2	1.47	.49	.53	37.3	32.8	80	80	in subset: 1
37	12	3.08	3.07	.17	.38	.48	-1.5	.48	-1.5	1.54	.37	.53	37.7	32.8	103	103	in subset: 1
37	12	3.08	3.07	.17	.38	1.44	1.1	1.43	1.1	.42	.48	.53	30.1	32.8	115	115	in subset: 1
38	12	3.17	3.10	.22	.38	.48	-1.5	.47	-1.5	1.62	.79	.53	33.9	33.5	120	120	in subset: 2
38	12	3.17	3.10	.22	.38	.77	-.5	.76	-.5	1.33	.70	.53	43.8	33.5	161	161	in subset: 2
38	12	3.17	3.16	.31	.38	.73	-.6	.73	-.6	1.27	.73	.53	33.5	32.1	34	34	in subset: 1
38	12	3.17	3.16	.31	.38	1.23	.6	1.23	.6	.72	.52	.53	32.4	32.1	38	38	in subset: 1
38	12	3.17	3.16	.31	.38	1.42	1.0	1.39	1.0	.53	.67	.53	29.9	32.1	46	46	in subset: 1
38	12	3.17	3.16	.31	.38	.64	-.9	.63	-.9	1.41	.49	.53	36.4	32.1	62	62	in subset: 1
38	12	3.17	3.16	.31	.38	1.51	1.2	1.49	1.2	.43	.62	.53	30.2	32.1	84	84	in subset: 1
38	12	3.17	3.16	.31	.38	1.95	2.0	1.93	2.0	-.16	.63	.53	25.1	32.1	90	90	in subset: 1
38	12	3.17	3.16	.31	.38	.38	-1.9	.39	-1.9	1.66	.57	.53	36.3	32.1	107	107	in subset: 1
38	12	3.17	3.16	.31	.38	2.47	2.8	2.45	2.8	-.71	.36	.53	23.1	32.1	113	113	in subset: 1
38	12	3.17	3.16	.31	.38	.76	-.5	.75	-.5	1.30	.56	.53	36.1	32.1	160	160	in subset: 1
38	12	3.17	3.16	.31	.38	1.17	.5	1.16	.5	.77	.22	.53	34.7	32.1	169	169	in subset: 1
38	12	3.17	3.16	.31	.38	.57	-1.1	.57	-1.1	1.48	.55	.53	37.2	32.1	174	174	in subset: 1
39	12	3.25	3.19	.36	.38	.52	-1.4	.50	-1.4	1.60	.80	.53	41.1	32.9	42	42	in subset: 2
39	12	3.25	3.19	.36	.38	.60	-1.0	.59	-1.1	1.47	.58	.53	37.5	32.9	100	100	in subset: 2
39	12	3.25	3.19	.36	.38	.80	-.4	.79	-.4	1.28	.58	.53	37.5	32.9	159	159	in subset: 2
39	12	3.25	3.25	.46	.38	.60	-1.0	.60	-1.1	1.46	.58	.53	35.2	31.3	29	29	in subset: 1
39	12	3.25	3.25	.46	.38	.56	-1.2	.56	-1.2	1.44	.15	.53	33.4	31.3	51	51	in subset: 1
39	12	3.25	3.25	.46	.38	.59	-1.1	.58	-1.1	1.48	.59	.53	34.6	31.3	71	71	in subset: 1
39	12	3.25	3.25	.46	.38	.40	-1.9	.40	-1.9	1.67	.61	.53	37.8	31.3	130	130	in subset: 1
40	12	3.33	3.28	.51	.38	.67	-.8	.68	-.8	1.40	.72	.53	40.6	32.1	163	163	in subset: 2
40	12	3.33	3.33	.60	.38	1.23	.6	1.20	.6	.71	.61	.53	29.3	30.4	9	9	in subset: 1
40	12	3.33	3.33	.60	.38	.45	-1.7	.45	-1.6	1.63	.59	.53	32.0	30.4	67	67	in subset: 1
40	12	3.33	3.33	.60	.38	.94	.0	.94	.0	1.09	.79	.53	34.2	30.4	114	114	in subset: 1
40	12	3.33	3.33	.60	.38	1.00	.1	1.00	.1	.98	.01	.53	30.8	30.4	153	153	in subset: 1
40	12	3.33	3.33	.60	.38	.41	-1.8	.42	-1.8	1.67	.63	.53	35.5	30.4	172	172	in subset: 1
41	12	3.42	3.36	.66	.38	1.89	1.9	1.89	1.9	.03	.52	.53	28.1	31.2	7	7	in subset: 2
41	12	3.42	3.42	.75	.38	.94	.0	.93	.0	1.07	.52	.53	33.3	29.3	41	41	in subset: 1
41	12	3.42	3.42	.75	.38	1.09	.3	1.06	.2	.88	.71	.53	28.7	29.3	73	73	in subset: 1
41	12	3.42	3.42	.75	.38	1.31	.8	1.28	.8	.62	.58	.53	25.6	29.3	132	132	in subset: 1
41	12	3.42	3.42	.75	.38	.65	-.9	.64	-.9	1.45	.76	.53	29.7	29.3	148	148	in subset: 1
42	12	3.50	3.50	.90	.38	.78	-.4	.78	-.4	1.25	.50	.53	28.6	28.2	60	60	in subset: 1
42	12	3.50	3.50	.90	.38	.37	-2.0	.35	-2.1	1.76	.71	.53	34.4	28.2	70	70	in subset: 1
42	12	3.50	3.50	.90	.38	.41	-1.8	.42	-1.8	1.70	.82	.53	32.3	28.2	95	95	in subset: 1
42	12	3.50	3.50	.90	.38	.85	-.2	.85	-.2	1.16	.61	.53	30.9	28.2	123	123	in subset: 1
42	12	3.50	3.50	.90	.38	.96	.0	.96	.0	1.07	.68	.53	31.5	28.2	149	149	in subset: 1
42	12	3.50	3.50	.90	.38	1.27	.7	1.24	.7	.74	.62	.53	26.0	28.2	177	177	in subset: 1
44	12	3.67	3.52	.93	.39	1.11	.3	1.05	.2	.92	.58	.52	33.3	29.6	175	175	in subset: 4
44	12	3.67	3.52	.93	.39	2.04	2.2	1.99	2.1	-.23	.44	.52	38.9	29.6	187	187	in subset: 4
43	12	3.58	3.59	1.05	.39	.17	-3.3	.16	-3.4	1.96	.78	.52	32.5	26.9	157	157	in subset: 1
43	12	3.58	3.59	1.05	.39	.97	.0	.96	.0	1.04	.34	.52	25.8	26.9	179	179	in subset: 1
44	12	3.67	3.68	1.20	.39	1.00	.1	.98	.0	1.01	.66	.52	28.2	25.6	21	21	in subset: 1
44	12	3.67	3.68	1.20	.39	1.40	1.0	1.32	.8	.54	.68	.52	23.2	25.6	31	31	in subset: 1
45	12	3.75	3.76	1.35	.39	.43	-1.7	.42	-1.8	1.66	.59	.52	26.6	24.2	109	109	in subset: 1
45	12	3.75	3.76	1.35	.39	.78	-.4	.82	-.3	1.19	.19	.52	26.2	24.2	152	152	in subset: 1
46	12	3.83	3.85	1.51	.40	3.03	3.5	2.98	3.5	-1.36	.49	.51	12.0	22.8	158	158	in subset: 1
47	12	3.92	3.94	1.67	.41	1.45	1.1	1.40	1.0	.62	.77	.51	18.2	21.3	97	97	in subset: 1

49	12	4.08	4.11	2.01	.42	1.86	1.8	1.71	1.6	.20	.63	.49	16.4	18.2	22 22	in subset: 1
49	12	4.08	4.11	2.01	.42	2.55	2.9	2.43	2.7	-.67	.40	.49	15.5	18.2	171 171	in subset: 1
34.0	12.0	2.83	2.80	-.31	.39	1.00	-.2	1.00	-.2		.50				Mean (Count: 190)	
5.6	.0	.47	.48	.85	.02	.70	1.4	.70	1.4		.24				S.D. (Population)	
5.6	.0	.47	.48	.85	.02	.70	1.4	.70	1.4		.24				S.D. (Sample)	

Model, Populn: RMSE .39 Adj (True) S.D. .76 Separation 1.93 Strata 2.91 Reliability (not inter-rater) .79
Model, Sample: RMSE .39 Adj (True) S.D. .76 Separation 1.94 Strata 2.92 Reliability (not inter-rater) .79
Model, Fixed (all same) chi-square: 824.0 d.f.: 189 significance (probability): .00
Model, Random (normal) chi-square: 157.5 d.f.: 188 significance (probability): .95
Inter-Rater agreement opportunities: 164244 Exact agreements: 54396 = 33.1% Expected: 51990.6 = 31.7%

Rasch Analysis of Expert Ratings 22/12/2014 16:45:26
Table 7.2.1 Dhivehi English Familiarity Level Measurement Report (arranged by MN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	N Dhivehi English Familiarity Level	
28	12	2.33	2.55	-.42	.40	.42	-1.7	.45	-1.6	1.63	.40	.51	3 Some Familiarity	in subset: 3
5569	1980	2.81	2.80	.02	.03	.94	-1.9	.95	-1.7	1.06	.66	.65	1 No Familiarity	in subset: 1
595	204	2.92	2.85	.11	.09	.78	-2.5	.77	-2.5	1.27	.69	.62	2 Limited Familiarity	in subset: 2
261	84	3.11	2.96	.29	.15	2.83	8.5	2.80	8.4	-1.10	.36	.60	4 Very Familiar	in subset: 4
1613.3	570.0	2.79	2.79	.00	.17	1.24	.6	1.24	.6		.53		Mean (Count: 4)	
2292.7	816.9	.29	.15	.26	.14	.94	4.6	.92	4.5		.15		S.D. (Population)	
2647.4	943.3	.33	.17	.30	.16	1.08	5.3	1.06	5.2		.17		S.D. (Sample)	

Model, Populn: RMSE .22 Adj (True) S.D. .14 Separation .63 Strata 1.18 Reliability .29
Model, Sample: RMSE .22 Adj (True) S.D. .20 Separation .93 Strata 1.58 Reliability .47
Model, Fixed (all same) chi-square: 5.3 d.f.: 3 significance (probability): .15
Model, Random (normal) chi-square: 1.4 d.f.: 2 significance (probability): .48

Rasch Analysis of Expert Ratings 22/12/2014 16:45:26
Table 7.3.1 Speaker Measurement Report (arranged by MN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	Nu Speaker
424	190	2.23	2.18	-1.10	.10	.74	-2.7	.74	-2.8	1.30	.52	.52	4 Speaker 5 Spanish-English Male 1
436	190	2.29	2.25	-.98	.10	1.35	3.2	1.37	3.3	.58	.35	.52	2 Speaker 3 Arabic-English Female 1
471	190	2.48	2.43	-.62	.10	1.00	.0	1.01	.0	.97	.44	.53	8 Speaker 9 Spanish-English Female 1
479	190	2.52	2.48	-.54	.10	1.30	2.7	1.29	2.7	.68	.40	.53	1 Speaker 2 Dhivehi-English Female 1
490	190	2.58	2.54	-.44	.10	.82	-1.8	.83	-1.8	1.18	.53	.53	6 Speaker 7 Spanish-English Male 2
505	190	2.66	2.62	-.29	.10	.85	-1.6	.85	-1.6	1.19	.51	.54	3 Speaker 4 Dhivehi-English Female 2
530	190	2.79	2.75	-.06	.10	1.56	4.8	1.55	4.8	.38	.59	.54	7 Speaker 8 Dhivehi-English Male 1
590	190	3.11	3.08	.50	.10	.84	-1.7	.83	-1.8	1.18	.65	.54	5 Speaker 6 Arabic-English Male 1
602	190	3.17	3.15	.61	.10	1.02	.2	1.02	.2	.99	.63	.54	11 Speaker 12 Arabic-English Female 2
618	190	3.25	3.24	.75	.10	.79	-2.3	.79	-2.2	1.24	.64	.54	12 Speaker 13 Dhivehi-English Male 2
654	190	3.44	3.44	1.09	.10	.78	-2.4	.77	-2.5	1.25	.60	.54	9 Speaker 10 Arabic-English 2
654	190	3.44	3.44	1.09	.10	.92	-.7	.91	-.8	1.10	.56	.54	10 Speaker 11 Spanish-English Female 2
537.8	190.0	2.83	2.80	.00	.10	1.00	-.2	1.00	-.2		.53		Mean (Count: 12)
79.0	.0	.42	.43	.75	.00	.25	2.4	.25	2.4		.09		S.D. (Population)
82.5	.0	.43	.45	.78	.00	.27	2.5	.27	2.6		.10		S.D. (Sample)

Model, Populn: RMSE .10 Adj (True) S.D. .74 Separation 7.55 Strata 10.40 Reliability .98
Model, Sample: RMSE .10 Adj (True) S.D. .77 Separation 7.89 Strata 10.85 Reliability .98
Model, Fixed (all same) chi-square: 686.7 d.f.: 11 significance (probability): .00
Model, Random (normal) chi-square: 10.8 d.f.: 10 significance (probability): .37

Rasch Analysis of Expert Ratings 22/12/2014 16:45:26
Table 8.1 Category Statistics.

Model = ?,?,?,R5

DATA Score	Category Counts Total	Used	%	Cum. %	QUALITY CONTROL Avge Meas	Exp. Meas	OUTFIT MnSq	RASCH-ANDRICH Thresholds Measure	S.E.	EXPECTATION Measure at Category (Mean)	-0.5	MOST PROBABLE from (Modal)	RASCH-THURSTONE Thresholds (Median)	Cat PEAK Prob
1	166	166	7%	7%	-1.49	-1.67	1.2			(-3.90)		low	low	100%
2	709	709	31%	38%	-.94	-.91	1.0	-2.74	.09	-1.78	-3.00	-2.74	-2.85	57%
3	842	842	37%	75%	-.19	-.14	1.0	-.69	.05	.05	-.81	-.69	-.76	50%
4	472	472	21%	96%	.70	.61	.9	.82	.06	1.80	.89	.82	.85	54%
5	91	91	4%	100%	1.33	1.31	1.0	2.61	.12	(3.80)	2.93	2.61	2.74	100%

Scale structure

```
Measr:-4.0             -2.0              0.0              2.0              4.0
        +                +                +                +                +
  Mode:<(^)-------12-------^-------23-----^-----34-------^-----45--------(^)5>

Median:<(^)------12--------^-------23-----^-----34-------^------45-------(^)5>

  Mean:<(^)-----12---------^------23------^------34------^--------45-----(^)5>
        +                +                +                +                +
Measr:-4.0             -2.0              0.0              2.0              4.0
```

Probability Curves

```
      -4.0             -2.0              0.0              2.0              4.0
       ++----------------+----------------+----------------+----------------++
     1 |                                                                    |
       |                                                                    |
       |                                                                    |
       |                                                                    |
       |                                                                   5|
     P |111                                                             555 |
     r |   1                                                           5    |
     o |    11                                                       55     |
     b |      11                                                   55       |
     a |        1     2222222222                      444444      5         |
     b |         11222          222    3333333      444     444 55          |
     i |         2211              2233       333 44           *4           |
     l |       22    1             332          4*3          55  44         |
     i |     22       11         33   22      44   33       5      444      |
     t |   22           1      33       22   4       33    55         44    |
     y | 22              11  33           244        3355              44   |
       |2                  **          4422         5533                44|
       |                333  111      444    222     55    333              |
       |             333        111444          22*55         333           |
       |      3333333         44444111111   555555 22222         333333     |
     0 |**********************55555555555***111111111111********************|
       ++----------------+----------------+----------------+----------------++
      -4.0             -2.0              0.0              2.0              4.0
```

Expected Score Ogive (Model ICC)

```
      -4.0             -2.0              0.0              2.0              4.0
       ++----------------+----------------+----------------+----------------++
     5 |                                                                    |
       |                                                              555555|
       |                                                        445555      |
       |                                                   44444            |
     4 |                                               4444                 |
       |                                           4444                     |
```

```
    |                                       3344                        |
    |                                    333                            |
  3 |                                 3333                              |
    |                             3333                                  |
    |                          223                                      |
    |                       2222                                        |
  2 |                  22222                                            |
    |             22222                                                 |
    |      1111222                                                      |
    |11111                                                              |
  1 |                                                                   |
    ++----------------+----------------+----------------+----------------++
  -4.0             -2.0              0.0              2.0              4.0
```

Rasch Analysis of Expert Ratings 22/12/2014 16:45:26
Table 4.1 Unexpected Responses (9 residuals sorted by u).

Cat	Score	Exp.	Resd	StRes	Num	rat	N	Dhivehi English Fam	Nu	Speaker
5	5	2.4	2.6	3.5	118	118	1	No Familiarity	7	Speaker 8 Dhivehi-English Male 1
1	1	3.6	-2.6	-3.4	158	158	1	No Familiarity	1	Speaker 2 Dhivehi-English Female 1
5	5	2.5	2.5	3.3	188	188	4	Very Familiar	1	Speaker 2 Dhivehi-English Female 1
5	5	2.5	2.5	3.3	189	189	4	Very Familiar	1	Speaker 2 Dhivehi-English Female 1
2	2	4.2	-2.2	-3.2	158	158	1	No Familiarity	11	Speaker 12 Arabic-English Female 2
1	1	3.5	-2.5	-3.2	189	189	4	Very Familiar	9	Speaker 10 Arabic-English 2
1	1	3.5	-2.5	-3.2	189	189	4	Very Familiar	10	Speaker 11 Spanish-English Female 2
5	5	2.7	2.3	3.1	188	188	4	Very Familiar	3	Speaker 4 Dhivehi-English Female 2
5	5	2.7	2.3	3.1	189	189	4	Very Familiar	3	Speaker 4 Dhivehi-English Female 2
Cat	Score	Exp.	Resd	StRes	Num	rat	N	Dhivehi English Fam	Nu	Speaker

The Facets analyses output of bias/interaction - Spanish-English

Rasch Analysis of Expert Ratings 08/01/2015 15:18:29
Table 9.3 Bias Iteration Report.

Bias/Interaction: 2. Spanish English Familiarity Level, 3. Speaker

There are empirically 48 Bias terms

Iteration	Max. Score Residual Elements	%	Categories	Max. Logit Change Elements	Steps
BIAS 1	-12.2210	-27.8		-.6419	
BIAS 2	-.4529	-1.0		-.0258	
BIAS 3	-.0009	.0		-.0001	

Rasch Analysis of Expert Ratings 08/01/2015 15:18:29
Table 11.3 Bias/Interaction Measurement Summary.

Bias/Interaction: 2. Spanish English Familiarity Level, 3. Speaker

Cat	Score	Exp.	Resd	StRes	
2.83	2.83	2.83	.00	.00	Mean (Count: 2280)
.97	.97	.64	.73	.99	S.D. (Population)
.97	.97	.64	.73	.99	S.D. (Sample)

```
Count of measurable responses          = 2280
Raw-score variance of observations      =   0.94 100.00%
Variance explained by Rasch measures    =   0.39  42.00%
Variance of residuals                   =   0.55  58.00%
Variance explained by bias/interactions =   0.01   1.40%
Variance remaining in residuals         =   0.53  56.61%
```

Rasch Analysis of Expert Ratings 08/01/2015 15:18:29
Table 12.3 Bias/Interaction Summary Report.

Bias/Interaction: 2. Spanish English Familiarity Level, 3. Speaker (higher score = higher bias measure)

```
Bias/Interaction Size:
              1       1   1   1 113321 44 221316112  1 11 1    11
 +--------------------Q---------S--------M--------S---------Q--------------------+
-1                                       0                                       1

Bias/Interaction Significance:
   1      1    1  213 21112 43 22 21412121 111  111    1
 +--------Q+--------S+---------M---------+S--------+Q--------+
-3        -2        -1         0         1         2         3
```

Rasch Analysis of Expert Ratings 08/01/2015 15:18:29
Table 13.3.1 Bias/Interaction Report (arranged by mN).

Bias/Interaction: 2. Spanish English Familiarity Level, 3. Speaker (higher score = higher bias measure)

Observd Score	Expctd Score	Observd Count	Obs-Exp Average	Bias Size	Model S.E.	t	d.f.	Prob.	Infit MnSq	Outfit MnSq	Sq	N	Spanish English Fam	measr	Nu	Speaker	measr
104	93.27	35	.31	.55	.22	2.44	34	.0200	1.0	1.0	24	4	Very Familiar	.13	6	Speaker 7 Spanish-English Male 2	-.44
50	44.69	20	.27	.52	.31	1.69	19	.1081	1.6	1.5	1	1	No Familiarity	-.26	1	Speaker 2 Dhivehi-English Female 1	-.54
131	123.31	35	.22	.41	.23	1.75	34	.0891	1.0	1.0	40	4	Very Familiar	.13	10	Speaker 11 Spanish-English Female 2	1.09
51	47.21	20	.19	.36	.31	1.18	19	.2516	.7	.8	9	1	No Familiarity	-.26	3	Speaker 4 Dhivehi-English Female 2	-.29
151	140.64	57	.18	.33	.18	1.88	56	.0657	1.5	1.5	2	2	Limited Familiarity	.00	1	Speaker 2 Dhivehi-English Female 1	-.54
42	39.47	20	.13	.28	.33	.85	19	.4057	1.7	1.6	13	1	No Familiarity	-.26	4	Speaker 5 Spanish-English Male 1	-1.10
193	184.38	78	.11	.21	.15	1.35	77	.1814	1.4	1.4	7	3	Some Familiarity	.12	2	Speaker 3 Arabic-English Female 1	-.98
208	199.14	78	.11	.21	.15	1.36	77	.1788	.7	.7	31	3	Some Familiarity	.12	8	Speaker 9 Spanish-English Female 1	-.62
154	148.37	57	.10	.18	.18	1.01	56	.3185	1.1	1.0	10	2	Limited Familiarity	.00	3	Speaker 4 Dhivehi-English Female 2	-.29

198	193.15	57	.09	.15	.18	.85	56	.3981	1.1	1.1	38	2 Limited Familiarity	.00	10 Speaker 11 Spanish-English Female 2	1.09
160	155.82	57	.07	.13	.18	.74	56	.4630	2.0	1.9	26	2 Limited Familiarity	.00	7 Speaker 8 Dhivehi-English Male 1	-.06
47	45.75	20	.06	.12	.31	.39	19	.6984	.9	.9	21	1 No Familiarity	-.26	6 Speaker 7 Spanish-English Male 2	-.44
92	89.73	35	.06	.12	.23	.52	34	.6052	1.1	1.2	32	4 Very Familiar	.13	8 Speaker 9 Spanish-English Female 1	-.62
197	193.15	57	.07	.12	.18	.68	56	.5019	.9	.9	34	2 Limited Familiarity	.00	9 Speaker 10 Arabic-English Male 2	1.09
259	253.78	78	.07	.12	.15	.78	77	.4389	.9	.9	43	3 Some Familiarity	.12	11 Speaker 12 Arabic-English Female 2	.61
83	80.95	35	.06	.11	.24	.49	34	.6294	.6	.7	16	4 Very Familiar	.13	4 Speaker 5 Spanish-English Male 1	-1.10
253	248.82	78	.05	.09	.15	.62	77	.5352	.6	.6	19	3 Some Familiarity	.12	5 Speaker 6 Arabic-English Male 1	.50
185	182.25	57	.05	.08	.17	.48	56	.6333	1.0	1.0	46	2 Limited Familiarity	.00	12 Speaker 13 Dhivehi-English Male 2	.75
264	260.39	78	.05	.08	.15	.54	77	.5904	.7	.7	47	3 Some Familiarity	.12	12 Speaker 13 Dhivehi-English Male 2	.75
102	100.67	35	.04	.07	.22	.30	34	.7664	1.3	1.3	28	4 Very Familiar	.13	7 Speaker 8 Dhivehi-English Male 1	-.06
181	179.31	78	.02	.04	.16	.27	77	.7905	.5	.5	15	3 Some Familiarity	.12	4 Speaker 5 Spanish-English Male 1	-1.10
50	49.66	20	.02	.03	.31	.10	19	.9188	1.8	1.8	25	1 No Familiarity	-.26	7 Speaker 8 Dhivehi-English Male 1	-.06
112	111.68	35	.01	.02	.22	.07	34	.9440	1.0	1.0	20	4 Very Familiar	.13	5 Speaker 6 Arabic-English Male 1	.50
44	43.92	20	.00	.01	.32	.02	19	.9804	1.6	1.5	29	1 No Familiarity	-.26	8 Speaker 9 Spanish-English Female 1	-.62
275	275.18	78	.00	.00	.15	-.03	77	.9780	.6	.6	35	3 Some Familiarity	.12	9 Speaker 10 Arabic-English Male 2	1.09
176	177.43	57	-.03	-.04	.17	-.25	56	.8044	1.1	1.0	42	2 Limited Familiarity	.00	11 Speaker 12 Arabic-English Female 2	.61
58	58.55	20	-.03	-.05	.30	-.16	19	.8724	.8	.8	45	1 No Familiarity	-.26	12 Speaker 13 Dhivehi-English Male 2	.75
95	96.05	35	-.03	-.05	.23	-.24	34	.8137	.7	.7	12	4 Very Familiar	.13	3 Speaker 4 Dhivehi-English Female 2	-.29
172	173.81	57	-.03	-.06	.17	-.32	56	.7530	.9	.9	18	2 Limited Familiarity	.00	5 Speaker 6 Arabic-English Male 1	.50
82	83.20	35	-.03	-.07	.24	-.28	34	.7802	1.1	1.1	8	4 Very Familiar	.13	2 Speaker 3 Arabic-English Female 1	-.98
122	123.31	35	-.04	-.07	.23	-.30	34	.7679	.9	.9	36	4 Very Familiar	.13	9 Speaker 10 Arabic-English Male 2	1.09
56	56.90	20	-.04	-.08	.30	-.27	19	.7906	1.3	1.3	41	1 No Familiarity	-.26	11 Speaker 12 Arabic-English Female 2	.61
199	202.50	78	-.04	-.08	.15	-.53	77	.5952	.9	.9	3	3 Some Familiarity	.12	1 Speaker 2 Dhivehi-English Female 1	-.54
218	223.87	78	-.08	-.13	.15	-.88	77	.3828	1.3	1.3	27	3 Some Familiarity	.12	7 Speaker 8 Dhivehi-English Male 1	-.06
111	113.87	35	-.08	-.14	.22	-.65	34	.5231	1.0	1.0	44	4 Very Familiar	.13	11 Speaker 12 Arabic-English Female 2	.61
139	143.91	57	-.09	-.16	.18	-.89	56	.3799	.9	.9	22	2 Limited Familiarity	.00	6 Speaker 7 Spanish-English Male 2	-.44
268	275.18	78	-.09	-.16	.15	-1.09	77	.2808	.7	.7	39	3 Some Familiarity	.12	10 Speaker 11 Spanish-English Female 2	1.09
200	207.12	78	-.09	-.17	.15	-1.08	77	.2837	.6	.6	23	3 Some Familiarity	.12	6 Speaker 7 Spanish-English Male 2	-.44
39	40.59	20	-.08	-.18	.34	-.53	19	.6037	1.6	1.7	5	1 No Familiarity	-.26	2 Speaker 3 Arabic-English Female 1	-.98
205	213.41	78	-.11	-.19	.15	-1.27	77	.2088	.7	.8	11	3 Some Familiarity	.12	3 Speaker 4 Dhivehi-English Female 2	-.29
60	62.31	20	-.12	-.20	.30	-.69	19	.5005	1.1	1.1	33	1 No Familiarity	-.26	9 Speaker 10 Arabic-English Male 2	1.09
122	127.90	57	-.10	-.21	.19	-1.10	56	.2744	1.3	1.3	6	2 Limited Familiarity	.00	2 Speaker 3 Arabic-English Female 1	-.98
118	124.35	57	-.11	-.23	.19	-1.20	56	.2351	.8	.8	14	2 Limited Familiarity	.00	4 Speaker 5 Spanish-English Male 1	-1.10
53	55.67	20	-.13	-.24	.30	-.80	19	.4319	1.2	1.2	17	1 No Familiarity	-.26	5 Speaker 6 Arabic-English Male 1	.50
111	116.79	35	-.17	-.29	.22	-1.30	34	.2017	.6	.6	48	4 Very Familiar	.13	12 Speaker 13 Dhivehi-English Male 2	.75
127	138.27	57	-.20	-.38	.19	-2.05	56	.0448	1.1	1.1	30	2 Limited Familiarity	.00	8 Speaker 9 Spanish-English Female 1	-.62
57	62.31	20	-.27	-.47	.30	-1.58	19	.1312	.9	1.0	37	1 No Familiarity	-.26	10 Speaker 11 Spanish-English Female 2	1.09
79	91.22	35	-.35	-.67	.24	-2.79	34	.0086	1.2	1.2	4	4 Very Familiar	.13	1 Speaker 2 Dhivehi-English Female 1	-.54
134.4	134.44	47.5	.00	.00	.22	.00			1.0	1.0	Mean (Count: 48)				
69.4	69.14	22.0	.13	.23	.06	1.05			.3	.3	S.D. (Population)				
70.2	69.87	22.2	.13	.24	.06	1.06			.3	.3	S.D. (Sample)				

Fixed (all = 0) chi-square: 53.2 d.f.: 48 significance (probability): .28

Rasch Analysis of Expert Ratings 08/01/2015 15:18:29
Table 14.3.1.2 Bias/Interaction Pairwise Report (arranged by mN).

Bias/Interaction: 2. Spanish English Familiarity Level, 3. Speaker

Target Nu Speaker	Target Measr	S.E.	Obs-Exp Average	Context N Spanish English Fam	Target Measr	S.E.	Obs-Exp Average	Context N Spanish English Fam	Target Contrast	Joint S.E.	t	Welch d.f.	Prob.
1 Speaker 2 Dhivehi-English Female 1	-.03	.31	.27	1 No Familiarity	-1.21	.24	-.35	4 Very Familiar	1.19	.39	3.05	44	.0039
1 Speaker 2 Dhivehi-English Female 1	-.21	.18	.18	2 Limited Familiarity	-1.21	.24	-.35	4 Very Familiar	1.00	.30	3.36	77	.0012
1 Speaker 2 Dhivehi-English Female 1	-.03	.31	.27	1 No Familiarity	-.63	.15	-.04	3 Some Familiarity	.60	.34	1.75	40	.0883
1 Speaker 2 Dhivehi-English Female 1	-.63	.15	-.04	3 Some Familiarity	-1.21	.24	-.35	4 Very Familiar	.59	.28	2.06	77	.0425
3 Speaker 4 Dhivehi-English Female 2	.07	.31	.19	1 No Familiarity	-.49	.15	-.11	3 Some Familiarity	.55	.34	1.62	40	.1124
4 Speaker 5 Spanish-English Male 1	-.82	.33	.13	1 No Familiarity	-1.33	.19	-.11	2 Limited Familiarity	.51	.38	1.34	43	.1864
1 Speaker 2 Dhivehi-English Female 1	-.21	.18	.18	2 Limited Familiarity	-.63	.15	-.04	3 Some Familiarity	.42	.23	1.77	125	.0790
3 Speaker 4 Dhivehi-English Female 2	.07	.31	.19	1 No Familiarity	-.35	.23	-.03	4 Very Familiar	.42	.38	1.09	44	.2813
8 Speaker 9 Spanish-English Female 1	-.61	.32	.00	1 No Familiarity	-1.01	.19	-.20	2 Limited Familiarity	.39	.37	1.06	42	.2959
12 Speaker 13 Dhivehi-English Male 2	.84	.17	.05	2 Limited Familiarity	.46	.22	-.17	4 Very Familiar	.38	.28	1.32	78	.1902
12 Speaker 13 Dhivehi-English Male 2	.83	.15	.05	3 Some Familiarity	.46	.22	-.17	4 Very Familiar	.37	.27	1.38	79	.1707
3 Speaker 4 Dhivehi-English Female 2	-.12	.18	.10	2 Limited Familiarity	-.49	.15	-.11	3 Some Familiarity	.37	.23	1.59	125	.1146
10 Speaker 11 Spanish-English Female 2	1.24	.18	.09	2 Limited Familiarity	.93	.15	-.09	3 Some Familiarity	.31	.23	1.35	125	.1787
6 Speaker 7 Spanish-English Male 2	-.31	.31	.06	1 No Familiarity	-.60	.15	-.09	3 Some Familiarity	.29	.35	.83	39	.4130
6 Speaker 7 Spanish-English Male 2	-.31	.31	.06	1 No Familiarity	-.60	.18	-.09	2 Limited Familiarity	.28	.36	.79	42	.4368
2 Speaker 3 Arabic-English Female 1	-.77	.15	.11	3 Some Familiarity	-1.04	.24	-.03	4 Very Familiar	.27	.28	.97	78	.3334

7	Speaker 8 Dhivehi-English Male 1	.07	.18	.07	2 Limited Familiarity	-.19	.15	-.08	3 Some Familiarity	.26	.23	1.13	125	.2599
11	Speaker 12 Arabic-English Female 2	.72	.15	.07	3 Some Familiarity	.46	.22	-.08	4 Very Familiar	.26	.27	.97	78	.3358
12	Speaker 13 Dhivehi-English Male 2	.71	.30	-.03	1 No Familiarity	.46	.22	-.17	4 Very Familiar	.24	.37	.65	44	.5185
4	Speaker 5 Spanish-English Male 1	-.82	.33	.13	1 No Familiarity	-1.06	.16	.02	3 Some Familiarity	.24	.36	.65	39	.5195
3	Speaker 4 Dhivehi-English Female 2	-.12	.18	.10	2 Limited Familiarity	-.35	.23	-.03	4 Very Familiar	.23	.29	.81	78	.4223
9	Speaker 10 Arabic-English Male 2	1.21	.18	.07	2 Limited Familiarity	1.02	.23	-.04	4 Very Familiar	.19	.29	.65	78	.5179
1	Speaker 2 Dhivehi-English Female 1	-.03	.31	.27	1 No Familiarity	-.21	.18	.18	2 Limited Familiarity	.18	.36	.52	42	.6075
3	Speaker 4 Dhivehi-English Female 2	.07	.31	.19	1 No Familiarity	-.12	.18	.10	2 Limited Familiarity	.18	.35	.52	42	.6074
7	Speaker 8 Dhivehi-English Male 1	-.03	.31	.02	1 No Familiarity	-.19	.15	-.08	3 Some Familiarity	.16	.34	.48	39	.6351
4	Speaker 5 Spanish-English Male 1	-.82	.33	.13	1 No Familiarity	-.99	.24	.06	4 Very Familiar	.16	.40	.41	43	.6874
9	Speaker 10 Arabic-English Male 2	1.21	.18	.07	2 Limited Familiarity	1.08	.15	.00	3 Some Familiarity	.12	.23	.53	125	.5962
11	Speaker 12 Arabic-English Female 2	.56	.17	-.03	2 Limited Familiarity	.46	.22	-.08	4 Very Familiar	.10	.28	.36	78	.7227
8	Speaker 9 Spanish-English Female 1	-.42	.15	.11	3 Some Familiarity	-.50	.23	.06	4 Very Familiar	.09	.27	.31	78	.7541
5	Speaker 6 Arabic-English Male 1	.59	.15	.05	3 Some Familiarity	.51	.22	.01	4 Very Familiar	.08	.27	.29	78	.7757
11	Speaker 12 Arabic-English Female 2	.53	.30	-.04	1 No Familiarity	.46	.22	-.08	4 Very Familiar	.06	.37	.17	44	.8656
9	Speaker 10 Arabic-English Male 2	1.08	.15	.00	3 Some Familiarity	1.02	.23	-.04	4 Very Familiar	.06	.27	.23	78	.8169
7	Speaker 8 Dhivehi-English Male 1	.07	.18	.07	2 Limited Familiarity	.01	.22	.04	4 Very Familiar	.06	.29	.22	78	.8264
2	Speaker 3 Arabic-English Female 1	-1.15	.34	-.08	1 No Familiarity	-1.19	.19	-.10	2 Limited Familiarity	.03	.39	.08	41	.9350
6	Speaker 7 Spanish-English Male 2	-.60	.18	-.09	2 Limited Familiarity	-.60	.15	-.09	3 Some Familiarity	.00	.24	.02	125	.9867
12	Speaker 13 Dhivehi-English Male 2	.84	.17	.05	2 Limited Familiarity	.83	.15	.05	3 Some Familiarity	.00	.23	.01	125	.9901
7	Speaker 8 Dhivehi-English Male 1	-.03	.31	.02	1 No Familiarity	.01	.22	.04	4 Very Familiar	-.04	.38	-.09	43	.9260
11	Speaker 12 Arabic-English Female 2	.53	.30	-.04	1 No Familiarity	.56	.17	-.03	2 Limited Familiarity	-.04	.35	-.11	42	.9148
5	Speaker 6 Arabic-English Male 1	.44	.17	-.03	2 Limited Familiarity	.51	.22	.01	4 Very Familiar	-.07	.28	-.25	78	.8031
4	Speaker 5 Spanish-English Male 1	-1.06	.16	.02	3 Some Familiarity	-.99	.24	.06	4 Very Familiar	-.07	.28	-.26	79	.7983
7	Speaker 8 Dhivehi-English Male 1	-.03	.31	.02	1 No Familiarity	.07	.18	.07	2 Limited Familiarity	-.10	.35	-.28	42	.7825
8	Speaker 9 Spanish-English Female 1	-.61	.32	.00	1 No Familiarity	-.50	.23	.06	4 Very Familiar	-.11	.39	-.28	43	.7785
2	Speaker 3 Arabic-English Female 1	-1.15	.34	-.08	1 No Familiarity	-1.04	.24	-.03	4 Very Familiar	-.11	.41	-.27	43	.7874
12	Speaker 13 Dhivehi-English Male 2	.71	.30	-.03	1 No Familiarity	.83	.15	.05	3 Some Familiarity	-.13	.33	-.39	40	.7003
12	Speaker 13 Dhivehi-English Male 2	.71	.30	-.03	1 No Familiarity	.84	.17	.05	2 Limited Familiarity	-.13	.35	-.38	42	.7039
9	Speaker 10 Arabic-English Male 2	.88	.30	-.12	1 No Familiarity	1.02	.23	-.04	4 Very Familiar	-.14	.37	-.37	44	.7161
3	Speaker 4 Dhivehi-English Female 2	-.49	.15	-.11	3 Some Familiarity	-.35	.23	-.03	4 Very Familiar	-.14	.27	-.51	79	.6132
2	Speaker 3 Arabic-English Female 1	-1.19	.19	-.10	2 Limited Familiarity	-1.04	.24	-.03	4 Very Familiar	-.14	.30	-.47	79	.6371
5	Speaker 6 Arabic-English Male 1	.44	.17	-.03	2 Limited Familiarity	.59	.15	.05	3 Some Familiarity	-.15	.23	-.64	125	.5202
11	Speaker 12 Arabic-English Female 2	.56	.17	-.03	2 Limited Familiarity	.72	.15	.07	3 Some Familiarity	-.16	.23	-.69	125	.4884
5	Speaker 6 Arabic-English Male 1	.25	.30	-.13	1 No Familiarity	.44	.17	-.03	2 Limited Familiarity	-.19	.35	-.54	42	.5936
11	Speaker 12 Arabic-English Female 2	.53	.30	-.04	1 No Familiarity	.72	.15	.07	3 Some Familiarity	-.20	.34	-.59	40	.5600
8	Speaker 9 Spanish-English Female 1	-.61	.32	.00	1 No Familiarity	-.42	.15	.11	3 Some Familiarity	-.20	.35	-.56	39	.5802
7	Speaker 8 Dhivehi-English Male 1	-.19	.15	-.08	3 Some Familiarity	.01	.22	.04	4 Very Familiar	-.20	.27	-.74	79	.4635
9	Speaker 10 Arabic-English Male 2	.88	.30	-.12	1 No Familiarity	1.08	.15	.00	3 Some Familiarity	-.20	.33	-.60	40	.5519
10	Speaker 11 Spanish-English Female 2	1.24	.18	.09	2 Limited Familiarity	1.50	.23	.22	4 Very Familiar	-.26	.29	-.88	77	.3806
5	Speaker 6 Arabic-English Male 1	.25	.30	-.13	1 No Familiarity	.51	.22	.01	4 Very Familiar	-.26	.38	-.69	44	.4953
4	Speaker 5 Spanish-English Male 1	-1.33	.19	-.11	2 Limited Familiarity	-1.06	.16	.02	3 Some Familiarity	-.27	.25	-1.10	124	.2742
10	Speaker 11 Spanish-English Female 2	.62	.30	-.27	1 No Familiarity	.93	.15	-.09	3 Some Familiarity	-.31	.34	-.92	40	.3616
9	Speaker 10 Arabic-English Male 2	.88	.30	-.12	1 No Familiarity	1.21	.18	.07	2 Limited Familiarity	-.32	.35	-.94	43	.3548
5	Speaker 6 Arabic-English Male 1	.25	.30	-.13	1 No Familiarity	.59	.15	.05	3 Some Familiarity	-.34	.34	-1.00	39	.3256
4	Speaker 5 Spanish-English Male 1	-1.33	.19	-.11	2 Limited Familiarity	-.99	.24	.06	4 Very Familiar	-.35	.30	-1.14	79	.2587
2	Speaker 3 Arabic-English Female 1	-1.15	.34	-.08	1 No Familiarity	-.77	.15	.11	3 Some Familiarity	-.39	.37	-1.04	38	.3046
2	Speaker 3 Arabic-English Female 1	-1.19	.19	-.10	2 Limited Familiarity	-.77	.15	.11	3 Some Familiarity	-.42	.25	-1.71	124	.0904
6	Speaker 7 Spanish-English Male 2	-.31	.31	.06	1 No Familiarity	.11	.22	.31	4 Very Familiar	-.42	.39	-1.10	43	.2762
8	Speaker 9 Spanish-English Female 1	-1.01	.19	-.20	2 Limited Familiarity	-.50	.23	.06	4 Very Familiar	-.50	.30	-1.70	79	.0921
10	Speaker 11 Spanish-English Female 2	.93	.15	-.09	3 Some Familiarity	1.50	.23	.22	4 Very Familiar	-.57	.28	-2.06	77	.0428
8	Speaker 9 Spanish-English Female 1	-1.01	.19	-.20	2 Limited Familiarity	-.42	.15	.11	3 Some Familiarity	-.59	.24	-2.45	124	.0157
10	Speaker 11 Spanish-English Female 2	.62	.30	-.27	1 No Familiarity	1.24	.18	.09	2 Limited Familiarity	-.62	.35	-1.79	42	.0804
6	Speaker 7 Spanish-English Male 2	-.60	.18	-.09	2 Limited Familiarity	.11	.22	.31	4 Very Familiar	-.71	.29	-2.45	79	.0164
6	Speaker 7 Spanish-English Male 2	-.60	.15	-.09	3 Some Familiarity	.11	.22	.31	4 Very Familiar	-.71	.27	-2.63	79	.0104
10	Speaker 11 Spanish-English Female 2	.62	.30	-.27	1 No Familiarity	1.50	.23	.22	4 Very Familiar	-.88	.38	-2.32	44	.0250

Rasch Analysis of Expert Ratings 08/01/2015 15:18:29
Table 14.3.1.3 Bias/Interaction Pairwise Report (arranged by mN).

Bias/Interaction: 2. Spanish English Familiarity Level, 3. Speaker

Target N Spanish English Fam	Target Measr	S.E.	Obs-Exp Average	Context Nu Speaker	Target Measr	S.E.	Obs-Exp Average	Context Nu Speaker	Target Contrast	Joint S.E.	t	Welch d.f.	Prob.
1 No Familiarity	.26	.31	.27	1 Speaker 2 Dhivehi-English Female 1	-.73	.30	-.27	10 Speaker 11 Spanish-English Female 2	.99	.43	2.31	37	.0266
4 Very Familiar	.68	.22	.31	6 Speaker 7 Spanish-English Male 2	-.16	.22	-.17	12 Speaker 13 Dhivehi-English Male 2	.84	.32	2.65	67	.0101
1 No Familiarity	.11	.31	.19	3 Speaker 4 Dhivehi-English Female 2	-.73	.30	-.27	10 Speaker 11 Spanish-English Female 2	.83	.43	1.95	37	.0590
1 No Familiarity	.26	.31	.27	1 Speaker 2 Dhivehi-English Female 1	-.50	.30	-.13	5 Speaker 6 Arabic-English Male 1	.76	.43	1.76	37	.0859
1 No Familiarity	.02	.33	.13	4 Speaker 5 Spanish-English Male 1	-.73	.30	-.27	10 Speaker 11 Spanish-English Female 2	.75	.44	1.69	37	.0988

1 No Familiarity	.26	.31	.27	1 Speaker 2 Dhivehi-English Female 1	-.46	.30	-.12	9 Speaker 10 Arabic-English Male 2	.72	.43	1.69	37	.0997
2 Limited Familiarity	.34	.18	.18	1 Speaker 2 Dhivehi-English Female 1	-.38	.19	-.20	8 Speaker 9 Spanish-English Female 1	.72	.26	2.78	111	.0064
4 Very Familiar	.54	.23	.22	10 Speaker 11 Spanish-English Female 2	-.16	.22	-.17	12 Speaker 13 Dhivehi-English Male 2	.70	.32	2.16	67	.0340
1 No Familiarity	.26	.31	.27	1 Speaker 2 Dhivehi-English Female 1	-.43	.34	-.08	2 Speaker 3 Arabic-English Female 1	.70	.46	1.53	37	.1357
4 Very Familiar	.68	.22	.31	6 Speaker 7 Spanish-English Male 2	-.02	.22	-.08	11 Speaker 12 Arabic-English Female 2	.69	.32	2.18	67	.0325
4 Very Familiar	.68	.22	.31	6 Speaker 7 Spanish-English Male 2	.06	.23	-.04	9 Speaker 10 Arabic-English Male 2	.62	.32	1.93	67	.0582
1 No Familiarity	.11	.31	.19	3 Speaker 4 Dhivehi-English Female 2	-.50	.30	-.13	5 Speaker 6 Arabic-English Male 1	.60	.43	1.41	37	.1683
1 No Familiarity	.26	.31	.27	1 Speaker 2 Dhivehi-English Female 1	-.34	.30	-.04	11 Speaker 12 Arabic-English Female 2	.60	.43	1.39	37	.1715
1 No Familiarity	-.13	.31	.06	6 Speaker 7 Spanish-English Male 2	-.73	.30	-.27	10 Speaker 11 Spanish-English Female 2	.60	.43	1.37	37	.1774
1 No Familiarity	.26	.31	.27	1 Speaker 2 Dhivehi-English Female 1	-.30	.30	-.03	12 Speaker 13 Dhivehi-English Male 2	.57	.43	1.32	37	.1941
1 No Familiarity	.11	.31	.19	3 Speaker 4 Dhivehi-English Female 2	-.46	.30	-.12	9 Speaker 10 Arabic-English Male 2	.57	.43	1.33	37	.1929
2 Limited Familiarity	.34	.18	.18	1 Speaker 2 Dhivehi-English Female 1	-.23	.19	-.11	4 Speaker 5 Spanish-English Male 1	.57	.26	2.16	111	.0333
2 Limited Familiarity	.18	.18	.10	3 Speaker 4 Dhivehi-English Female 2	-.38	.19	-.20	8 Speaker 9 Spanish-English Female 1	.56	.26	2.18	111	.0312
4 Very Familiar	.54	.23	.22	10 Speaker 11 Spanish-English Female 2	-.02	.22	-.08	11 Speaker 12 Arabic-English Female 2	.55	.32	1.71	67	.0920
2 Limited Familiarity	.34	.18	.18	1 Speaker 2 Dhivehi-English Female 1	-.21	.19	-.10	2 Speaker 3 Arabic-English Female 1	.54	.26	2.09	111	.0390
1 No Familiarity	.02	.33	.13	4 Speaker 5 Spanish-English Male 1	-.50	.30	-.13	5 Speaker 6 Arabic-English Male 1	.52	.45	1.17	37	.2496
2 Limited Familiarity	.13	.18	.07	7 Speaker 8 Dhivehi-English Male 1	-.38	.19	-.20	8 Speaker 9 Spanish-English Female 1	.52	.26	2.00	111	.0477
1 No Familiarity	.26	.31	.27	1 Speaker 2 Dhivehi-English Female 1	-.25	.32	.00	8 Speaker 9 Spanish-English Female 1	.51	.44	1.15	37	.2577
1 No Familiarity	-.22	.31	.02	7 Speaker 8 Dhivehi-English Male 1	-.73	.30	-.27	10 Speaker 11 Spanish-English Female 2	.50	.43	1.18	37	.2475
2 Limited Familiarity	.34	.18	.18	1 Speaker 2 Dhivehi-English Female 1	-.16	.18	-.09	6 Speaker 7 Spanish-English Male 2	.50	.25	1.95	111	.0542
1 No Familiarity	.26	.31	.27	1 Speaker 2 Dhivehi-English Female 1	-.22	.31	.02	7 Speaker 8 Dhivehi-English Male 1	.49	.43	1.12	38	.2700
1 No Familiarity	.02	.33	.13	4 Speaker 5 Spanish-English Male 1	-.46	.30	-.12	9 Speaker 10 Arabic-English Male 2	.48	.44	1.09	37	.2823
4 Very Familiar	.68	.22	.31	6 Speaker 7 Spanish-English Male 2	.19	.22	.04	7 Speaker 8 Dhivehi-English Male 1	.48	.32	1.51	67	.1350
1 No Familiarity	-.25	.32	.00	8 Speaker 9 Spanish-English Female 1	-.73	.30	-.27	10 Speaker 11 Spanish-English Female 2	.48	.44	1.10	37	.2804
1 No Familiarity	.11	.31	.19	3 Speaker 4 Dhivehi-English Female 2	-.34	.30	-.04	11 Speaker 12 Arabic-English Female 2	.44	.43	1.03	37	.3086
4 Very Familiar	.68	.22	.31	6 Speaker 7 Spanish-English Male 2	.25	.23	.06	8 Speaker 9 Spanish-English Female 1	.43	.32	1.34	67	.1857
4 Very Familiar	.25	.23	.06	8 Speaker 9 Spanish-English Female 1	-.16	.22	-.17	12 Speaker 13 Dhivehi-English Male 2	.41	.32	1.28	67	.2036
1 No Familiarity	.11	.31	.19	3 Speaker 4 Dhivehi-English Female 2	-.30	.30	-.03	12 Speaker 13 Dhivehi-English Male 2	.41	.43	.96	37	.3436
2 Limited Familiarity	.18	.18	.10	3 Speaker 4 Dhivehi-English Female 2	-.23	.19	-.11	4 Speaker 5 Spanish-English Male 1	.41	.26	1.56	111	.1205
4 Very Familiar	.24	.24	.06	4 Speaker 5 Spanish-English Male 1	-.16	.22	-.17	12 Speaker 13 Dhivehi-English Male 2	.41	.33	1.25	67	.2154
3 Some Familiarity	.33	.15	.11	2 Speaker 3 Arabic-English Female 1	-.07	.15	-.11	3 Speaker 4 Dhivehi-English Female 2	.40	.22	1.85	153	.0662
1 No Familiarity	.26	.31	.27	1 Speaker 2 Dhivehi-English Female 1	-.13	.31	.06	6 Speaker 7 Spanish-English Male 2	.40	.44	.90	37	.3735
2 Limited Familiarity	.34	.18	.18	1 Speaker 2 Dhivehi-English Female 1	-.05	.17	-.03	5 Speaker 6 Arabic-English Male 1	.39	.25	1.56	111	.1213
2 Limited Familiarity	.34	.18	.18	1 Speaker 2 Dhivehi-English Female 1	-.04	.17	-.03	11 Speaker 12 Arabic-English Female 2	.38	.25	1.51	111	.1327
3 Some Familiarity	.33	.15	.11	2 Speaker 3 Arabic-English Female 1	-.04	.15	-.09	6 Speaker 7 Spanish-English Male 2	.37	.22	1.72	153	.0878
3 Some Familiarity	.33	.15	.11	2 Speaker 3 Arabic-English Female 1	-.04	.15	-.09	10 Speaker 11 Spanish-English Female 2	.37	.22	1.72	153	.0867
3 Some Familiarity	.33	.15	.11	8 Speaker 9 Spanish-English Female 1	-.04	.15	-.09	10 Speaker 11 Spanish-English Female 2	.37	.21	1.73	153	.0860
4 Very Familiar	.19	.22	.04	7 Speaker 8 Dhivehi-English Male 1	-.16	.22	-.17	12 Speaker 13 Dhivehi-English Male 2	.36	.32	1.13	67	.2620
1 No Familiarity	.02	.33	.13	4 Speaker 5 Spanish-English Male 1	-.34	.30	-.04	11 Speaker 12 Arabic-English Female 2	.36	.44	.81	37	.4241
1 No Familiarity	.11	.31	.19	3 Speaker 4 Dhivehi-English Female 2	-.25	.32	.00	8 Speaker 9 Spanish-English Female 1	.35	.44	.80	37	.4297
3 Some Familiarity	.33	.15	.11	2 Speaker 3 Arabic-English Female 1	-.01	.15	-.08	7 Speaker 8 Dhivehi-English Male 1	.34	.22	1.58	153	.1165
2 Limited Familiarity	.18	.18	.10	3 Speaker 4 Dhivehi-English Female 2	-.16	.18	-.09	6 Speaker 7 Spanish-English Male 2	.34	.25	1.34	111	.1842
1 No Familiarity	.11	.31	.19	3 Speaker 4 Dhivehi-English Female 2	-.22	.31	.02	7 Speaker 8 Dhivehi-English Male 1	.33	.43	.76	37	.4516
2 Limited Familiarity	-.05	.17	-.03	5 Speaker 6 Arabic-English Male 1	-.38	.19	-.20	8 Speaker 9 Spanish-English Female 1	.33	.26	1.29	111	.2011
1 No Familiarity	-.13	.31	.06	6 Speaker 7 Spanish-English Male 2	-.46	.30	-.12	9 Speaker 10 Arabic-English Male 2	.33	.43	.76	37	.4529
1 No Familiarity	.02	.33	.13	4 Speaker 5 Spanish-English Male 1	-.30	.30	-.03	12 Speaker 13 Dhivehi-English Male 2	.33	.44	.74	37	.4657
4 Very Familiar	.14	.22	.01	5 Speaker 6 Arabic-English Male 1	-.16	.22	-.17	12 Speaker 13 Dhivehi-English Male 2	.31	.32	.97	68	.3353
1 No Familiarity	-.43	.34	-.08	2 Speaker 3 Arabic-English Female 1	-.73	.30	-.27	10 Speaker 11 Spanish-English Female 2	.29	.45	.65	37	.5188
1 No Familiarity	.02	.33	.13	4 Speaker 5 Spanish-English Male 1	-.25	.32	.00	8 Speaker 9 Spanish-English Female 1	.27	.46	.59	37	.5591
1 No Familiarity	-.46	.30	-.12	9 Speaker 10 Arabic-English Male 2	-.73	.30	-.27	10 Speaker 11 Spanish-English Female 2	.27	.42	.63	37	.5303
4 Very Familiar	.25	.23	.06	8 Speaker 9 Spanish-English Female 1	-.02	.22	-.08	11 Speaker 12 Arabic-English Female 2	.26	.32	.82	67	.4127
4 Very Familiar	.24	.24	.06	4 Speaker 5 Spanish-English Male 1	-.02	.22	-.08	11 Speaker 12 Arabic-English Female 2	.26	.33	.80	67	.4279
3 Some Familiarity	.22	.15	.05	5 Speaker 6 Arabic-English Male 1	-.04	.15	-.09	6 Speaker 7 Spanish-English Male 2	.26	.21	1.21	153	.2290
3 Some Familiarity	.22	.15	.05	5 Speaker 6 Arabic-English Male 1	-.04	.15	-.09	10 Speaker 11 Spanish-English Female 2	.26	.21	1.21	153	.2282
2 Limited Familiarity	.34	.18	.18	1 Speaker 2 Dhivehi-English Female 1	.09	.17	.05	12 Speaker 13 Dhivehi-English Male 2	.25	.25	1.00	111	.3177
1 No Familiarity	.02	.33	.13	4 Speaker 5 Spanish-English Male 1	-.22	.31	.02	7 Speaker 8 Dhivehi-English Male 1	.25	.45	.55	37	.5867
1 No Familiarity	.26	.31	.27	1 Speaker 2 Dhivehi-English Female 1	.02	.33	.13	4 Speaker 5 Spanish-English Male 1	.24	.45	.54	37	.5946
1 No Familiarity	.11	.31	.19	3 Speaker 4 Dhivehi-English Female 2	-.13	.31	.06	6 Speaker 7 Spanish-English Male 2	.24	.44	.54	37	.5890
4 Very Familiar	.07	.23	-.03	3 Speaker 4 Dhivehi-English Female 2	-.16	.22	-.17	12 Speaker 13 Dhivehi-English Male 2	.24	.32	.75	67	.4587
1 No Familiarity	-.22	.31	.02	7 Speaker 8 Dhivehi-English Male 1	-.46	.30	-.12	9 Speaker 10 Arabic-English Male 2	.24	.43	.55	37	.5840
2 Limited Familiarity	.18	.18	.10	3 Speaker 4 Dhivehi-English Female 2	-.05	.17	-.03	5 Speaker 6 Arabic-English Male 1	.23	.25	.94	111	.3498
1 No Familiarity	-.50	.30	-.13	5 Speaker 6 Arabic-English Male 1	-.73	.30	-.27	10 Speaker 11 Spanish-English Female 2	.23	.43	.54	37	.5941
4 Very Familiar	.06	.24	-.03	2 Speaker 3 Arabic-English Female 1	-.16	.22	-.17	12 Speaker 13 Dhivehi-English Male 2	.23	.33	.69	67	.4913
3 Some Familiarity	.22	.15	.05	5 Speaker 6 Arabic-English Male 1	-.01	.15	-.08	7 Speaker 8 Dhivehi-English Male 1	.22	.21	1.06	153	.2900
4 Very Familiar	.06	.23	-.04	9 Speaker 10 Arabic-English Male 2	-.16	.22	-.17	12 Speaker 13 Dhivehi-English Male 2	.22	.32	.70	67	.4847
2 Limited Familiarity	-.16	.18	-.09	6 Speaker 7 Spanish-English Male 2	-.38	.19	-.20	8 Speaker 9 Spanish-English Female 1	.22	.26	.86	111	.3937
2 Limited Familiarity	.18	.18	.10	3 Speaker 4 Dhivehi-English Female 2	-.04	.17	-.03	11 Speaker 12 Arabic-English Female 2	.22	.25	.89	111	.3744
2 Limited Familiarity	.34	.18	.18	1 Speaker 2 Dhivehi-English Female 1	.12	.18	.07	9 Speaker 10 Arabic-English Male 2	.22	.25	.86	111	.3928
1 No Familiarity	-.25	.32	.00	8 Speaker 9 Spanish-English Female 1	-.46	.30	-.12	9 Speaker 10 Arabic-English Male 2	.21	.44	.49	37	.6298
3 Some Familiarity	.33	.15	.11	2 Speaker 3 Arabic-English Female 1	.12	.15	.00	9 Speaker 10 Arabic-English Male 2	.21	.22	.98	153	.3273
4 Very Familiar	.19	.22	.04	7 Speaker 8 Dhivehi-English Male 1	-.02	.22	-.08	11 Speaker 12 Arabic-English Female 2	.21	.32	.67	67	.5067
3 Some Familiarity	.33	.15	.11	8 Speaker 9 Spanish-English Female 1	.12	.15	.00	9 Speaker 10 Arabic-English Male 2	.21	.21	.98	153	.3287

3 Some Familiarity	.17	.16	.02	4 Speaker 5 Spanish-English Male 1	-.04	.15	-.09	6 Speaker 7 Spanish-English Male 2	.21	.22	.94	153	.3470
3 Some Familiarity	.17	.16	.02	4 Speaker 5 Spanish-English Male 1	-.04	.15	-.09	10 Speaker 11 Spanish-English Female 2	.21	.22	.94	153	.3474
2 Limited Familiarity	.34	.18	.18	1 Speaker 2 Dhivehi-English Female 1	.13	.18	.07	7 Speaker 8 Dhivehi-English Male 1	.20	.25	.82	111	.4167
1 No Familiarity	-.13	.31	.06	6 Speaker 7 Spanish-English Male 2	-.34	.30	-.04	11 Speaker 12 Arabic-English Female 2	.20	.43	.47	37	.6409
2 Limited Familiarity	.15	.18	.09	10 Speaker 11 Spanish-English Female 2	-.04	.17	-.03	11 Speaker 12 Arabic-English Female 2	.19	.25	.78	111	.4367
4 Very Familiar	.25	.23	.06	8 Speaker 9 Spanish-English Female 1	.06	.23	-.04	9 Speaker 10 Arabic-English Male 2	.19	.32	.58	67	.5640
2 Limited Familiarity	.34	.18	.18	1 Speaker 2 Dhivehi-English Female 1	.15	.18	.09	10 Speaker 11 Spanish-English Female 2	.18	.25	.73	111	.4650
4 Very Familiar	.24	.24	.06	4 Speaker 5 Spanish-English Male 1	.06	.23	-.04	9 Speaker 10 Arabic-English Male 2	.18	.33	.56	67	.5794
2 Limited Familiarity	-.21	.19	-.10	2 Speaker 3 Arabic-English Female 1	-.38	.19	-.20	8 Speaker 9 Spanish-English Female 1	.17	.27	.65	111	.5145
3 Some Familiarity	.17	.16	.02	4 Speaker 5 Spanish-English Male 1	-.01	.15	-.08	7 Speaker 8 Dhivehi-English Male 1	.17	.22	.80	153	.4257
2 Limited Familiarity	.13	.18	.07	7 Speaker 8 Dhivehi-English Male 1	-.04	.17	-.03	11 Speaker 12 Arabic-English Female 2	.17	.25	.70	111	.4854
1 No Familiarity	-.13	.31	.06	6 Speaker 7 Spanish-English Male 2	-.30	.30	-.03	12 Speaker 13 Dhivehi-English Male 2	.17	.43	.40	37	.6937
3 Some Familiarity	.33	.15	.11	2 Speaker 3 Arabic-English Female 1	.17	.16	.02	4 Speaker 5 Spanish-English Male 1	.17	.22	.75	153	.4522
2 Limited Familiarity	.12	.18	.07	9 Speaker 10 Arabic-English Male 2	-.04	.17	-.03	11 Speaker 12 Arabic-English Female 2	.16	.25	.66	111	.5135
4 Very Familiar	.14	.22	.01	5 Speaker 6 Arabic-English Male 1	-.02	.22	-.08	11 Speaker 12 Arabic-English Female 2	.16	.32	.51	67	.6144
3 Some Familiarity	.12	.15	.00	9 Speaker 10 Arabic-English Male 2	-.04	.15	-.09	10 Speaker 11 Spanish-English Female 2	.16	.21	.75	153	.4568
1 No Familiarity	.26	.31	.27	1 Speaker 2 Dhivehi-English Female 1	.11	.31	.19	3 Speaker 4 Dhivehi-English Female 2	.16	.43	.36	37	.7198
2 Limited Familiarity	.34	.18	.18	1 Speaker 2 Dhivehi-English Female 1	.18	.18	.10	3 Speaker 4 Dhivehi-English Female 2	.16	.25	.62	111	.5366
1 No Familiarity	.02	.33	.13	4 Speaker 5 Spanish-English Male 1	-.13	.31	.06	6 Speaker 7 Spanish-English Male 2	.15	.45	.34	37	.7346
2 Limited Familiarity	-.23	.19	-.11	4 Speaker 5 Spanish-English Male 1	-.38	.19	-.20	8 Speaker 9 Spanish-English Female 1	.15	.27	.57	111	.5701
4 Very Familiar	-.02	.22	-.08	11 Speaker 12 Arabic-English Female 2	-.16	.22	-.17	12 Speaker 13 Dhivehi-English Male 2	.15	.32	.46	68	.6439
4 Very Familiar	.68	.22	.31	6 Speaker 7 Spanish-English Male 2	.54	.23	.22	10 Speaker 11 Spanish-English Female 2	.14	.32	.43	67	.6684
4 Very Familiar	.19	.22	.04	7 Speaker 8 Dhivehi-English Male 1	.06	.23	-.04	9 Speaker 10 Arabic-English Male 2	.13	.32	.42	67	.6743
3 Some Familiarity	.33	.15	.11	2 Speaker 3 Arabic-English Female 1	.20	.15	.05	12 Speaker 13 Dhivehi-English Male 2	.13	.22	.59	153	.5552
3 Some Familiarity	.33	.15	.11	8 Speaker 9 Spanish-English Female 1	.20	.15	.05	12 Speaker 13 Dhivehi-English Male 2	.12	.21	.58	153	.5596
3 Some Familiarity	.33	.15	.11	2 Speaker 3 Arabic-English Female 1	.22	.15	.05	5 Speaker 6 Arabic-English Male 1	.12	.21	.54	153	.5917
1 No Familiarity	-.13	.31	.06	6 Speaker 7 Spanish-English Male 2	-.25	.32	.00	8 Speaker 9 Spanish-English Female 1	.12	.45	.26	37	.7985
1 No Familiarity	-.22	.31	.02	7 Speaker 8 Dhivehi-English Male 1	-.34	.30	-.04	11 Speaker 12 Arabic-English Female 2	.11	.43	.26	37	.7947
3 Some Familiarity	.04	.15	-.04	1 Speaker 2 Dhivehi-English Female 1	-.07	.15	-.11	3 Speaker 4 Dhivehi-English Female 2	.11	.22	.51	153	.6077
2 Limited Familiarity	-.05	.17	-.03	5 Speaker 6 Arabic-English Male 1	-.16	.18	-.09	6 Speaker 7 Spanish-English Male 2	.11	.25	.42	111	.6757
4 Very Familiar	.24	.24	.06	4 Speaker 5 Spanish-English Male 1	.14	.22	.01	5 Speaker 6 Arabic-English Male 1	.10	.33	.30	67	.7623
3 Some Familiarity	.22	.15	.05	5 Speaker 6 Arabic-English Male 1	.12	.15	.00	9 Speaker 10 Arabic-English Male 2	.10	.21	.46	153	.6483
2 Limited Familiarity	.18	.18	.10	3 Speaker 4 Dhivehi-English Female 2	.09	.17	.05	12 Speaker 13 Dhivehi-English Male 2	.09	.25	.38	111	.7046
3 Some Familiarity	.33	.15	.11	2 Speaker 3 Arabic-English Female 1	.24	.15	.07	11 Speaker 12 Arabic-English Female 2	.09	.21	.43	153	.6693
1 No Familiarity	-.13	.31	.06	6 Speaker 7 Spanish-English Male 2	-.22	.31	.02	7 Speaker 8 Dhivehi-English Male 1	.09	.44	.21	37	.8362
4 Very Familiar	.07	.23	-.03	3 Speaker 4 Dhivehi-English Female 2	-.02	.22	-.08	11 Speaker 12 Arabic-English Female 2	.09	.32	.28	67	.7772
3 Some Familiarity	.33	.15	.11	8 Speaker 9 Spanish-English Female 1	.24	.15	.07	11 Speaker 12 Arabic-English Female 2	.09	.21	.42	153	.6753
1 No Familiarity	-.25	.32	.00	8 Speaker 9 Spanish-English Female 1	-.34	.30	-.04	11 Speaker 12 Arabic-English Female 2	.09	.44	.20	37	.8408
1 No Familiarity	.11	.31	.19	3 Speaker 4 Dhivehi-English Female 2	.02	.33	.13	4 Speaker 5 Spanish-English Male 1	.08	.45	.19	37	.8521
4 Very Familiar	.14	.22	.01	5 Speaker 6 Arabic-English Male 1	.06	.23	-.04	9 Speaker 10 Arabic-English Male 2	.08	.32	.26	67	.7946
3 Some Familiarity	.04	.15	-.04	1 Speaker 2 Dhivehi-English Female 1	-.04	.15	-.09	6 Speaker 7 Spanish-English Male 2	.08	.22	.39	153	.7004
3 Some Familiarity	.04	.15	-.04	1 Speaker 2 Dhivehi-English Female 1	-.04	.15	-.09	10 Speaker 11 Spanish-English Female 2	.08	.21	.38	153	.7045
1 No Familiarity	-.22	.31	.02	7 Speaker 8 Dhivehi-English Male 1	-.30	.30	-.03	12 Speaker 13 Dhivehi-English Male 2	.08	.43	.19	37	.8522
4 Very Familiar	.06	.24	-.03	2 Speaker 3 Arabic-English Female 1	-.02	.22	-.08	11 Speaker 12 Arabic-English Female 2	.08	.33	.24	67	.8110
4 Very Familiar	.06	.23	-.04	9 Speaker 10 Arabic-English Male 2	-.02	.22	-.08	11 Speaker 12 Arabic-English Female 2	.08	.32	.24	67	.8100
2 Limited Familiarity	.15	.18	.09	10 Speaker 11 Spanish-English Female 2	.09	.17	.05	12 Speaker 13 Dhivehi-English Male 2	.07	.25	.27	111	.7893
1 No Familiarity	-.43	.34	-.08	2 Speaker 3 Arabic-English Female 1	-.50	.30	-.13	5 Speaker 6 Arabic-English Male 1	.07	.45	.14	37	.8868
2 Limited Familiarity	.18	.18	.10	3 Speaker 4 Dhivehi-English Female 2	.12	.18	.07	9 Speaker 10 Arabic-English Male 2	.06	.25	.24	111	.8132
1 No Familiarity	-.25	.32	.00	8 Speaker 9 Spanish-English Female 1	-.30	.30	-.03	12 Speaker 13 Dhivehi-English Male 2	.06	.44	.13	37	.8979
3 Some Familiarity	.04	.15	-.04	1 Speaker 2 Dhivehi-English Female 1	-.01	.15	-.08	7 Speaker 8 Dhivehi-English Male 1	.05	.21	.23	153	.8154
2 Limited Familiarity	.18	.18	.10	3 Speaker 4 Dhivehi-English Female 2	.13	.18	.07	7 Speaker 8 Dhivehi-English Male 1	.05	.25	.19	111	.8469
4 Very Familiar	.24	.24	.06	4 Speaker 5 Spanish-English Male 1	.19	.22	.04	7 Speaker 8 Dhivehi-English Male 1	.05	.33	.15	67	.8850
2 Limited Familiarity	.13	.18	.07	7 Speaker 8 Dhivehi-English Male 1	.09	.17	.05	12 Speaker 13 Dhivehi-English Male 2	.05	.25	.19	111	.8523
3 Some Familiarity	.17	.16	.02	4 Speaker 5 Spanish-English Male 1	.12	.15	.00	9 Speaker 10 Arabic-English Male 2	.05	.22	.21	153	.8328
2 Limited Familiarity	.12	.18	.07	9 Speaker 10 Arabic-English Male 2	.09	.17	.05	12 Speaker 13 Dhivehi-English Male 2	.04	.25	.14	111	.8869
3 Some Familiarity	.24	.15	.07	11 Speaker 12 Arabic-English Female 2	.20	.15	.05	12 Speaker 13 Dhivehi-English Male 2	.04	.21	.17	153	.8678
3 Some Familiarity	-.01	.15	-.08	7 Speaker 8 Dhivehi-English Male 1	-.04	.15	-.09	10 Speaker 11 Spanish-English Female 2	.03	.21	.15	153	.8829
2 Limited Familiarity	.18	.18	.10	3 Speaker 4 Dhivehi-English Female 2	.15	.18	.09	10 Speaker 11 Spanish-English Female 2	.03	.25	.11	111	.9108
1 No Familiarity	-.43	.34	-.08	2 Speaker 3 Arabic-English Female 1	-.46	.30	-.12	9 Speaker 10 Arabic-English Male 2	.03	.45	.06	37	.9535
1 No Familiarity	-.22	.31	.02	7 Speaker 8 Dhivehi-English Male 1	-.25	.32	.00	8 Speaker 9 Spanish-English Female 1	.02	.44	.05	37	.9575
2 Limited Familiarity	-.21	.19	-.10	2 Speaker 3 Arabic-English Female 1	-.23	.19	-.11	4 Speaker 5 Spanish-English Male 1	.02	.27	.08	111	.9370
4 Very Familiar	.07	.23	-.03	3 Speaker 4 Dhivehi-English Female 2	.06	.23	-.04	9 Speaker 10 Arabic-English Male 2	.01	.32	.04	68	.9663
3 Some Familiarity	.22	.15	.05	5 Speaker 6 Arabic-English Male 1	.20	.15	.05	12 Speaker 13 Dhivehi-English Male 2	.01	.21	.06	153	.9553
2 Limited Familiarity	.13	.18	.07	7 Speaker 8 Dhivehi-English Male 1	.12	.18	.07	9 Speaker 10 Arabic-English Male 2	.01	.25	.04	111	.9652
3 Some Familiarity	.33	.15	.11	2 Speaker 3 Arabic-English Female 1	.33	.15	.11	8 Speaker 9 Spanish-English Female 1	.00	.22	.01	153	.9903
4 Very Familiar	.06	.24	-.03	2 Speaker 3 Arabic-English Female 1	.06	.23	-.04	9 Speaker 10 Arabic-English Male 2	.00	.33	.00	67	.9973
3 Some Familiarity	-.04	.15	-.09	6 Speaker 7 Spanish-English Male 2	-.04	.15	-.09	10 Speaker 11 Spanish-English Female 2	.00	.21	-.01	153	.9930
4 Very Familiar	.24	.24	.06	4 Speaker 5 Spanish-English Male 1	.25	.23	.06	8 Speaker 9 Spanish-English Female 1	.00	.33	-.01	67	.9885
2 Limited Familiarity	-.05	.17	-.03	5 Speaker 6 Arabic-English Male 1	-.04	.17	-.03	11 Speaker 12 Arabic-English Female 2	-.01	.25	-.05	111	.9619
4 Very Familiar	.06	.24	-.03	2 Speaker 3 Arabic-English Female 1	.07	.23	-.03	3 Speaker 4 Dhivehi-English Female 2	-.01	.33	-.04	67	.9697
2 Limited Familiarity	.13	.18	.07	7 Speaker 8 Dhivehi-English Male 1	.15	.18	.09	10 Speaker 11 Spanish-English Female 2	-.02	.25	-.08	111	.9354
3 Some Familiarity	.22	.15	.05	5 Speaker 6 Arabic-English Male 1	.24	.15	.07	11 Speaker 12 Arabic-English Female 2	-.02	.21	-.11	153	.9118

3 Some Familiarity	-.07	.15	-.11	3 Speaker 4 Dhivehi-English Female 2	-.04	.15	-.09	6 Speaker 7 Spanish-English Male 2	-.03	.22	-.13	153	.8984
3 Some Familiarity	-.07	.15	-.11	3 Speaker 4 Dhivehi-English Female 2	-.04	.15	-.09	10 Speaker 11 Spanish-English Female 2	-.03	.21	-.14	153	.8905
2 Limited Familiarity	.12	.18	.07	9 Speaker 10 Arabic-English Male 2	.15	.18	.09	10 Speaker 11 Spanish-English Female 2	-.03	.25	-.12	111	.9009
1 No Familiarity	-.34	.30	-.04	11 Speaker 12 Arabic-English Female 2	-.30	.30	-.03	12 Speaker 13 Dhivehi-English Male 2	-.03	.42	-.08	37	.9399
3 Some Familiarity	-.04	.15	-.09	6 Speaker 7 Spanish-English Male 2	-.01	.15	-.08	7 Speaker 8 Dhivehi-English Male 1	-.03	.21	-.16	153	.8770
1 No Familiarity	-.50	.30	-.13	5 Speaker 6 Arabic-English Male 1	-.46	.30	-.12	9 Speaker 10 Arabic-English Male 2	-.04	.43	-.09	37	.9281
3 Some Familiarity	.17	.16	.02	4 Speaker 5 Spanish-English Male 1	.20	.15	.05	12 Speaker 13 Dhivehi-English Male 2	-.04	.22	-.18	153	.8579
2 Limited Familiarity	-.21	.19	-.10	2 Speaker 3 Arabic-English Female 1	-.16	.18	-.09	6 Speaker 7 Spanish-English Male 2	-.05	.26	-.19	111	.8527
3 Some Familiarity	.17	.16	.02	4 Speaker 5 Spanish-English Male 1	.22	.15	.05	5 Speaker 6 Arabic-English Male 1	-.05	.22	-.23	153	.8149
4 Very Familiar	.14	.22	.01	5 Speaker 6 Arabic-English Male 1	.19	.22	.04	7 Speaker 8 Dhivehi-English Male 1	-.05	.32	-.16	67	.8717
4 Very Familiar	.19	.22	.04	7 Speaker 8 Dhivehi-English Male 1	.25	.23	.06	8 Speaker 9 Spanish-English Female 1	-.05	.32	-.16	67	.8716
3 Some Familiarity	-.07	.15	-.11	3 Speaker 4 Dhivehi-English Female 2	-.01	.15	-.08	7 Speaker 8 Dhivehi-English Male 1	-.06	.21	-.28	153	.7763
4 Very Familiar	.07	.23	-.03	3 Speaker 4 Dhivehi-English Female 2	.14	.22	.01	5 Speaker 6 Arabic-English Male 1	-.07	.32	-.22	67	.8275
2 Limited Familiarity	-.23	.19	-.11	4 Speaker 5 Spanish-English Male 1	-.16	.18	-.09	6 Speaker 7 Spanish-English Male 2	-.07	.27	-.27	111	.7909
3 Some Familiarity	.17	.16	.02	4 Speaker 5 Spanish-English Male 1	.24	.15	.07	11 Speaker 12 Arabic-English Female 2	-.07	.22	-.34	153	.7328
3 Some Familiarity	.04	.15	-.04	1 Speaker 2 Dhivehi-English Female 1	.12	.15	.00	9 Speaker 10 Arabic-English Male 2	-.08	.22	-.36	153	.7192
4 Very Familiar	.06	.24	-.03	2 Speaker 3 Arabic-English Female 1	.14	.22	.01	5 Speaker 6 Arabic-English Male 1	-.08	.33	-.25	67	.8013
3 Some Familiarity	.12	.15	.00	9 Speaker 10 Arabic-English Male 2	.20	.15	.05	12 Speaker 13 Dhivehi-English Male 2	-.09	.21	-.40	153	.6896
1 No Familiarity	-.43	.34	-.08	2 Speaker 3 Arabic-English Female 1	-.34	.30	-.04	11 Speaker 12 Arabic-English Female 2	-.10	.45	-.22	37	.8304
4 Very Familiar	.14	.22	.01	5 Speaker 6 Arabic-English Male 1	.25	.23	.06	8 Speaker 9 Spanish-English Female 1	-.10	.32	-.32	67	.7476
3 Some Familiarity	.22	.15	.05	5 Speaker 6 Arabic-English Male 1	.33	.15	.11	8 Speaker 9 Spanish-English Female 1	-.11	.21	-.53	153	.5966
2 Limited Familiarity	-.16	.18	-.09	6 Speaker 7 Spanish-English Male 2	-.04	.17	-.03	11 Speaker 12 Arabic-English Female 2	-.12	.25	-.47	111	.6418
3 Some Familiarity	.12	.15	.00	9 Speaker 10 Arabic-English Male 2	.24	.15	.07	11 Speaker 12 Arabic-English Female 2	-.12	.21	-.57	153	.5718
4 Very Familiar	.07	.23	-.03	3 Speaker 4 Dhivehi-English Female 2	.19	.22	.04	7 Speaker 8 Dhivehi-English Male 1	-.12	.32	-.38	67	.7055
3 Some Familiarity	.04	.15	-.04	1 Speaker 2 Dhivehi-English Female 1	.17	.16	.02	4 Speaker 5 Spanish-English Male 1	-.12	.22	-.56	153	.5743
1 No Familiarity	-.46	.30	-.12	9 Speaker 10 Arabic-English Male 2	-.34	.30	-.04	11 Speaker 12 Arabic-English Female 2	-.12	.42	-.29	37	.7712
2 Limited Familiarity	-.04	.17	-.03	11 Speaker 12 Arabic-English Female 2	.09	.17	.05	12 Speaker 13 Dhivehi-English Male 2	-.13	.25	-.52	111	.6074
3 Some Familiarity	-.01	.15	-.08	7 Speaker 8 Dhivehi-English Male 1	.12	.15	.00	9 Speaker 10 Arabic-English Male 2	-.13	.21	-.60	153	.5500
1 No Familiarity	-.43	.34	-.08	2 Speaker 3 Arabic-English Female 1	-.30	.30	-.03	12 Speaker 13 Dhivehi-English Male 2	-.13	.45	-.29	37	.7754
4 Very Familiar	.06	.24	-.03	2 Speaker 3 Arabic-English Female 1	.19	.22	.04	7 Speaker 8 Dhivehi-English Male 1	-.13	.33	-.41	67	.6830
2 Limited Familiarity	-.05	.17	-.03	5 Speaker 6 Arabic-English Male 1	.09	.17	.05	12 Speaker 13 Dhivehi-English Male 2	-.14	.25	-.56	112	.5747
2 Limited Familiarity	-.21	.19	-.10	2 Speaker 3 Arabic-English Female 1	-.05	.17	-.03	5 Speaker 6 Arabic-English Male 1	-.15	.26	-.60	111	.5502
1 No Familiarity	-.46	.30	-.12	9 Speaker 10 Arabic-English Male 2	-.30	.30	-.03	12 Speaker 13 Dhivehi-English Male 2	-.16	.42	-.37	37	.7136
3 Some Familiarity	-.04	.15	-.09	6 Speaker 7 Spanish-English Male 2	.12	.15	.00	9 Speaker 10 Arabic-English Male 2	-.16	.22	-.75	153	.4555
1 No Familiarity	-.50	.30	-.13	5 Speaker 6 Arabic-English Male 1	-.34	.30	-.04	11 Speaker 12 Arabic-English Female 2	-.16	.43	-.38	37	.7053
3 Some Familiarity	.04	.15	-.04	1 Speaker 2 Dhivehi-English Female 1	.20	.15	.05	12 Speaker 13 Dhivehi-English Male 2	-.16	.21	-.76	153	.4488
3 Some Familiarity	.17	.16	.02	4 Speaker 5 Spanish-English Male 1	.33	.15	.11	8 Speaker 9 Spanish-English Female 1	-.16	.22	-.75	153	.4553
2 Limited Familiarity	-.21	.19	-.10	2 Speaker 3 Arabic-English Female 1	-.04	.17	-.03	11 Speaker 12 Arabic-English Female 2	-.17	.26	-.65	111	.5201
4 Very Familiar	.07	.23	-.03	3 Speaker 4 Dhivehi-English Female 2	.24	.24	.06	4 Speaker 5 Spanish-English Male 1	-.17	.33	-.52	67	.6080
4 Very Familiar	.07	.23	-.03	3 Speaker 4 Dhivehi-English Female 2	.25	.23	.06	8 Speaker 9 Spanish-English Female 1	-.17	.32	-.54	67	.5926
2 Limited Familiarity	-.05	.17	-.03	5 Speaker 6 Arabic-English Male 1	.12	.18	.07	9 Speaker 10 Arabic-English Male 2	-.17	.25	-.70	111	.4837
3 Some Familiarity	.04	.15	-.04	1 Speaker 2 Dhivehi-English Female 1	.22	.15	.05	5 Speaker 6 Arabic-English Male 1	-.17	.21	-.82	153	.4153
2 Limited Familiarity	-.23	.19	-.11	4 Speaker 5 Spanish-English Male 1	-.05	.17	-.03	5 Speaker 6 Arabic-English Male 1	-.18	.26	-.68	111	.4995
4 Very Familiar	.06	.24	-.03	2 Speaker 3 Arabic-English Female 1	.24	.24	.06	4 Speaker 5 Spanish-English Male 1	-.18	.33	-.54	67	.5890
2 Limited Familiarity	-.05	.17	-.03	5 Speaker 6 Arabic-English Male 1	.13	.18	.07	7 Speaker 8 Dhivehi-English Male 1	-.19	.25	-.75	111	.4565
4 Very Familiar	.06	.24	-.03	2 Speaker 3 Arabic-English Female 1	.25	.23	.06	8 Speaker 9 Spanish-English Female 1	-.19	.33	-.57	67	.5739
1 No Familiarity	-.43	.34	-.08	2 Speaker 3 Arabic-English Female 1	-.25	.32	.00	8 Speaker 9 Spanish-English Female 1	-.19	.47	-.40	37	.6914
2 Limited Familiarity	-.23	.19	-.11	4 Speaker 5 Spanish-English Male 1	-.04	.17	-.03	11 Speaker 12 Arabic-English Female 2	-.19	.26	-.72	111	.4711
3 Some Familiarity	-.07	.15	-.11	3 Speaker 4 Dhivehi-English Female 2	.12	.15	.00	9 Speaker 10 Arabic-English Male 2	-.19	.21	-.88	153	.3807
1 No Familiarity	-.50	.30	-.13	5 Speaker 6 Arabic-English Male 1	-.30	.30	-.03	12 Speaker 13 Dhivehi-English Male 2	-.19	.43	-.46	37	.6500
3 Some Familiarity	.04	.15	-.04	1 Speaker 2 Dhivehi-English Female 1	.24	.15	.07	11 Speaker 12 Arabic-English Female 2	-.20	.21	-.93	153	.3563
2 Limited Familiarity	-.05	.17	-.03	5 Speaker 6 Arabic-English Male 1	.15	.18	.09	10 Speaker 11 Spanish-English Female 2	-.21	.25	-.83	111	.4096
1 No Familiarity	-.43	.34	-.08	2 Speaker 3 Arabic-English Female 1	-.22	.31	.02	7 Speaker 8 Dhivehi-English Male 1	-.21	.46	-.46	37	.6482
3 Some Familiarity	-.01	.15	-.08	7 Speaker 8 Dhivehi-English Male 1	.20	.15	.05	12 Speaker 13 Dhivehi-English Male 2	-.21	.21	-1.00	153	.3173
3 Some Familiarity	-.07	.15	-.11	3 Speaker 4 Dhivehi-English Female 2	.17	.16	.02	4 Speaker 5 Spanish-English Male 1	-.23	.22	-1.07	153	.2854
3 Some Familiarity	-.04	.15	-.09	10 Speaker 11 Spanish-English Female 2	.20	.15	.05	12 Speaker 13 Dhivehi-English Male 2	-.24	.21	-1.15	153	.2516
2 Limited Familiarity	-.16	.18	-.09	6 Speaker 7 Spanish-English Male 2	.09	.17	.05	12 Speaker 13 Dhivehi-English Male 2	-.24	.25	-.97	111	.3339
3 Some Familiarity	-.04	.15	-.09	6 Speaker 7 Spanish-English Male 2	.20	.15	.05	12 Speaker 13 Dhivehi-English Male 2	-.25	.21	-1.15	153	.2522
3 Some Familiarity	-.01	.15	-.08	7 Speaker 8 Dhivehi-English Male 1	.24	.15	.07	11 Speaker 12 Arabic-English Female 2	-.25	.21	-1.17	153	.2434
1 No Familiarity	-.50	.30	-.13	5 Speaker 6 Arabic-English Male 1	-.25	.32	.00	8 Speaker 9 Spanish-English Female 1	-.25	.44	-.57	37	.5721
3 Some Familiarity	-.07	.15	-.11	3 Speaker 4 Dhivehi-English Female 2	.20	.15	.05	12 Speaker 13 Dhivehi-English Male 2	-.27	.21	-1.28	153	.2017
1 No Familiarity	-.50	.30	-.13	5 Speaker 6 Arabic-English Male 1	-.22	.31	.02	7 Speaker 8 Dhivehi-English Male 1	-.28	.43	-.64	37	.5277
3 Some Familiarity	-.04	.15	-.09	10 Speaker 11 Spanish-English Female 2	.24	.15	.07	11 Speaker 12 Arabic-English Female 2	-.28	.21	-1.32	153	.1892
2 Limited Familiarity	-.16	.18	-.09	6 Speaker 7 Spanish-English Male 2	.12	.18	.07	9 Speaker 10 Arabic-English Male 2	-.28	.25	-1.11	111	.2711
3 Some Familiarity	-.04	.15	-.09	6 Speaker 7 Spanish-English Male 2	.24	.15	.07	11 Speaker 12 Arabic-English Female 2	-.28	.21	-1.32	153	.1901
3 Some Familiarity	-.07	.15	-.11	3 Speaker 4 Dhivehi-English Female 2	.22	.15	.05	5 Speaker 6 Arabic-English Male 1	-.29	.21	-1.34	153	.1819
3 Some Familiarity	.04	.15	-.04	1 Speaker 2 Dhivehi-English Female 1	.33	.15	.11	8 Speaker 9 Spanish-English Female 1	-.29	.22	-1.33	153	.1843
4 Very Familiar	.25	.23	.06	8 Speaker 9 Spanish-English Female 1	.54	.23	.22	10 Speaker 11 Spanish-English Female 2	-.29	.33	-.89	67	.3793
3 Some Familiarity	.04	.15	-.04	1 Speaker 2 Dhivehi-English Female 1	.33	.15	.11	2 Speaker 3 Arabic-English Female 1	-.29	.22	-1.33	153	.1845
2 Limited Familiarity	-.16	.18	-.09	6 Speaker 7 Spanish-English Male 2	.13	.18	.07	7 Speaker 8 Dhivehi-English Male 1	-.29	.25	-1.15	111	.2527
4 Very Familiar	.24	.24	.06	4 Speaker 5 Spanish-English Male 1	.54	.23	.22	10 Speaker 11 Spanish-English Female 2	-.29	.33	-.89	67	.3783
2 Limited Familiarity	-.21	.19	-.10	2 Speaker 3 Arabic-English Female 1	.09	.17	.05	12 Speaker 13 Dhivehi-English Male 2	-.29	.26	-1.14	111	.2578

1 No Familiarity	-.43	.34	-.08	2 Speaker 3 Arabic-English Female 1	-.13	.31	.06	6 Speaker 7 Spanish-English Male 2	-.30	.46	-.65	37	.5168
3 Some Familiarity	-.07	.15	-.11	3 Speaker 4 Dhivehi-English Female 2	.24	.15	.07	11 Speaker 12 Arabic-English Female 2	-.31	.21	-1.45	153	.1493
2 Limited Familiarity	-.16	.18	-.09	6 Speaker 7 Spanish-English Male 2	.15	.18	.09	10 Speaker 11 Spanish-English Female 2	-.31	.25	-1.23	111	.2220
2 Limited Familiarity	-.23	.19	-.11	4 Speaker 5 Spanish-English Male 1	.09	.17	.05	12 Speaker 13 Dhivehi-English Male 2	-.32	.26	-1.21	111	.2282
2 Limited Familiarity	-.21	.19	-.10	2 Speaker 3 Arabic-English Female 1	.12	.18	.07	9 Speaker 10 Arabic-English Male 2	-.33	.26	-1.27	111	.2071
3 Some Familiarity	-.01	.15	-.08	7 Speaker 8 Dhivehi-English Male 1	.33	.15	.11	8 Speaker 9 Spanish-English Female 1	-.34	.21	-1.58	153	.1158
2 Limited Familiarity	-.21	.19	-.10	2 Speaker 3 Arabic-English Female 1	.13	.18	.07	7 Speaker 8 Dhivehi-English Male 1	-.34	.26	-1.31	111	.1922
4 Very Familiar	.19	.22	.04	7 Speaker 8 Dhivehi-English Male 1	.54	.23	.22	10 Speaker 11 Spanish-English Female 2	-.34	.32	-1.05	67	.2961
2 Limited Familiarity	-.38	.19	-.20	8 Speaker 9 Spanish-English Female 1	-.04	.17	-.03	11 Speaker 12 Arabic-English Female 2	-.34	.26	-1.33	111	.1854
2 Limited Familiarity	-.23	.19	-.11	4 Speaker 5 Spanish-English Male 1	.12	.18	.07	9 Speaker 10 Arabic-English Male 2	-.35	.26	-1.34	111	.1824
2 Limited Familiarity	-.21	.19	-.10	2 Speaker 3 Arabic-English Female 1	.15	.18	.09	10 Speaker 11 Spanish-English Female 2	-.36	.26	-1.39	111	.1678
2 Limited Familiarity	-.23	.19	-.11	4 Speaker 5 Spanish-English Male 1	.13	.18	.07	7 Speaker 8 Dhivehi-English Male 1	-.36	.26	-1.38	111	.1689
1 No Familiarity	-.50	.30	-.13	5 Speaker 6 Arabic-English Male 1	-.13	.31	.06	6 Speaker 7 Spanish-English Male 2	-.37	.44	-.84	37	.4056
3 Some Familiarity	-.04	.15	-.09	6 Speaker 7 Spanish-English Male 2	.33	.15	.11	8 Speaker 9 Spanish-English Female 1	-.37	.22	-1.72	153	.0871
4 Very Familiar	-.54	.24	-.35	1 Speaker 2 Dhivehi-English Female 1	-.16	.22	-.17	12 Speaker 13 Dhivehi-English Male 2	-.38	.33	-1.15	67	.2557
2 Limited Familiarity	-.23	.19	-.11	4 Speaker 5 Spanish-English Male 1	.15	.18	.09	10 Speaker 11 Spanish-English Female 2	-.38	.26	-1.46	111	.1470
2 Limited Familiarity	-.21	.19	-.10	2 Speaker 3 Arabic-English Female 1	.18	.18	.10	3 Speaker 4 Dhivehi-English Female 2	-.39	.26	-1.49	111	.1381
1 No Familiarity	-.73	.30	-.27	10 Speaker 11 Spanish-English Female 2	-.34	.30	-.04	11 Speaker 12 Arabic-English Female 2	-.39	.42	-.92	37	.3617
4 Very Familiar	.14	.22	.01	5 Speaker 6 Arabic-English Male 1	.54	.23	.22	10 Speaker 11 Spanish-English Female 2	-.39	.32	-1.21	67	.2293
3 Some Familiarity	-.07	.15	-.11	3 Speaker 4 Dhivehi-English Female 2	.33	.15	.11	8 Speaker 9 Spanish-English Female 1	-.40	.21	-1.86	153	.0655
1 No Familiarity	-.73	.30	-.27	10 Speaker 11 Spanish-English Female 2	-.30	.30	-.03	12 Speaker 13 Dhivehi-English Male 2	-.42	.42	-1.00	37	.3231
4 Very Familiar	.24	.24	.06	4 Speaker 5 Spanish-English Male 1	.68	.22	.31	6 Speaker 7 Spanish-English Male 2	-.43	.33	-1.33	67	.1872
1 No Familiarity	-.43	.34	-.08	2 Speaker 3 Arabic-English Female 1	.02	.33	.13	4 Speaker 5 Spanish-English Male 1	-.46	.47	-.97	37	.3381
4 Very Familiar	.07	.23	-.03	3 Speaker 4 Dhivehi-English Female 2	.54	.23	.22	10 Speaker 11 Spanish-English Female 2	-.46	.33	-1.42	67	.1602
2 Limited Familiarity	-.38	.19	-.20	8 Speaker 9 Spanish-English Female 1	.09	.17	.05	12 Speaker 13 Dhivehi-English Male 2	-.47	.26	-1.83	111	.0702
4 Very Familiar	.06	.24	-.03	2 Speaker 3 Arabic-English Female 1	.54	.23	.22	10 Speaker 11 Spanish-English Female 2	-.48	.33	-1.43	67	.1573
4 Very Familiar	.06	.23	-.04	9 Speaker 10 Arabic-English Male 2	.54	.23	.22	10 Speaker 11 Spanish-English Female 2	-.48	.33	-1.46	67	.1485
2 Limited Familiarity	-.38	.19	-.20	8 Speaker 9 Spanish-English Female 1	.12	.18	.07	9 Speaker 10 Arabic-English Male 2	-.50	.26	-1.96	111	.0527
4 Very Familiar	-.54	.24	-.35	1 Speaker 2 Dhivehi-English Female 1	-.02	.22	-.08	11 Speaker 12 Arabic-English Female 2	-.52	.33	-1.60	67	.1153
4 Very Familiar	.14	.22	.01	5 Speaker 6 Arabic-English Male 1	.68	.22	.31	6 Speaker 7 Spanish-English Male 2	-.53	.32	-1.68	67	.0982
2 Limited Familiarity	-.38	.19	-.20	8 Speaker 9 Spanish-English Female 1	.15	.18	.09	10 Speaker 11 Spanish-English Female 2	-.54	.26	-2.08	111	.0400
1 No Familiarity	-.43	.34	-.08	2 Speaker 3 Arabic-English Female 1	.11	.31	.19	3 Speaker 4 Dhivehi-English Female 2	-.54	.46	-1.18	37	.2437
4 Very Familiar	-.54	.24	-.35	1 Speaker 2 Dhivehi-English Female 1	.06	.23	-.04	9 Speaker 10 Arabic-English Male 2	-.60	.33	-1.82	67	.0734
4 Very Familiar	-.54	.24	-.35	1 Speaker 2 Dhivehi-English Female 1	.06	.24	-.03	2 Speaker 3 Arabic-English Female 1	-.60	.34	-1.79	67	.0783
4 Very Familiar	.07	.23	-.03	3 Speaker 4 Dhivehi-English Female 2	.68	.22	.31	6 Speaker 7 Spanish-English Male 2	-.60	.32	-1.88	67	.0638
4 Very Familiar	-.54	.24	-.35	1 Speaker 2 Dhivehi-English Female 1	.07	.23	-.03	3 Speaker 4 Dhivehi-English Female 2	-.61	.33	-1.86	67	.0672
4 Very Familiar	.06	.24	-.03	2 Speaker 3 Arabic-English Female 1	.68	.22	.31	6 Speaker 7 Spanish-English Male 2	-.61	.33	-1.89	67	.0637
4 Very Familiar	-.54	.24	-.35	1 Speaker 2 Dhivehi-English Female 1	.14	.22	.01	5 Speaker 6 Arabic-English Male 1	-.68	.33	-2.08	67	.0409
4 Very Familiar	-.54	.24	-.35	1 Speaker 2 Dhivehi-English Female 1	.19	.22	.04	7 Speaker 8 Dhivehi-English Male 1	-.74	.33	-2.24	67	.0285
4 Very Familiar	-.54	.24	-.35	1 Speaker 2 Dhivehi-English Female 1	.24	.24	.06	4 Speaker 5 Spanish-English Male 1	-.78	.34	-2.33	67	.0228
4 Very Familiar	-.54	.24	-.35	1 Speaker 2 Dhivehi-English Female 1	.25	.23	.06	8 Speaker 9 Spanish-English Female 1	-.79	.33	-2.38	67	.0203
4 Very Familiar	-.54	.24	-.35	1 Speaker 2 Dhivehi-English Female 1	.54	.23	.22	10 Speaker 11 Spanish-English Female 2	-1.08	.33	-3.22	67	.0020
4 Very Familiar	-.54	.24	-.35	1 Speaker 2 Dhivehi-English Female 1	.68	.22	.31	6 Speaker 7 Spanish-English Male 2	-1.22	.33	-3.71	67	.0004

The Facets analyses output of bias/interaction - Arabic-English

```
Rasch Analysis of Expert Ratings 08/01/2015 15:25:07
Table 9.3  Bias Iteration Report.

Bias/Interaction: 2. Arabic English Familiarity Level, 3. Speaker

There are empirically 48 Bias terms
+-----------------------------------------------------------+
| Iteration      Max. Score Residual      Max. Logit Change |
|             Elements    %  Categories   Elements    Steps |
|-----------------------------------------------------------|
| BIAS   1    -10.3168  19.1                .6330           |
| BIAS   2      -.2457   -.7               -.0218           |
| BIAS   3      -.0002    .0                .0000           |
+-----------------------------------------------------------+

Rasch Analysis of Expert Ratings 08/01/2015 15:25:07
Table 11.3  Bias/Interaction Measurement Summary.

Bias/Interaction: 2. Arabic English Familiarity Level, 3. Speaker

+--------------------------------------------------+
| Cat  Score  Exp.  Resd StRes|                    |
|-----------------------------+--------------------|
| 2.83  2.83  2.83   .00  .00 | Mean (Count: 2280) |
|  .97   .97   .63   .73  .99 | S.D. (Population)  |
|  .97   .97   .63   .73  .99 | S.D. (Sample)      |
+--------------------------------------------------+
Count of measurable responses          = 2280
Raw-score variance of observations      =   0.94 100.00%
Variance explained by Rasch measures    =   0.39  42.00%
Variance of residuals                   =   0.55  58.00%
Variance explained by bias/interactions =   0.01   1.00%
Variance remaining in residuals         =   0.54  57.01%

Rasch Analysis of Expert Ratings 08/01/2015 15:25:07
Table 12.3  Bias/Interaction Summary Report.

Bias/Interaction: 2. Arabic English Familiarity Level, 3. Speaker (higher score = higher bias measure)

Bias/Interaction Size:
                 1            22 134124 6 222141322  1   1       1
 +-----------------------Q-------S-------M-------S-------Q-----------------------+
-1                                       0                                       1

Bias/Interaction Significance:
           1  1   112 21 452 1411142 21122 113      1
 +---------+-Q-------+S--------M--------S+-------Q-+---------+
-3        -2        -1         0         1         2         3


Rasch Analysis of Expert Ratings 08/01/2015 15:25:07
Table 13.3.1  Bias/Interaction Report (arranged by mN).

Bias/Interaction: 2. Arabic English Familiarity Level, 3. Speaker (higher score = higher bias measure)

+------------------------------------------------------------------------------------------------------------------------------------------------------------------+
|Observd  Expctd  Observd  Obs-Exp|  Bias  Model                    |Infit Outfit|    Arabic English Familiarity LeSpeaker                                         |
|  Score   Score    Count  Average|  Size   S.E.      t   d.f. Prob. | MnSq  MnSq | Sq N Arabic English Fami  measr Nu Speaker                                 measr |
|---------------------------------+---------------------------------+------------+---------------------------------------------------------------------------------|
|   55      48.32    20        .33|   .61   .30   2.05     19 .0544 |  1.3   1.3 |  8 4 Very Familiar          .11  2 Speaker 3 Arabic-English Female 1     -.98 |
|   63      58.49    20        .23|   .39   .29   1.33     19 .1991 |  1.7   1.7 | 28 4 Very Familiar          .11  7 Speaker 8 Dhivehi-English Male 1      -.06 |
|  109     102.58    40        .16|   .29   .21   1.37     39 .1774 |   .7    .7 |  9 1 No Familiarity        -.10  3 Speaker 4 Dhivehi-English Female 2    -.29 |
|  227     218.03    68        .13|   .23   .16   1.44     67 .1559 |   .5    .5 | 19 3 Some Familiarity       .08  5 Speaker 6 Arabic-English Male 1        .50 |
|  231     222.32    68        .13|   .22   .16   1.39     67 .1691 |  1.0   1.0 | 43 3 Some Familiarity       .08 11 Speaker 12 Arabic-English Female 2     .61 |
|  140     133.55    62        .10|   .21   .18   1.18     61 .2437 |   .7    .7 | 14 2 Limited Familiarity   -.09  4 Speaker 5 Spanish-English Male 1     -1.10 |
|  112     107.78    40        .11|   .19   .21    .89     39 .3771 |  2.0   2.0 | 25 1 No Familiarity        -.10  7 Speaker 8 Dhivehi-English Male 1      -.06 |
|  101      97.20    40        .10|   .18   .22    .83     39 .4137 |  1.9   1.9 |  1 1 No Familiarity        -.10  1 Speaker 2 Dhivehi-English Female 1    -.54 |
|  165     159.34    62        .09|   .17   .17    .98     61 .3330 |   .9    .9 | 10 2 Limited Familiarity   -.09  3 Speaker 4 Dhivehi-English Female 2    -.29 |
```

247	240.82	68	.09	.17	.16	1.01	67	.3180	.8	.8	35	3 Some Familiarity	.08	9 Speaker 10 Arabic-English Male 2	1.09
201	195.84	62	.08	.15	.17	.87	61	.3888	1.0	.9	46	2 Limited Familiarity	-.09	12 Speaker 13 Dhivehi-English Male 2	.75
137	133.90	40	.08	.14	.21	.65	39	.5190	1.0	1.0	37	1 No Familiarity	-.10	10 Speaker 11 Spanish-English Female 2	1.09
91	88.34	40	.07	.13	.22	.60	39	.5535	1.2	1.2	5	1 No Familiarity	-.10	2 Speaker 3 Arabic-English Female 1	-.98
73	71.61	20	.07	.13	.30	.42	19	.6808	.8	.8	40	4 Very Familiar	.11	10 Speaker 11 Spanish-English Female 2	1.09
129	126.25	40	.07	.12	.21	.57	39	.5697	.5	.5	45	1 No Familiarity	-.10	12 Speaker 13 Dhivehi-English Male 2	.75
48	47.01	20	.05	.09	.31	.31	19	.7617	.7	.7	16	4 Very Familiar	.11	4 Speaker 5 Spanish-English Male 1	-1.10
53	52.13	20	.04	.08	.30	.26	19	.7955	1.0	1.0	32	4 Very Familiar	.11	8 Speaker 9 Spanish-English Female 1	-.62
157	154.54	62	.04	.07	.17	.43	61	.6709	.8	.9	22	2 Limited Familiarity	-.09	6 Speaker 7 Spanish-English Male 2	-.44
153	151.03	62	.03	.06	.17	.34	61	.7318	1.2	1.2	2	2 Limited Familiarity	-.09	1 Speaker 2 Dhivehi-English Female 1	-.54
177	174.91	68	.03	.06	.16	.34	67	.7325	.7	.7	31	3 Some Familiarity	.08	8 Speaker 9 Spanish-English Female 1	-.62
242	240.82	68	.02	.03	.16	.19	67	.8480	.8	.8	39	3 Some Familiarity	.08	10 Speaker 11 Spanish-English Female 2	1.09
100	99.47	40	.01	.02	.22	.11	39	.9101	.8	.8	21	1 No Familiarity	-.10	6 Speaker 7 Spanish-English Male 2	-.44
66	66.15	20	-.01	-.01	.29	-.04	19	.9647	.8	.8	44	4 Very Familiar	.11	11 Speaker 12 Arabic-English Female 2	.61
148	148.48	62	-.01	-.01	.18	-.08	61	.9327	1.3	1.3	30	2 Limited Familiarity	-.09	8 Speaker 9 Spanish-English Female 1	-.62
54	54.18	20	-.01	-.02	.30	-.05	19	.9567	.8	.8	24	4 Very Familiar	.11	6 Speaker 7 Spanish-English Male 2	-.44
207	207.62	62	-.01	-.02	.17	-.10	61	.9173	.7	.7	34	2 Limited Familiarity	-.09	9 Speaker 10 Arabic-English Male 2	1.09
186	186.73	62	-.01	-.02	.17	-.12	61	.9030	.7	.7	18	2 Limited Familiarity	-.09	5 Speaker 6 Arabic-English Male 1	.50
227	228.04	68	-.02	-.03	.16	-.17	67	.8683	.7	.7	47	3 Some Familiarity	.08	12 Speaker 13 Dhivehi-English Male 2	.75
188	190.62	62	-.04	-.07	.17	-.44	61	.6606	1.0	1.0	42	2 Limited Familiarity	-.09	11 Speaker 12 Arabic-English Female 2	.61
179	181.85	68	-.04	-.08	.16	-.46	67	.6444	.8	.8	23	3 Some Familiarity	.08	6 Speaker 7 Spanish-English Male 2	-.44
155	157.65	68	-.04	-.08	.17	-.45	67	.6547	.6	.6	15	3 Some Familiarity	.08	4 Speaker 5 Spanish-English Male 1	-1.10
159	162.07	68	-.05	-.09	.17	-.52	67	.6082	1.4	1.4	7	3 Some Familiarity	.08	2 Speaker 3 Arabic-English Female 1	-.98
193	196.40	68	-.05	-.09	.16	-.55	67	.5871	1.1	1.1	27	3 Some Familiarity	.08	7 Speaker 8 Dhivehi-English Male 1	-.06
174	177.83	68	-.06	-.10	.16	-.63	67	.5329	1.0	.9	3	3 Some Familiarity	.08	1 Speaker 2 Dhivehi-English Female 1	-.54
93	95.54	40	-.06	-.12	.22	-.56	39	.5814	1.1	1.1	29	1 No Familiarity	-.10	8 Speaker 9 Spanish-English Female 1	-.62
70	71.61	20	-.08	-.14	.30	-.48	19	.6353	.5	.5	36	4 Very Familiar	.11	9 Speaker 10 Arabic-English Male 2	1.09
162	167.36	62	-.09	-.16	.17	-.91	61	.3644	1.6	1.6	26	2 Limited Familiarity	-.09	7 Speaker 8 Dhivehi-English Male 1	-.06
202	207.62	62	-.09	-.16	.17	-.95	61	.3465	1.0	1.0	38	2 Limited Familiarity	-.09	10 Speaker 11 Spanish-English Female 2	1.09
54	55.80	20	-.09	-.16	.30	-.54	19	.5985	.4	.4	12	4 Very Familiar	.11	3 Speaker 4 Dhivehi-English Female 2	-.29
63	64.88	20	-.09	-.16	.29	-.55	19	.5859	1.0	1.0	20	4 Very Familiar	.11	5 Speaker 6 Arabic-English Male 1	.50
130	133.90	40	-.10	-.17	.21	-.82	39	.4189	1.0	1.0	33	1 No Familiarity	-.10	9 Speaker 10 Arabic-English Male 2	1.09
51	52.99	20	-.10	-.18	.30	-.60	19	.5561	1.2	1.3	4	4 Very Familiar	.11	1 Speaker 2 Dhivehi-English Female 1	-.54
131	137.35	62	-.10	-.21	.18	-1.15	61	.2558	1.3	1.3	6	2 Limited Familiarity	-.09	2 Speaker 3 Arabic-English Female 1	-.98
117	122.87	40	-.15	-.26	.21	-1.22	39	.2281	1.2	1.2	41	1 No Familiarity	-.10	11 Speaker 12 Arabic-English Female 2	.61
81	85.88	40	-.12	-.26	.23	-1.11	39	.2756	1.1	1.0	13	1 No Familiarity	-.10	4 Speaker 5 Spanish-English Male 1	-1.10
177	187.32	68	-.15	-.27	.16	-1.67	67	.1001	.9	1.0	11	3 Some Familiarity	.08	3 Speaker 4 Dhivehi-English Female 2	-.29
114	120.34	40	-.16	-.28	.21	-1.32	39	.1932	1.5	1.5	17	1 No Familiarity	-.10	5 Speaker 6 Arabic-English Male 1	.50
61	67.84	20	-.34	-.59	.29	-2.02	19	.0581	.9	.9	48	4 Very Familiar	.11	12 Speaker 13 Dhivehi-English Male 2	.75
134.4	134.44	47.5	.00	.00	.21	.00			1.0	1.0	Mean (Count: 48)				
58.5	58.13	19.0	.11	.20	.05	.88			.4	.4	S.D. (Population)				
59.1	58.75	19.2	.11	.20	.05	.89			.4	.4	S.D. (Sample)				

Fixed (all = 0) chi-square: 37.5 d.f.: 48 significance (probability): .86

Rasch Analysis of Expert Ratings 08/01/2015 15:25:07
Table 14.3.1.2 Bias/Interaction Pairwise Report (arranged by mN).

Bias/Interaction: 2. Arabic English Familiarity Level, 3. Speaker

Target Nu Speaker	Target Measr	S.E.	Obs-Exp Average	Context N Arabic English Fami	Target Measr	S.E.	Obs-Exp Average	Context N Arabic English Fami	Target Contrast	Joint S.E.	t	Welch d.f.	Prob.
12 Speaker 13 Dhivehi-English Male 2	.90	.17	.08	2 Limited Familiarity	.16	.29	-.34	4 Very Familiar	.74	.34	2.18	42	.0348
12 Speaker 13 Dhivehi-English Male 2	.87	.21	.07	1 No Familiarity	.16	.29	-.34	4 Very Familiar	.71	.36	1.98	44	.0544
12 Speaker 13 Dhivehi-English Male 2	.73	.16	-.02	3 Some Familiarity	.16	.29	-.34	4 Very Familiar	.57	.34	1.69	41	.0986
3 Speaker 4 Dhivehi-English Female 2	.00	.21	.16	1 No Familiarity	-.57	.16	-.15	3 Some Familiarity	.56	.27	2.11	91	.0379
3 Speaker 4 Dhivehi-English Female 2	.00	.21	.16	1 No Familiarity	-.45	.30	-.09	4 Very Familiar	.45	.37	1.23	44	.2252
3 Speaker 4 Dhivehi-English Female 2	-.13	.17	.09	2 Limited Familiarity	-.57	.16	-.15	3 Some Familiarity	.44	.24	1.86	127	.0655
5 Speaker 6 Arabic-English Male 1	.73	.16	.13	3 Some Familiarity	.33	.29	-.09	4 Very Familiar	.39	.34	1.17	41	.2470
1 Speaker 2 Dhivehi-English Female 1	-.37	.22	.10	1 No Familiarity	-.73	.30	-.10	4 Very Familiar	.36	.37	.97	44	.3390
7 Speaker 8 Dhivehi-English Male 1	.13	.21	.11	1 No Familiarity	-.21	.17	-.09	2 Limited Familiarity	.35	.27	1.27	90	.2073
2 Speaker 3 Arabic-English Female 1	-.84	.22	.07	1 No Familiarity	-1.19	.18	-.10	2 Limited Familiarity	.34	.29	1.19	90	.2364
3 Speaker 4 Dhivehi-English Female 2	-.13	.17	.09	2 Limited Familiarity	-.45	.30	-.09	4 Very Familiar	.33	.34	.95	42	.3479
9 Speaker 10 Arabic-English Male 2	1.25	.16	.09	3 Some Familiarity	.94	.30	-.08	4 Very Familiar	.31	.34	.91	42	.3692
10 Speaker 11 Spanish-English Female 2	1.23	.21	.08	1 No Familiarity	.93	.17	-.09	2 Limited Familiarity	.30	.27	1.10	89	.2740
4 Speaker 5 Spanish-English Male 1	-.89	.18	.10	2 Limited Familiarity	-1.18	.17	-.04	3 Some Familiarity	.29	.25	1.16	127	.2470
1 Speaker 2 Dhivehi-English Female 1	-.37	.22	.10	1 No Familiarity	-.65	.16	-.06	3 Some Familiarity	.28	.27	1.04	90	.3025
7 Speaker 8 Dhivehi-English Male 1	.13	.21	.11	1 No Familiarity	-.15	.16	-.05	3 Some Familiarity	.28	.26	1.04	90	.3006

1 Speaker 2 Dhivehi-English Female 1	-.48	.17	.03	2 Limited Familiarity	-.73	.30	-.10	4 Very Familiar	.24	.35	.69	42	.4934
11 Speaker 12 Arabic-English Female 2	.83	.16	.13	3 Some Familiarity	.59	.29	-.01	4 Very Familiar	.24	.34	.71	41	.4842
2 Speaker 3 Arabic-English Female 1	-.84	.22	.07	1 No Familiarity	-1.06	.17	-.05	3 Some Familiarity	.22	.28	.79	90	.4331
12 Speaker 13 Dhivehi-English Male 2	.90	.17	.08	2 Limited Familiarity	.73	.16	-.02	3 Some Familiarity	.17	.23	.74	127	.4590
1 Speaker 2 Dhivehi-English Female 1	-.48	.17	.03	2 Limited Familiarity	-.65	.16	-.06	3 Some Familiarity	.16	.24	.68	127	.4975
6 Speaker 7 Spanish-English Male 2	-.36	.17	.04	2 Limited Familiarity	-.51	.16	-.04	3 Some Familiarity	.15	.24	.63	127	.5307
12 Speaker 13 Dhivehi-English Male 2	.87	.21	.07	1 No Familiarity	.73	.16	-.02	3 Some Familiarity	.15	.26	.56	90	.5796
5 Speaker 6 Arabic-English Male 1	.48	.17	-.01	2 Limited Familiarity	.33	.29	-.09	4 Very Familiar	.14	.34	.42	42	.6764
9 Speaker 10 Arabic-English Male 2	1.07	.17	-.01	2 Limited Familiarity	.94	.30	-.08	4 Very Familiar	.13	.34	.37	42	.7148
3 Speaker 4 Dhivehi-English Female 2	.00	.21	.16	1 No Familiarity	-.13	.17	.09	2 Limited Familiarity	.12	.27	.45	89	.6511
1 Speaker 2 Dhivehi-English Female 1	-.37	.22	.10	1 No Familiarity	-.48	.17	.03	2 Limited Familiarity	.12	.28	.43	90	.6718
4 Speaker 5 Spanish-English Male 1	-.89	.18	.10	2 Limited Familiarity	-1.01	.31	.05	4 Very Familiar	.12	.36	.33	42	.7462
10 Speaker 11 Spanish-English Female 2	1.23	.21	.08	1 No Familiarity	1.12	.16	.02	3 Some Familiarity	.11	.27	.40	91	.6925
6 Speaker 7 Spanish-English Male 2	-.41	.22	.01	1 No Familiarity	-.51	.16	-.04	3 Some Familiarity	.10	.27	.37	90	.7119
6 Speaker 7 Spanish-English Male 2	-.36	.17	.04	2 Limited Familiarity	-.45	.30	-.01	4 Very Familiar	.09	.35	.26	42	.7951
1 Speaker 2 Dhivehi-English Female 1	-.65	.16	-.06	3 Some Familiarity	-.73	.30	-.10	4 Very Familiar	.08	.34	.23	41	.8207
6 Speaker 7 Spanish-English Male 2	-.41	.22	.01	1 No Familiarity	-.45	.30	-.01	4 Very Familiar	.04	.37	.11	44	.9121
10 Speaker 11 Spanish-English Female 2	1.23	.21	.08	1 No Familiarity	1.21	.30	.07	4 Very Familiar	.01	.37	.03	44	.9767
8 Speaker 9 Spanish-English Female 1	-.57	.16	.03	3 Some Familiarity	-.54	.30	.04	4 Very Familiar	-.02	.34	-.07	41	.9472
12 Speaker 13 Dhivehi-English Male 2	.87	.21	.07	1 No Familiarity	.90	.17	.08	2 Limited Familiarity	-.03	.27	-.10	89	.9221
9 Speaker 10 Arabic-English Male 2	.92	.21	-.10	1 No Familiarity	.94	.30	-.08	4 Very Familiar	-.03	.36	-.07	44	.9409
6 Speaker 7 Spanish-English Male 2	-.41	.22	.01	1 No Familiarity	-.36	.17	.04	2 Limited Familiarity	-.05	.28	-.18	89	.8586
6 Speaker 7 Spanish-English Male 2	-.51	.16	-.04	3 Some Familiarity	-.45	.30	-.01	4 Very Familiar	-.06	.34	-.17	41	.8629
11 Speaker 12 Arabic-English Female 2	.53	.17	-.04	2 Limited Familiarity	.59	.29	-.01	4 Very Familiar	-.06	.34	-.18	42	.8583
7 Speaker 8 Dhivehi-English Male 1	-.21	.17	-.09	2 Limited Familiarity	-.15	.16	-.05	3 Some Familiarity	-.07	.24	-.29	127	.7694
8 Speaker 9 Spanish-English Female 1	-.64	.18	-.01	2 Limited Familiarity	-.57	.16	.03	3 Some Familiarity	-.07	.24	-.30	127	.7679
8 Speaker 9 Spanish-English Female 1	-.64	.18	-.01	2 Limited Familiarity	-.54	.30	.04	4 Very Familiar	-.09	.35	-.27	43	.7887
10 Speaker 11 Spanish-English Female 2	1.12	.16	.02	3 Some Familiarity	1.21	.30	.07	4 Very Familiar	-.09	.34	-.28	41	.7838
8 Speaker 9 Spanish-English Female 1	-.75	.22	-.06	1 No Familiarity	-.64	.18	-.01	2 Limited Familiarity	-.11	.28	-.38	89	.7037
3 Speaker 4 Dhivehi-English Female 2	-.57	.16	-.15	3 Some Familiarity	-.45	.30	-.09	4 Very Familiar	-.11	.34	-.33	41	.7428
5 Speaker 6 Arabic-English Male 1	.22	.21	-.16	1 No Familiarity	.33	.29	-.09	4 Very Familiar	-.11	.36	-.32	44	.7520
2 Speaker 3 Arabic-English Female 1	-1.19	.18	-.10	2 Limited Familiarity	-1.06	.17	-.05	3 Some Familiarity	-.12	.25	-.50	126	.6200
9 Speaker 10 Arabic-English Male 2	.92	.21	-.10	1 No Familiarity	1.07	.17	-.01	2 Limited Familiarity	-.15	.27	-.57	89	.5702
4 Speaker 5 Spanish-English Male 1	-1.18	.17	-.04	3 Some Familiarity	-1.01	.31	.05	4 Very Familiar	-.17	.35	-.49	42	.6292
8 Speaker 9 Spanish-English Female 1	-.75	.22	-.06	1 No Familiarity	-.57	.16	.03	3 Some Familiarity	-.18	.27	-.65	89	.5167
4 Speaker 5 Spanish-English Male 1	-1.36	.23	-.12	1 No Familiarity	-1.18	.17	-.04	3 Some Familiarity	-.18	.29	-.63	89	.5323
11 Speaker 12 Arabic-English Female 2	.35	.21	-.15	1 No Familiarity	.53	.17	-.04	2 Limited Familiarity	-.18	.27	-.68	89	.4993
9 Speaker 10 Arabic-English Male 2	1.07	.17	-.01	2 Limited Familiarity	1.25	.16	.09	3 Some Familiarity	-.18	.24	-.78	127	.4391
10 Speaker 11 Spanish-English Female 2	.93	.17	-.09	2 Limited Familiarity	1.12	.16	.02	3 Some Familiarity	-.19	.23	-.82	127	.4164
8 Speaker 9 Spanish-English Female 1	-.75	.22	-.06	1 No Familiarity	-.54	.30	.04	4 Very Familiar	-.20	.37	-.54	45	.5913
7 Speaker 8 Dhivehi-English Male 1	.13	.21	.11	1 No Familiarity	.33	.29	.23	4 Very Familiar	-.20	.36	-.56	44	.5770
11 Speaker 12 Arabic-English Female 2	.35	.21	-.15	1 No Familiarity	.59	.29	-.01	4 Very Familiar	-.24	.36	-.67	44	.5049
5 Speaker 6 Arabic-English Male 1	.48	.17	-.01	2 Limited Familiarity	.73	.16	.13	3 Some Familiarity	-.25	.23	-1.08	127	.2823
5 Speaker 6 Arabic-English Male 1	.22	.21	-.16	1 No Familiarity	.48	.17	-.01	2 Limited Familiarity	-.26	.27	-.96	89	.3414
10 Speaker 11 Spanish-English Female 2	.93	.17	-.09	2 Limited Familiarity	1.21	.30	.07	4 Very Familiar	-.29	.35	-.83	42	.4130
11 Speaker 12 Arabic-English Female 2	.53	.17	-.04	2 Limited Familiarity	.83	.16	.13	3 Some Familiarity	-.30	.23	-1.28	127	.2028
9 Speaker 10 Arabic-English Male 2	.92	.21	-.10	1 No Familiarity	1.25	.16	.09	3 Some Familiarity	-.34	.27	-1.26	91	.2095
4 Speaker 5 Spanish-English Male 1	-1.36	.23	-.12	1 No Familiarity	-1.01	.31	.05	4 Very Familiar	-.35	.39	-.91	45	.3669
4 Speaker 5 Spanish-English Male 1	-1.36	.23	-.12	1 No Familiarity	-.89	.18	.10	2 Limited Familiarity	-.47	.29	-1.59	88	.1144
2 Speaker 3 Arabic-English Female 1	-.84	.22	.07	1 No Familiarity	-.36	.30	.33	4 Very Familiar	-.48	.37	-1.29	45	.2046
7 Speaker 8 Dhivehi-English Male 1	-.15	.16	-.05	3 Some Familiarity	.33	.29	.23	4 Very Familiar	-.48	.34	-1.43	41	.1606
11 Speaker 12 Arabic-English Female 2	.35	.21	-.15	1 No Familiarity	.83	.16	.13	3 Some Familiarity	-.48	.26	-1.82	90	.0723
5 Speaker 6 Arabic-English Male 1	.22	.21	-.16	1 No Familiarity	.73	.16	.13	3 Some Familiarity	-.51	.26	-1.92	90	.0576
7 Speaker 8 Dhivehi-English Male 1	-.21	.17	-.09	2 Limited Familiarity	.33	.29	.23	4 Very Familiar	-.55	.34	-1.61	43	.1148
2 Speaker 3 Arabic-English Female 1	-1.06	.17	-.05	3 Some Familiarity	-.36	.30	.33	4 Very Familiar	-.70	.34	-2.04	42	.0479
2 Speaker 3 Arabic-English Female 1	-1.19	.18	-.10	2 Limited Familiarity	-.36	.30	.33	4 Very Familiar	-.82	.35	-2.35	44	.0235

Rasch Analysis of Expert Ratings 08/01/2015 15:25:07
Table 14.3.1.3 Bias/Interaction Pairwise Report (arranged by mN).

Bias/Interaction: 2. Arabic English Familiarity Level, 3. Speaker

Target N Arabic English Fami	Target Measr	S.E.	Obs-Exp Average	Context Nu Speaker	Target Measr	S.E.	Obs-Exp Average	Context Nu Speaker	Target Contrast	Joint S.E.	t	Welch d.f.	Prob.
4 Very Familiar	.72	.30	.33	2 Speaker 3 Arabic-English Female 1	-.48	.29	-.34	12 Speaker 13 Dhivehi-English Male 2	1.20	.42	2.88	37	.0067
4 Very Familiar	.50	.29	.23	7 Speaker 8 Dhivehi-English Male 1	-.48	.29	-.34	12 Speaker 13 Dhivehi-English Male 2	.98	.42	2.37	37	.0233
4 Very Familiar	.72	.30	.33	2 Speaker 3 Arabic-English Female 1	-.05	.29	-.09	5 Speaker 6 Arabic-English Male 1	.77	.42	1.85	37	.0725
4 Very Familiar	.72	.30	.33	2 Speaker 3 Arabic-English Female 1	-.05	.30	-.09	3 Speaker 4 Dhivehi-English Female 2	.77	.42	1.83	37	.0759
4 Very Familiar	.72	.30	.33	2 Speaker 3 Arabic-English Female 1	-.03	.30	-.08	9 Speaker 10 Arabic-English Male 2	.75	.42	1.79	37	.0815

4 Very Familiar	.24	.30	.07	10 Speaker 11 Spanish-English Female 2	-.48	.29	-.34	12 Speaker 13 Dhivehi-English Male 2	.72	.42	1.71	37	.0963
4 Very Familiar	.21	.31	.05	4 Speaker 5 Spanish-English Male 1	-.48	.29	-.34	12 Speaker 13 Dhivehi-English Male 2	.69	.43	1.61	37	.1149
4 Very Familiar	.19	.30	.04	8 Speaker 9 Spanish-English Female 1	-.48	.29	-.34	12 Speaker 13 Dhivehi-English Male 2	.67	.42	1.60	37	.1184
4 Very Familiar	.72	.30	.33	2 Speaker 3 Arabic-English Female 1	.09	.30	-.01	6 Speaker 7 Spanish-English Male 2	.63	.42	1.49	37	.1457
4 Very Familiar	.72	.30	.33	2 Speaker 3 Arabic-English Female 1	.10	.29	-.01	11 Speaker 12 Arabic-English Female 2	.62	.42	1.49	37	.1449
4 Very Familiar	.10	.29	-.01	11 Speaker 12 Arabic-English Female 2	-.48	.29	-.34	12 Speaker 13 Dhivehi-English Male 2	.58	.42	1.39	37	.1720
4 Very Familiar	.09	.30	-.01	6 Speaker 7 Spanish-English Male 2	-.48	.29	-.34	12 Speaker 13 Dhivehi-English Male 2	.58	.42	1.37	37	.1775
1 No Familiarity	.19	.21	.16	3 Speaker 4 Dhivehi-English Female 2	-.37	.21	-.16	5 Speaker 6 Arabic-English Male 1	.57	.30	1.91	77	.0602
1 No Familiarity	.19	.21	.16	3 Speaker 4 Dhivehi-English Female 2	-.35	.23	-.12	4 Speaker 5 Spanish-English Male 1	.55	.31	1.74	77	.0855
1 No Familiarity	.19	.21	.16	3 Speaker 4 Dhivehi-English Female 2	-.35	.21	-.15	11 Speaker 12 Arabic-English Female 2	.55	.30	1.84	77	.0700
4 Very Familiar	.50	.29	.23	7 Speaker 8 Dhivehi-English Male 1	-.03	.30	-.08	9 Speaker 10 Arabic-English Male 2	.53	.42	1.28	37	.2093
4 Very Familiar	.72	.30	.33	2 Speaker 3 Arabic-English Female 1	.19	.30	.04	8 Speaker 9 Spanish-English Female 1	.53	.42	1.26	37	.2164
4 Very Familiar	.72	.30	.33	2 Speaker 3 Arabic-English Female 1	.21	.31	.05	4 Speaker 5 Spanish-English Male 1	.52	.43	1.20	37	.2364
4 Very Familiar	.72	.30	.33	2 Speaker 3 Arabic-English Female 1	.24	.30	.07	10 Speaker 11 Spanish-English Female 2	.49	.42	1.14	37	.2605
1 No Familiarity	.19	.21	.16	3 Speaker 4 Dhivehi-English Female 2	-.27	.21	-.10	9 Speaker 10 Arabic-English Male 2	.46	.30	1.55	77	.1250
1 No Familiarity	.08	.22	.10	1 Speaker 2 Dhivehi-English Female 1	-.37	.21	-.16	5 Speaker 6 Arabic-English Male 1	.46	.30	1.52	77	.1336
4 Very Familiar	-.03	.30	-.08	9 Speaker 10 Arabic-English Male 2	-.48	.29	-.34	12 Speaker 13 Dhivehi-English Male 2	.45	.42	1.07	37	.2898
1 No Familiarity	.09	.21	.11	7 Speaker 8 Dhivehi-English Male 1	-.35	.21	-.15	11 Speaker 12 Arabic-English Female 2	.44	.30	1.50	77	.1384
1 No Familiarity	.08	.22	.10	1 Speaker 2 Dhivehi-English Female 1	-.35	.23	-.12	4 Speaker 5 Spanish-English Male 1	.43	.32	1.37	77	.1738
1 No Familiarity	.08	.22	.10	1 Speaker 2 Dhivehi-English Female 1	-.35	.21	-.15	11 Speaker 12 Arabic-English Female 2	.43	.30	1.45	77	.1522
4 Very Familiar	-.05	.30	-.09	3 Speaker 4 Dhivehi-English Female 2	-.48	.29	-.34	12 Speaker 13 Dhivehi-English Male 2	.43	.42	1.03	37	.3087
4 Very Familiar	-.05	.29	-.09	5 Speaker 6 Arabic-English Male 1	-.48	.29	-.34	12 Speaker 13 Dhivehi-English Male 2	.43	.42	1.03	37	.3076
1 No Familiarity	.19	.21	.16	3 Speaker 4 Dhivehi-English Female 2	-.22	.22	-.06	8 Speaker 9 Spanish-English Female 1	.41	.31	1.35	77	.1803
4 Very Familiar	-.07	.30	-.10	1 Speaker 2 Dhivehi-English Female 1	-.48	.29	-.34	12 Speaker 13 Dhivehi-English Male 2	.41	.42	.97	37	.3362
1 No Familiarity	.04	.22	.07	2 Speaker 3 Arabic-English Female 1	-.37	.21	-.16	5 Speaker 6 Arabic-English Male 1	.41	.31	1.34	77	.1830
4 Very Familiar	.50	.29	.23	7 Speaker 8 Dhivehi-English Male 1	.10	.29	-.01	11 Speaker 12 Arabic-English Female 2	.40	.42	.97	37	.3377
1 No Familiarity	.04	.21	.08	10 Speaker 11 Spanish-English Female 2	-.35	.21	-.15	11 Speaker 12 Arabic-English Female 2	.39	.30	1.32	77	.1891
1 No Familiarity	.04	.22	.07	2 Speaker 3 Arabic-English Female 1	-.35	.23	-.12	4 Speaker 5 Spanish-English Male 1	.39	.32	1.21	77	.2290
1 No Familiarity	.04	.22	.07	2 Speaker 3 Arabic-English Female 1	-.35	.21	-.15	11 Speaker 12 Arabic-English Female 2	.39	.31	1.27	77	.2062
2 Limited Familiarity	.12	.18	.10	4 Speaker 5 Spanish-English Male 1	-.25	.17	-.09	10 Speaker 11 Spanish-English Female 2	.37	.25	1.51	121	.1343
2 Limited Familiarity	.12	.18	.10	4 Speaker 5 Spanish-English Male 1	-.25	.17	-.09	7 Speaker 8 Dhivehi-English Male 1	.37	.25	1.48	121	.1409
1 No Familiarity	.09	.21	.11	7 Speaker 8 Dhivehi-English Male 1	-.27	.21	-.10	9 Speaker 10 Arabic-English Male 2	.36	.30	1.21	77	.2301
1 No Familiarity	.08	.22	.10	1 Speaker 2 Dhivehi-English Female 1	-.27	.21	-.10	9 Speaker 10 Arabic-English Male 2	.35	.30	1.16	77	.2489
2 Limited Familiarity	.08	.17	.09	3 Speaker 4 Dhivehi-English Female 2	-.25	.17	-.09	10 Speaker 11 Spanish-English Female 2	.33	.24	1.36	121	.1761
2 Limited Familiarity	.08	.17	.09	3 Speaker 4 Dhivehi-English Female 2	-.25	.17	-.09	7 Speaker 8 Dhivehi-English Male 1	.32	.24	1.34	121	.1840
3 Some Familiarity	.31	.16	.13	5 Speaker 6 Arabic-English Male 1	-.01	.16	-.05	7 Speaker 8 Dhivehi-English Male 1	.32	.23	1.40	133	.1638
4 Very Familiar	.50	.29	.23	7 Speaker 8 Dhivehi-English Male 1	.19	.30	.04	8 Speaker 9 Spanish-English Female 1	.31	.42	.74	37	.4623
1 No Familiarity	.09	.21	.11	7 Speaker 8 Dhivehi-English Male 1	-.22	.22	-.06	8 Speaker 9 Spanish-English Female 1	.31	.30	1.02	77	.3113
3 Some Familiarity	.31	.16	.13	5 Speaker 6 Arabic-English Male 1	.00	.16	-.04	6 Speaker 7 Spanish-English Male 2	.31	.23	1.34	133	.1834
1 No Familiarity	.04	.22	.07	2 Speaker 3 Arabic-English Female 1	-.27	.21	-.10	9 Speaker 10 Arabic-English Male 2	.30	.30	1.00	77	.3228
1 No Familiarity	.08	.22	.10	1 Speaker 2 Dhivehi-English Female 1	-.22	.22	-.06	8 Speaker 9 Spanish-English Female 1	.30	.31	.97	77	.3326
2 Limited Familiarity	.12	.18	.10	4 Speaker 5 Spanish-English Male 1	-.17	.17	-.04	11 Speaker 12 Arabic-English Female 2	.29	.25	1.16	121	.2481
1 No Familiarity	-.07	.22	.01	6 Speaker 7 Spanish-English Male 2	-.35	.21	-.15	11 Speaker 12 Arabic-English Female 2	.28	.30	.93	77	.3529
2 Limited Familiarity	-.03	.17	.03	1 Speaker 2 Dhivehi-English Female 1	-.30	.18	-.10	2 Speaker 3 Arabic-English Female 1	.27	.25	1.07	121	.2870
1 No Familiarity	.19	.21	.16	3 Speaker 4 Dhivehi-English Female 2	-.07	.22	.01	6 Speaker 7 Spanish-English Male 2	.27	.30	.88	77	.3814
4 Very Familiar	.50	.29	.23	7 Speaker 8 Dhivehi-English Male 1	.24	.30	.07	10 Speaker 11 Spanish-English Female 2	.27	.42	.63	37	.5334
4 Very Familiar	.21	.31	.05	4 Speaker 5 Spanish-English Male 1	-.05	.29	-.09	5 Speaker 6 Arabic-English Male 1	.26	.43	.61	37	.5488
3 Some Familiarity	.31	.16	.13	5 Speaker 6 Arabic-English Male 1	.05	.16	-.02	12 Speaker 13 Dhivehi-English Male 2	.26	.23	1.13	134	.2594
1 No Familiarity	.04	.22	.07	2 Speaker 3 Arabic-English Female 1	-.22	.22	-.06	8 Speaker 9 Spanish-English Female 1	.26	.31	.82	77	.4171
3 Some Familiarity	.30	.16	.13	11 Speaker 12 Arabic-English Female 2	.05	.16	-.02	12 Speaker 13 Dhivehi-English Male 2	.25	.23	1.10	133	.2725
2 Limited Familiarity	.08	.17	.09	3 Speaker 4 Dhivehi-English Female 2	-.17	.17	-.04	11 Speaker 12 Arabic-English Female 2	.24	.24	1.01	121	.3168
4 Very Familiar	.21	.31	.05	4 Speaker 5 Spanish-English Male 1	-.03	.30	-.08	9 Speaker 10 Arabic-English Male 2	.24	.43	.56	37	.5815
2 Limited Familiarity	-.02	.17	.04	6 Speaker 7 Spanish-English Male 2	-.25	.17	-.09	10 Speaker 11 Spanish-English Female 2	.23	.24	.97	121	.3352
2 Limited Familiarity	.12	.18	.10	4 Speaker 5 Spanish-English Male 1	-.11	.17	-.01	5 Speaker 6 Arabic-English Male 1	.23	.25	.94	121	.3479
2 Limited Familiarity	-.02	.17	.04	6 Speaker 7 Spanish-English Male 2	-.25	.17	-.09	7 Speaker 8 Dhivehi-English Male 1	.23	.24	.95	121	.3456
2 Limited Familiarity	.12	.18	.10	4 Speaker 5 Spanish-English Male 1	-.11	.17	-.01	9 Speaker 10 Arabic-English Male 2	.23	.25	.93	121	.3549
2 Limited Familiarity	.12	.18	.10	4 Speaker 5 Spanish-English Male 1	-.11	.18	-.01	8 Speaker 9 Spanish-English Female 1	.23	.25	.90	121	.3702
4 Very Familiar	.19	.30	.04	8 Speaker 9 Spanish-English Female 1	-.03	.30	-.08	9 Speaker 10 Arabic-English Male 2	.22	.42	.53	37	.6020
4 Very Familiar	.72	.30	.33	2 Speaker 3 Arabic-English Female 1	.50	.29	.23	7 Speaker 8 Dhivehi-English Male 1	.22	.42	.53	37	.6026
2 Limited Familiarity	-.03	.17	.03	1 Speaker 2 Dhivehi-English Female 1	-.25	.17	-.09	10 Speaker 11 Spanish-English Female 2	.22	.24	.91	121	.3664
2 Limited Familiarity	-.03	.17	.03	1 Speaker 2 Dhivehi-English Female 1	-.25	.17	-.09	7 Speaker 8 Dhivehi-English Male 1	.22	.24	.89	121	.3770
3 Some Familiarity	.31	.16	.13	5 Speaker 6 Arabic-English Male 1	.11	.16	.02	10 Speaker 11 Spanish-English Female 2	.20	.23	.87	133	.3857
1 No Familiarity	-.07	.22	.01	6 Speaker 7 Spanish-English Male 2	-.27	.21	-.10	9 Speaker 10 Arabic-English Male 2	.20	.30	.65	77	.5175
3 Some Familiarity	.24	.16	.09	9 Speaker 10 Arabic-English Male 2	.05	.16	-.02	12 Speaker 13 Dhivehi-English Male 2	.19	.23	.84	133	.4049
2 Limited Familiarity	.08	.17	.09	3 Speaker 4 Dhivehi-English Female 2	-.11	.17	-.01	5 Speaker 6 Arabic-English Male 1	.19	.24	.78	121	.4360
3 Some Familiarity	-.01	.17	-.05	2 Speaker 3 Arabic-English Female 1	-.20	.16	-.15	3 Speaker 4 Dhivehi-English Female 2	.19	.24	.79	133	.4304
2 Limited Familiarity	.08	.17	.09	3 Speaker 4 Dhivehi-English Female 2	-.11	.17	-.01	9 Speaker 10 Arabic-English Male 2	.18	.24	.77	121	.4441
2 Limited Familiarity	.08	.17	.09	3 Speaker 4 Dhivehi-English Female 2	-.11	.18	-.01	8 Speaker 9 Spanish-English Female 1	.18	.25	.74	121	.4602
3 Some Familiarity	.31	.16	.13	5 Speaker 6 Arabic-English Male 1	.13	.16	.03	8 Speaker 9 Spanish-English Female 1	.17	.23	.76	133	.4483
1 No Familiarity	.19	.21	.16	3 Speaker 4 Dhivehi-English Female 2	.02	.21	.07	12 Speaker 13 Dhivehi-English Male 2	.17	.30	.57	77	.5677
3 Some Familiarity	-.03	.16	-.06	1 Speaker 2 Dhivehi-English Female 1	-.20	.16	-.15	3 Speaker 4 Dhivehi-English Female 2	.17	.23	.73	133	.4652
1 No Familiarity	.19	.21	.16	3 Speaker 4 Dhivehi-English Female 2	.04	.21	.08	10 Speaker 11 Spanish-English Female 2	.15	.30	.51	77	.6088

1 No Familiarity	.08	.22	.10	1 Speaker 2 Dhivehi-English Female 1	-.07	.22	.01	6 Speaker 7 Spanish-English Male 2	.15	.30	.50	77	.6164
2 Limited Familiarity	-.02	.17	.04	6 Speaker 7 Spanish-English Male 2	-.17	.17	-.04	11 Speaker 12 Arabic-English Female 2	.15	.24	.61	121	.5406
1 No Familiarity	-.07	.22	.01	6 Speaker 7 Spanish-English Male 2	-.22	.22	-.06	8 Speaker 9 Spanish-English Female 1	.15	.31	.48	77	.6347
2 Limited Familiarity	-.11	.18	-.01	8 Speaker 9 Spanish-English Female 1	-.25	.17	-.09	10 Speaker 11 Spanish-English Female 2	.14	.24	.59	121	.5533
2 Limited Familiarity	-.11	.17	-.01	9 Speaker 10 Arabic-English Male 2	-.25	.17	-.09	10 Speaker 11 Spanish-English Female 2	.14	.24	.60	121	.5522
4 Very Familiar	.24	.30	.07	10 Speaker 11 Spanish-English Female 2	.10	.29	-.01	11 Speaker 12 Arabic-English Female 2	.14	.42	.33	37	.7431
2 Limited Familiarity	-.11	.17	-.01	5 Speaker 6 Arabic-English Male 1	-.25	.17	-.09	10 Speaker 11 Spanish-English Female 2	.14	.24	.58	121	.5599
2 Limited Familiarity	.12	.18	.10	4 Speaker 5 Spanish-English Male 1	-.02	.17	.04	6 Speaker 7 Spanish-English Male 2	.14	.25	.55	121	.5829
2 Limited Familiarity	-.11	.17	-.01	5 Speaker 6 Arabic-English Male 1	-.25	.17	-.09	7 Speaker 8 Dhivehi-English Male 1	.14	.24	.57	121	.5716
2 Limited Familiarity	-.03	.17	.03	1 Speaker 2 Dhivehi-English Female 1	-.17	.17	-.04	11 Speaker 12 Arabic-English Female 2	.13	.24	.55	121	.5806
3 Some Familiarity	.24	.16	.09	9 Speaker 10 Arabic-English Male 2	.11	.16	.02	10 Speaker 11 Spanish-English Female 2	.13	.23	.58	133	.5640
1 No Familiarity	-.22	.22	-.06	8 Speaker 9 Spanish-English Female 1	-.35	.21	-.15	11 Speaker 12 Arabic-English Female 2	.13	.30	.44	77	.6611
4 Very Familiar	.09	.30	-.01	6 Speaker 7 Spanish-English Male 2	-.03	.30	-.08	9 Speaker 10 Arabic-English Male 2	.13	.42	.30	37	.7650
4 Very Familiar	.21	.31	.05	4 Speaker 5 Spanish-English Male 1	.09	.30	-.01	6 Speaker 7 Spanish-English Male 2	.11	.43	.26	37	.7970
1 No Familiarity	.04	.22	.07	2 Speaker 3 Arabic-English Female 1	-.07	.22	.01	6 Speaker 7 Spanish-English Male 2	.11	.31	.35	77	.7276
4 Very Familiar	.21	.31	.05	4 Speaker 5 Spanish-English Male 1	.10	.29	-.01	11 Speaker 12 Arabic-English Female 2	.11	.43	.25	37	.8014
1 No Familiarity	.19	.21	.16	3 Speaker 4 Dhivehi-English Female 2	.09	.21	.11	7 Speaker 8 Dhivehi-English Male 1	.10	.30	.34	77	.7325
2 Limited Familiarity	.08	.17	.09	3 Speaker 4 Dhivehi-English Female 2	-.02	.17	.04	6 Speaker 7 Spanish-English Male 2	.09	.24	.38	121	.7029
4 Very Familiar	.19	.30	.04	8 Speaker 9 Spanish-English Female 1	.10	.29	-.01	11 Speaker 12 Arabic-English Female 2	.09	.42	.22	37	.8278
2 Limited Familiarity	-.02	.17	.04	6 Speaker 7 Spanish-English Male 2	-.11	.17	-.01	9 Speaker 10 Arabic-English Male 2	.09	.24	.38	121	.7058
2 Limited Familiarity	-.02	.17	.04	6 Speaker 7 Spanish-English Male 2	-.11	.18	-.01	8 Speaker 9 Spanish-English Female 1	.09	.25	.36	121	.7196
1 No Familiarity	-.27	.21	-.10	9 Speaker 10 Arabic-English Male 2	-.35	.21	-.15	11 Speaker 12 Arabic-English Female 2	.09	.30	.29	77	.7731
3 Some Familiarity	.13	.16	.03	8 Speaker 9 Spanish-English Female 1	.05	.16	-.02	12 Speaker 13 Dhivehi-English Male 2	.08	.23	.36	133	.7182
2 Limited Familiarity	-.03	.17	.03	1 Speaker 2 Dhivehi-English Female 1	-.11	.17	-.01	5 Speaker 6 Arabic-English Male 1	.08	.24	.33	121	.7400
2 Limited Familiarity	-.03	.17	.03	1 Speaker 2 Dhivehi-English Female 1	-.11	.17	-.01	9 Speaker 10 Arabic-English Male 2	.08	.24	.32	121	.7497
2 Limited Familiarity	-.03	.17	.03	1 Speaker 2 Dhivehi-English Female 1	-.11	.18	-.01	8 Speaker 9 Spanish-English Female 1	.07	.25	.30	121	.7628
1 No Familiarity	.09	.21	.11	7 Speaker 8 Dhivehi-English Male 1	.02	.21	.07	12 Speaker 13 Dhivehi-English Male 2	.07	.30	.23	77	.8186
3 Some Familiarity	.31	.16	.13	5 Speaker 6 Arabic-English Male 1	.24	.16	.09	9 Speaker 10 Arabic-English Male 2	.07	.23	.28	133	.7764
2 Limited Familiarity	.12	.18	.10	4 Speaker 5 Spanish-English Male 1	.05	.17	.08	12 Speaker 13 Dhivehi-English Male 2	.07	.25	.26	121	.7919
2 Limited Familiarity	-.11	.18	-.01	8 Speaker 9 Spanish-English Female 1	-.17	.17	-.04	11 Speaker 12 Arabic-English Female 2	.06	.24	.24	121	.8081
3 Some Familiarity	.11	.16	.02	10 Speaker 11 Spanish-English Female 2	.05	.16	-.02	12 Speaker 13 Dhivehi-English Male 2	.06	.23	.25	133	.7999
1 No Familiarity	.08	.22	.10	1 Speaker 2 Dhivehi-English Female 1	.02	.21	.07	12 Speaker 13 Dhivehi-English Male 2	.06	.30	.19	77	.8474
2 Limited Familiarity	-.11	.17	-.01	9 Speaker 10 Arabic-English Male 2	-.17	.17	-.04	11 Speaker 12 Arabic-English Female 2	.06	.24	.24	121	.8127
2 Limited Familiarity	-.11	.17	-.01	5 Speaker 6 Arabic-English Male 1	-.17	.17	-.04	11 Speaker 12 Arabic-English Female 2	.05	.24	.23	121	.8221
1 No Familiarity	.09	.21	.11	7 Speaker 8 Dhivehi-English Male 1	.04	.21	.08	10 Speaker 11 Spanish-English Female 2	.05	.30	.17	77	.8644
1 No Familiarity	-.22	.22	-.06	8 Speaker 9 Spanish-English Female 1	-.27	.21	-.10	9 Speaker 10 Arabic-English Male 2	.05	.30	.16	77	.8744
1 No Familiarity	.08	.22	.10	1 Speaker 2 Dhivehi-English Female 1	.04	.22	.07	2 Speaker 3 Arabic-English Female 1	.04	.31	.15	77	.8847
1 No Familiarity	.08	.22	.10	1 Speaker 2 Dhivehi-English Female 1	.04	.21	.08	10 Speaker 11 Spanish-English Female 2	.04	.30	.14	77	.8929
3 Some Familiarity	.13	.16	.03	8 Speaker 9 Spanish-English Female 1	.11	.16	.02	10 Speaker 11 Spanish-English Female 2	.02	.23	.11	133	.9147
2 Limited Familiarity	.08	.17	.09	3 Speaker 4 Dhivehi-English Female 2	.05	.17	.08	12 Speaker 13 Dhivehi-English Male 2	.02	.24	.09	121	.9309
1 No Familiarity	-.35	.23	-.12	4 Speaker 5 Spanish-English Male 1	-.37	.21	-.16	5 Speaker 6 Arabic-English Male 1	.02	.31	.07	77	.9477
1 No Familiarity	.04	.21	.08	10 Speaker 11 Spanish-English Female 2	.02	.21	.07	12 Speaker 13 Dhivehi-English Male 2	.02	.30	.06	77	.9538
4 Very Familiar	.21	.31	.05	4 Speaker 5 Spanish-English Male 1	.19	.30	.04	8 Speaker 9 Spanish-English Female 1	.02	.43	.04	37	.9707
1 No Familiarity	.04	.22	.07	2 Speaker 3 Arabic-English Female 1	.02	.21	.07	12 Speaker 13 Dhivehi-English Male 2	.01	.30	.04	77	.9663
3 Some Familiarity	.00	.16	-.04	6 Speaker 7 Spanish-English Male 2	-.01	.16	-.05	7 Speaker 8 Dhivehi-English Male 1	.01	.23	.05	133	.9578
3 Some Familiarity	.00	.17	-.04	4 Speaker 5 Spanish-English Male 1	-.01	.16	-.05	7 Speaker 8 Dhivehi-English Male 1	.01	.23	.05	133	.9615
3 Some Familiarity	.31	.16	.13	5 Speaker 6 Arabic-English Male 1	.30	.16	.13	11 Speaker 12 Arabic-English Female 2	.01	.23	.03	133	.9768
4 Very Familiar	-.05	.30	-.09	3 Speaker 4 Dhivehi-English Female 2	-.05	.29	-.09	5 Speaker 6 Arabic-English Male 1	.00	.42	.01	37	.9948
2 Limited Familiarity	-.25	.17	-.09	7 Speaker 8 Dhivehi-English Male 1	-.25	.17	-.09	10 Speaker 11 Spanish-English Female 2	.00	.24	.01	121	.9909
2 Limited Familiarity	-.11	.18	-.01	8 Speaker 9 Spanish-English Female 1	-.11	.17	-.01	9 Speaker 10 Arabic-English Male 2	.00	.24	.01	121	.9913
3 Some Familiarity	-.01	.17	-.05	2 Speaker 3 Arabic-English Female 1	-.01	.16	-.05	7 Speaker 8 Dhivehi-English Male 1	.00	.23	.00	133	.9970
3 Some Familiarity	.00	.17	-.04	4 Speaker 5 Spanish-English Male 1	.00	.16	-.04	6 Speaker 7 Spanish-English Male 2	.00	.24	.00	133	.9973
1 No Familiarity	-.35	.23	-.12	4 Speaker 5 Spanish-English Male 1	-.35	.21	-.15	11 Speaker 12 Arabic-English Female 2	.00	.31	.00	77	.9975
2 Limited Familiarity	-.11	.17	-.01	5 Speaker 6 Arabic-English Male 1	-.11	.17	-.01	9 Speaker 10 Arabic-English Male 2	.00	.24	-.01	121	.9901
4 Very Familiar	.09	.30	-.01	6 Speaker 7 Spanish-English Male 2	.10	.29	-.01	11 Speaker 12 Arabic-English Female 2	.00	.42	-.01	37	.9939
1 No Familiarity	.04	.22	.07	2 Speaker 3 Arabic-English Female 1	.04	.21	.08	10 Speaker 11 Spanish-English Female 2	.00	.31	-.01	77	.9888
2 Limited Familiarity	-.11	.17	-.01	5 Speaker 6 Arabic-English Male 1	-.11	.18	-.01	8 Speaker 9 Spanish-English Female 1	-.01	.24	-.02	121	.9815
1 No Familiarity	.08	.22	.10	1 Speaker 2 Dhivehi-English Female 1	.09	.21	.11	7 Speaker 8 Dhivehi-English Male 1	-.01	.30	-.03	77	.9727
3 Some Familiarity	-.01	.17	-.05	2 Speaker 3 Arabic-English Female 1	.00	.17	-.04	4 Speaker 5 Spanish-English Male 1	-.01	.24	-.04	133	.9653
3 Some Familiarity	-.01	.17	-.05	2 Speaker 3 Arabic-English Female 1	.00	.16	-.04	6 Speaker 7 Spanish-English Male 2	-.01	.23	-.05	133	.9618
2 Limited Familiarity	-.03	.17	.03	1 Speaker 2 Dhivehi-English Female 1	-.02	.17	.04	6 Speaker 7 Spanish-English Male 2	-.01	.25	-.06	121	.9550
3 Some Familiarity	-.03	.16	-.06	1 Speaker 2 Dhivehi-English Female 1	-.01	.16	-.05	7 Speaker 8 Dhivehi-English Male 1	-.02	.23	-.07	133	.9475
3 Some Familiarity	-.03	.16	-.06	1 Speaker 2 Dhivehi-English Female 1	-.01	.17	-.05	2 Speaker 3 Arabic-English Female 1	-.02	.24	-.07	133	.9457
4 Very Familiar	-.05	.30	-.09	3 Speaker 4 Dhivehi-English Female 2	-.03	.30	-.08	9 Speaker 10 Arabic-English Male 2	-.02	.42	-.04	37	.9688
4 Very Familiar	-.07	.30	-.10	1 Speaker 2 Dhivehi-English Female 1	-.05	.29	-.09	5 Speaker 6 Arabic-English Male 1	-.02	.42	-.04	37	.9650
4 Very Familiar	-.05	.29	-.09	5 Speaker 6 Arabic-English Male 1	-.03	.30	-.08	9 Speaker 10 Arabic-English Male 2	-.02	.42	-.05	37	.9633
4 Very Familiar	-.07	.30	-.10	1 Speaker 2 Dhivehi-English Female 1	-.05	.30	-.09	3 Speaker 4 Dhivehi-English Female 2	-.02	.43	-.05	37	.9601
1 No Familiarity	-.37	.21	-.16	5 Speaker 6 Arabic-English Male 1	-.35	.21	-.15	11 Speaker 12 Arabic-English Female 2	-.02	.30	-.07	77	.9421
3 Some Familiarity	-.03	.16	-.06	1 Speaker 2 Dhivehi-English Female 1	.00	.17	-.04	4 Speaker 5 Spanish-English Male 1	-.03	.24	-.11	133	.9110
3 Some Familiarity	-.03	.16	-.06	1 Speaker 2 Dhivehi-English Female 1	.00	.16	-.04	6 Speaker 7 Spanish-English Male 2	-.03	.23	-.12	133	.9063
4 Very Familiar	.21	.31	.05	4 Speaker 5 Spanish-English Male 1	.24	.30	.07	10 Speaker 11 Spanish-English Female 2	-.03	.43	-.07	37	.9426
4 Very Familiar	-.07	.30	-.10	1 Speaker 2 Dhivehi-English Female 1	-.03	.30	-.08	9 Speaker 10 Arabic-English Male 2	-.04	.43	-.09	37	.9291

2 Limited Familiarity	.08	.17	.09	3 Speaker 4 Dhivehi-English Female 2	.12	.18	.10	4 Speaker 5 Spanish-English Male 1	-.04	.25	-.18	121	.8588
4 Very Familiar	.19	.30	.04	8 Speaker 9 Spanish-English Female 1	.24	.30	.07	10 Speaker 11 Spanish-English Female 2	-.05	.43	-.11	37	.9123
3 Some Familiarity	.00	.16	-.04	6 Speaker 7 Spanish-English Male 2	.05	.16	-.02	12 Speaker 13 Dhivehi-English Male 2	-.05	.23	-.21	133	.8311
3 Some Familiarity	.00	.17	-.04	4 Speaker 5 Spanish-English Male 1	.05	.16	-.02	12 Speaker 13 Dhivehi-English Male 2	-.05	.23	-.21	133	.8320
2 Limited Familiarity	-.30	.18	-.10	2 Speaker 3 Arabic-English Female 1	-.25	.17	-.09	10 Speaker 11 Spanish-English Female 2	-.05	.25	-.21	121	.8373
2 Limited Familiarity	-.30	.18	-.10	2 Speaker 3 Arabic-English Female 1	-.25	.17	-.09	7 Speaker 8 Dhivehi-English Male 1	-.05	.25	-.21	121	.8303
1 No Familiarity	.04	.22	.07	2 Speaker 3 Arabic-English Female 1	.09	.21	.11	7 Speaker 8 Dhivehi-English Male 1	-.06	.31	-.18	77	.8570
3 Some Familiarity	.24	.16	.09	9 Speaker 10 Arabic-English Male 2	.30	.16	.13	11 Speaker 12 Arabic-English Female 2	-.06	.23	-.26	133	.7988
3 Some Familiarity	-.01	.17	-.05	2 Speaker 3 Arabic-English Female 1	.05	.16	-.02	12 Speaker 13 Dhivehi-English Male 2	-.06	.23	-.26	133	.7966
3 Some Familiarity	-.01	.16	-.05	7 Speaker 8 Dhivehi-English Male 1	.05	.16	-.02	12 Speaker 13 Dhivehi-English Male 2	-.06	.23	-.27	133	.7887
2 Limited Familiarity	-.02	.17	.04	6 Speaker 7 Spanish-English Male 2	.05	.17	.08	12 Speaker 13 Dhivehi-English Male 2	-.07	.24	-.30	121	.7655
3 Some Familiarity	-.03	.16	-.06	1 Speaker 2 Dhivehi-English Female 1	.05	.16	-.02	12 Speaker 13 Dhivehi-English Male 2	-.08	.23	-.33	133	.7406
2 Limited Familiarity	-.25	.17	-.09	7 Speaker 8 Dhivehi-English Male 1	-.17	.17	-.04	11 Speaker 12 Arabic-English Female 2	-.08	.24	-.34	121	.7311
2 Limited Familiarity	-.25	.17	-.09	10 Speaker 11 Spanish-English Female 2	-.17	.17	-.04	11 Speaker 12 Arabic-English Female 2	-.09	.24	-.36	121	.7198
2 Limited Familiarity	-.03	.17	.03	1 Speaker 2 Dhivehi-English Female 1	.05	.17	.08	12 Speaker 13 Dhivehi-English Male 2	-.09	.24	-.36	121	.7230
1 No Familiarity	-.35	.23	-.12	4 Speaker 5 Spanish-English Male 1	-.27	.21	-.10	9 Speaker 10 Arabic-English Male 2	-.09	.31	-.28	77	.7828
2 Limited Familiarity	-.11	.17	-.01	5 Speaker 6 Arabic-English Male 1	-.02	.17	.04	6 Speaker 7 Spanish-English Male 2	-.09	.24	-.39	121	.6962
1 No Familiarity	-.07	.22	.01	6 Speaker 7 Spanish-English Male 2	.02	.21	.07	12 Speaker 13 Dhivehi-English Male 2	-.10	.30	-.32	77	.7518
4 Very Familiar	.09	.30	-.01	6 Speaker 7 Spanish-English Male 2	.19	.30	.04	8 Speaker 9 Spanish-English Female 1	-.10	.42	-.23	37	.8232
2 Limited Familiarity	-.03	.17	.03	1 Speaker 2 Dhivehi-English Female 1	.08	.17	.09	3 Speaker 4 Dhivehi-English Female 2	-.11	.24	-.44	121	.6623
3 Some Familiarity	.00	.16	-.04	6 Speaker 7 Spanish-English Male 2	.11	.16	.02	10 Speaker 11 Spanish-English Female 2	-.11	.23	-.46	133	.6433
1 No Familiarity	-.37	.21	-.16	5 Speaker 6 Arabic-English Male 1	-.27	.21	-.10	9 Speaker 10 Arabic-English Male 2	-.11	.30	-.36	77	.7185
3 Some Familiarity	.00	.17	-.04	4 Speaker 5 Spanish-English Male 1	.11	.16	.02	10 Speaker 11 Spanish-English Female 2	-.11	.24	-.46	133	.6480
3 Some Familiarity	.13	.16	.03	8 Speaker 9 Spanish-English Female 1	.24	.16	.09	9 Speaker 10 Arabic-English Male 2	-.11	.23	-.47	133	.6389
1 No Familiarity	-.07	.22	.01	6 Speaker 7 Spanish-English Male 2	.04	.21	.08	10 Speaker 11 Spanish-English Female 2	-.11	.30	-.37	77	.7099
1 No Familiarity	.08	.22	.10	1 Speaker 2 Dhivehi-English Female 1	.19	.21	.16	3 Speaker 4 Dhivehi-English Female 2	-.11	.30	-.37	77	.7097
3 Some Familiarity	-.01	.17	-.05	2 Speaker 3 Arabic-English Female 1	.11	.16	.02	10 Speaker 11 Spanish-English Female 2	-.12	.23	-.50	133	.6150
3 Some Familiarity	-.01	.16	-.05	7 Speaker 8 Dhivehi-English Male 1	.11	.16	.02	10 Speaker 11 Spanish-English Female 2	-.12	.23	-.52	133	.6037
4 Very Familiar	-.03	.30	-.08	9 Speaker 10 Arabic-English Male 2	.10	.29	-.01	11 Speaker 12 Arabic-English Female 2	-.13	.42	-.31	37	.7575
3 Some Familiarity	.00	.16	-.04	6 Speaker 7 Spanish-English Male 2	.13	.16	.03	8 Speaker 9 Spanish-English Female 1	-.13	.23	-.57	133	.5693
3 Some Familiarity	.00	.17	-.04	4 Speaker 5 Spanish-English Male 1	.13	.16	.03	8 Speaker 9 Spanish-English Female 1	-.13	.24	-.56	133	.5752
3 Some Familiarity	-.03	.16	-.06	1 Speaker 2 Dhivehi-English Female 1	.11	.16	.02	10 Speaker 11 Spanish-English Female 2	-.13	.23	-.58	133	.5626
1 No Familiarity	-.35	.23	-.12	4 Speaker 5 Spanish-English Male 1	-.22	.22	-.06	8 Speaker 9 Spanish-English Female 1	-.13	.32	-.42	77	.6754
2 Limited Familiarity	-.30	.18	-.10	2 Speaker 3 Arabic-English Female 1	-.17	.17	-.04	11 Speaker 12 Arabic-English Female 2	-.14	.25	-.55	121	.5839
2 Limited Familiarity	-.25	.17	-.09	7 Speaker 8 Dhivehi-English Male 1	-.11	.17	-.01	9 Speaker 10 Arabic-English Male 2	-.14	.24	-.58	121	.5639
2 Limited Familiarity	-.25	.17	-.09	7 Speaker 8 Dhivehi-English Male 1	-.11	.18	-.01	8 Speaker 9 Spanish-English Female 1	-.14	.25	-.58	121	.5647
4 Very Familiar	.09	.30	-.01	6 Speaker 7 Spanish-English Male 2	.24	.30	.07	10 Speaker 11 Spanish-English Female 2	-.14	.43	-.34	37	.7391
3 Some Familiarity	-.01	.17	-.05	2 Speaker 3 Arabic-English Female 1	.13	.16	.03	8 Speaker 9 Spanish-English Female 1	-.14	.24	-.61	133	.5437
4 Very Familiar	-.05	.30	-.09	3 Speaker 4 Dhivehi-English Female 2	.09	.30	-.01	6 Speaker 7 Spanish-English Male 2	-.14	.42	-.34	38	.7359
3 Some Familiarity	-.01	.16	-.05	7 Speaker 8 Dhivehi-English Male 1	.13	.16	.03	8 Speaker 9 Spanish-English Female 1	-.14	.23	-.63	133	.5315
4 Very Familiar	-.05	.29	-.09	5 Speaker 6 Arabic-English Male 1	.09	.30	-.01	6 Speaker 7 Spanish-English Male 2	-.15	.42	-.35	37	.7288
4 Very Familiar	-.05	.30	-.09	3 Speaker 4 Dhivehi-English Female 2	.10	.29	-.01	11 Speaker 12 Arabic-English Female 2	-.15	.42	-.35	37	.7283
4 Very Familiar	-.05	.29	-.09	5 Speaker 6 Arabic-English Male 1	.10	.29	-.01	11 Speaker 12 Arabic-English Female 2	-.15	.42	-.36	37	.7211
2 Limited Familiarity	-.03	.17	.03	1 Speaker 2 Dhivehi-English Female 1	.12	.18	.10	4 Speaker 5 Spanish-English Male 1	-.15	.25	-.60	121	.5468
1 No Familiarity	-.37	.21	-.16	5 Speaker 6 Arabic-English Male 1	-.22	.22	-.06	8 Speaker 9 Spanish-English Female 1	-.16	.30	-.51	77	.6112
1 No Familiarity	.04	.22	.07	2 Speaker 3 Arabic-English Female 1	.19	.21	.16	3 Speaker 4 Dhivehi-English Female 2	-.16	.31	-.51	77	.6086
3 Some Familiarity	-.03	.16	-.06	1 Speaker 2 Dhivehi-English Female 1	.13	.16	.03	8 Speaker 9 Spanish-English Female 1	-.16	.23	-.69	133	.4937
2 Limited Familiarity	-.11	.18	-.01	8 Speaker 9 Spanish-English Female 1	.05	.17	.08	12 Speaker 13 Dhivehi-English Male 2	-.16	.24	-.66	121	.5098
1 No Familiarity	-.07	.22	.01	6 Speaker 7 Spanish-English Male 2	.09	.21	.11	7 Speaker 8 Dhivehi-English Male 1	-.16	.30	-.54	77	.5889
2 Limited Familiarity	-.11	.17	-.01	9 Speaker 10 Arabic-English Male 2	.05	.17	.08	12 Speaker 13 Dhivehi-English Male 2	-.16	.24	-.69	121	.4936
4 Very Familiar	-.07	.30	-.10	1 Speaker 2 Dhivehi-English Female 1	.09	.30	-.01	6 Speaker 7 Spanish-English Male 2	-.17	.43	-.39	37	.7003
2 Limited Familiarity	-.11	.17	-.01	5 Speaker 6 Arabic-English Male 1	.05	.17	.08	12 Speaker 13 Dhivehi-English Male 2	-.17	.24	-.70	121	.4850
3 Some Familiarity	.13	.16	.03	8 Speaker 9 Spanish-English Female 1	.30	.16	.13	11 Speaker 12 Arabic-English Female 2	-.17	.23	-.73	133	.4663
4 Very Familiar	-.07	.30	-.10	1 Speaker 2 Dhivehi-English Female 1	.10	.29	-.01	11 Speaker 12 Arabic-English Female 2	-.17	.42	-.40	37	.6927
3 Some Familiarity	-.20	.16	-.15	3 Speaker 4 Dhivehi-English Female 2	-.01	.16	-.05	7 Speaker 8 Dhivehi-English Male 1	-.19	.23	-.81	133	.4215
2 Limited Familiarity	-.30	.18	-.10	2 Speaker 3 Arabic-English Female 1	-.11	.17	-.01	5 Speaker 6 Arabic-English Male 1	-.19	.25	-.76	121	.4464
3 Some Familiarity	.11	.16	.02	10 Speaker 11 Spanish-English Female 2	.30	.16	.13	11 Speaker 12 Arabic-English Female 2	-.19	.23	-.84	133	.4023
2 Limited Familiarity	-.30	.18	-.10	2 Speaker 3 Arabic-English Female 1	-.11	.17	-.01	9 Speaker 10 Arabic-English Male 2	-.19	.25	-.77	121	.4401
2 Limited Familiarity	-.30	.18	-.10	2 Speaker 3 Arabic-English Female 1	-.11	.18	-.01	8 Speaker 9 Spanish-English Female 1	-.20	.25	-.77	121	.4427
3 Some Familiarity	-.20	.16	-.15	3 Speaker 4 Dhivehi-English Female 2	.00	.17	-.04	4 Speaker 5 Spanish-English Male 1	-.20	.24	-.83	133	.4071
3 Some Familiarity	-.20	.16	-.15	3 Speaker 4 Dhivehi-English Female 2	.00	.16	-.04	6 Speaker 7 Spanish-English Male 2	-.20	.23	-.85	133	.3951
2 Limited Familiarity	-.17	.17	-.04	11 Speaker 12 Arabic-English Female 2	.05	.17	.08	12 Speaker 13 Dhivehi-English Male 2	-.22	.24	-.93	121	.3564
4 Very Familiar	-.05	.30	-.09	3 Speaker 4 Dhivehi-English Female 2	.19	.30	.04	8 Speaker 9 Spanish-English Female 1	-.24	.42	-.56	37	.5760
3 Some Familiarity	.00	.16	-.04	6 Speaker 7 Spanish-English Male 2	.24	.16	.09	9 Speaker 10 Arabic-English Male 2	-.24	.23	-1.04	133	.3000
3 Some Familiarity	.00	.17	-.04	4 Speaker 5 Spanish-English Male 1	.24	.16	.09	9 Speaker 10 Arabic-English Male 2	-.24	.24	-1.02	133	.3087
4 Very Familiar	-.05	.29	-.09	5 Speaker 6 Arabic-English Male 1	.19	.30	.04	8 Speaker 9 Spanish-English Female 1	-.24	.42	-.58	37	.5684
1 No Familiarity	-.22	.22	-.06	8 Speaker 9 Spanish-English Female 1	.02	.21	.07	12 Speaker 13 Dhivehi-English Male 2	-.24	.30	-.80	77	.4273
3 Some Familiarity	-.20	.16	-.15	3 Speaker 4 Dhivehi-English Female 2	.05	.16	-.02	12 Speaker 13 Dhivehi-English Male 2	-.25	.23	-1.07	133	.2851
3 Some Familiarity	-.01	.17	-.05	2 Speaker 3 Arabic-English Female 1	.24	.16	.09	9 Speaker 10 Arabic-English Male 2	-.25	.24	-1.07	133	.2862
3 Some Familiarity	-.01	.16	-.05	7 Speaker 8 Dhivehi-English Male 1	.24	.16	.09	9 Speaker 10 Arabic-English Male 2	-.25	.23	-1.10	133	.2731
4 Very Familiar	-.05	.30	-.09	3 Speaker 4 Dhivehi-English Female 2	.21	.31	.05	4 Speaker 5 Spanish-English Male 1	-.26	.43	-.59	37	.5563
1 No Familiarity	-.22	.22	-.06	8 Speaker 9 Spanish-English Female 1	.04	.21	.08	10 Speaker 11 Spanish-English Female 2	-.26	.30	-.85	77	.3971

4 Very Familiar	-.07	.30	-.10	1 Speaker 2 Dhivehi-English Female 1	.19	.30	.04	8 Speaker 9 Spanish-English Female 1	-.26	.43	-.61	37	.5452
3 Some Familiarity	-.03	.16	-.06	1 Speaker 2 Dhivehi-English Female 1	.24	.16	.09	9 Speaker 10 Arabic-English Male 2	-.27	.23	-1.15	133	.2504
4 Very Familiar	-.03	.30	-.08	9 Speaker 10 Arabic-English Male 2	.24	.30	.07	10 Speaker 11 Spanish-English Female 2	-.27	.42	-.64	37	.5288
4 Very Familiar	-.07	.30	-.10	1 Speaker 2 Dhivehi-English Female 1	.21	.31	.05	4 Speaker 5 Spanish-English Male 1	-.28	.43	-.64	37	.5265
1 No Familiarity	-.35	.23	-.12	4 Speaker 5 Spanish-English Male 1	-.07	.22	.01	6 Speaker 7 Spanish-English Male 2	-.28	.32	-.89	77	.3771
2 Limited Familiarity	-.30	.18	-.10	2 Speaker 3 Arabic-English Female 1	-.02	.17	.04	6 Speaker 7 Spanish-English Male 2	-.28	.25	-1.13	121	.2615
4 Very Familiar	-.05	.30	-.09	3 Speaker 4 Dhivehi-English Female 2	.24	.30	.07	10 Speaker 11 Spanish-English Female 2	-.29	.43	-.67	37	.5047
4 Very Familiar	-.05	.29	-.09	5 Speaker 6 Arabic-English Male 1	.24	.30	.07	10 Speaker 11 Spanish-English Female 2	-.29	.42	-.69	37	.4970
1 No Familiarity	-.27	.21	-.10	9 Speaker 10 Arabic-English Male 2	.02	.21	.07	12 Speaker 13 Dhivehi-English Male 2	-.29	.30	-.98	77	.3286
4 Very Familiar	.21	.31	.05	4 Speaker 5 Spanish-English Male 1	.50	.29	.23	7 Speaker 8 Dhivehi-English Male 1	-.30	.43	-.70	37	.4912
3 Some Familiarity	.00	.16	-.04	6 Speaker 7 Spanish-English Male 2	.30	.16	.13	11 Speaker 12 Arabic-English Female 2	-.30	.23	-1.31	133	.1936
3 Some Familiarity	.00	.17	-.04	4 Speaker 5 Spanish-English Male 1	.30	.16	.13	11 Speaker 12 Arabic-English Female 2	-.30	.23	-1.28	133	.2022
1 No Familiarity	-.37	.21	-.16	5 Speaker 6 Arabic-English Male 1	-.07	.22	.01	6 Speaker 7 Spanish-English Male 2	-.30	.30	-1.00	77	.3181
2 Limited Familiarity	-.25	.17	-.09	7 Speaker 8 Dhivehi-English Male 1	.05	.17	.08	12 Speaker 13 Dhivehi-English Male 2	-.30	.24	-1.26	121	.2100
3 Some Familiarity	-.20	.16	-.15	3 Speaker 4 Dhivehi-English Female 2	.11	.16	.02	10 Speaker 11 Spanish-English Female 2	-.30	.23	-1.32	133	.1900
2 Limited Familiarity	-.25	.17	-.09	10 Speaker 11 Spanish-English Female 2	.05	.17	.08	12 Speaker 13 Dhivehi-English Male 2	-.31	.24	-1.28	121	.2014
3 Some Familiarity	.00	.17	-.04	4 Speaker 5 Spanish-English Male 1	.31	.16	.13	5 Speaker 6 Arabic-English Male 1	-.31	.23	-1.31	133	.1919
1 No Familiarity	-.27	.21	-.10	9 Speaker 10 Arabic-English Male 2	.04	.21	.08	10 Speaker 11 Spanish-English Female 2	-.31	.30	-1.04	77	.3028
4 Very Familiar	-.07	.30	-.10	1 Speaker 2 Dhivehi-English Female 1	.24	.30	.07	10 Speaker 11 Spanish-English Female 2	-.31	.43	-.72	37	.4765
3 Some Familiarity	-.01	.17	-.05	2 Speaker 3 Arabic-English Female 1	.30	.16	.13	11 Speaker 12 Arabic-English Female 2	-.31	.23	-1.33	133	.1850
3 Some Familiarity	-.01	.16	-.05	7 Speaker 8 Dhivehi-English Male 1	.30	.16	.13	11 Speaker 12 Arabic-English Female 2	-.31	.23	-1.37	133	.1733
3 Some Familiarity	-.01	.17	-.05	2 Speaker 3 Arabic-English Female 1	.31	.16	.13	5 Speaker 6 Arabic-English Male 1	-.32	.23	-1.36	133	.1753
3 Some Familiarity	-.03	.16	-.06	1 Speaker 2 Dhivehi-English Female 1	.30	.16	.13	11 Speaker 12 Arabic-English Female 2	-.33	.23	-1.42	133	.1578
3 Some Familiarity	-.20	.16	-.15	3 Speaker 4 Dhivehi-English Female 2	.13	.16	.03	8 Speaker 9 Spanish-English Female 1	-.33	.23	-1.42	134	.1574
3 Some Familiarity	-.03	.16	-.06	1 Speaker 2 Dhivehi-English Female 1	.31	.16	.13	5 Speaker 6 Arabic-English Male 1	-.33	.23	-1.45	133	.1490
2 Limited Familiarity	-.30	.18	-.10	2 Speaker 3 Arabic-English Female 1	.05	.17	.08	12 Speaker 13 Dhivehi-English Male 2	-.36	.25	-1.43	121	.1547
1 No Familiarity	-.35	.21	-.15	11 Speaker 12 Arabic-English Female 2	.02	.21	.07	12 Speaker 13 Dhivehi-English Male 2	-.38	.30	-1.27	77	.2073
1 No Familiarity	-.35	.23	-.12	4 Speaker 5 Spanish-English Male 1	.02	.21	.07	12 Speaker 13 Dhivehi-English Male 2	-.38	.31	-1.21	77	.2315
2 Limited Familiarity	-.30	.18	-.10	2 Speaker 3 Arabic-English Female 1	.08	.17	.09	3 Speaker 4 Dhivehi-English Female 2	-.38	.25	-1.50	121	.1351
1 No Familiarity	-.35	.23	-.12	4 Speaker 5 Spanish-English Male 1	.04	.21	.08	10 Speaker 11 Spanish-English Female 2	-.39	.31	-1.26	77	.2127
1 No Familiarity	-.37	.21	-.16	5 Speaker 6 Arabic-English Male 1	.02	.21	.07	12 Speaker 13 Dhivehi-English Male 2	-.40	.30	-1.34	77	.1832
4 Very Familiar	.09	.30	-.01	6 Speaker 7 Spanish-English Male 2	.50	.29	.23	7 Speaker 8 Dhivehi-English Male 1	-.41	.42	-.97	37	.3374
1 No Familiarity	-.37	.21	-.16	5 Speaker 6 Arabic-English Male 1	.04	.21	.08	10 Speaker 11 Spanish-English Female 2	-.41	.30	-1.40	77	.1669
2 Limited Familiarity	-.30	.18	-.10	2 Speaker 3 Arabic-English Female 1	.12	.18	.10	4 Speaker 5 Spanish-English Male 1	-.42	.26	-1.64	121	.1029
3 Some Familiarity	-.20	.16	-.15	3 Speaker 4 Dhivehi-English Female 2	.24	.16	.09	9 Speaker 10 Arabic-English Male 2	-.44	.23	-1.89	133	.0610
1 No Familiarity	-.35	.23	-.12	4 Speaker 5 Spanish-English Male 1	.09	.21	.11	7 Speaker 8 Dhivehi-English Male 1	-.45	.31	-1.42	77	.1598
1 No Familiarity	-.37	.21	-.16	5 Speaker 6 Arabic-English Male 1	.09	.21	.11	7 Speaker 8 Dhivehi-English Male 1	-.47	.30	-1.57	77	.1211
3 Some Familiarity	-.20	.16	-.15	3 Speaker 4 Dhivehi-English Female 2	.30	.16	.13	11 Speaker 12 Arabic-English Female 2	-.50	.23	-2.16	133	.0323
3 Some Familiarity	-.20	.16	-.15	3 Speaker 4 Dhivehi-English Female 2	.31	.16	.13	5 Speaker 6 Arabic-English Male 1	-.50	.23	-2.20	133	.0299
4 Very Familiar	-.05	.30	-.09	3 Speaker 4 Dhivehi-English Female 2	.50	.29	.23	7 Speaker 8 Dhivehi-English Male 1	-.55	.42	-1.31	37	.1967
4 Very Familiar	-.05	.29	-.09	5 Speaker 6 Arabic-English Male 1	.50	.29	.23	7 Speaker 8 Dhivehi-English Male 1	-.55	.42	-1.33	38	.1906
4 Very Familiar	-.07	.30	-.10	1 Speaker 2 Dhivehi-English Female 1	.50	.29	.23	7 Speaker 8 Dhivehi-English Male 1	-.57	.42	-1.36	37	.1832
4 Very Familiar	-.07	.30	-.10	1 Speaker 2 Dhivehi-English Female 1	.72	.30	.33	2 Speaker 3 Arabic-English Female 1	-.79	.43	-1.86	37	.0702

The Facets analyses output of bias/interaction - Dhivehi-English

```
Rasch Analysis of Expert Ratings 08/01/2015 15:30:27
Table 9.3  Bias Iteration Report.

Bias/Interaction: 2. Dhivehi English Familiarity Level, 3. Speaker

There are empirically 48 Bias terms
+------------------------------------------------------------+
| Iteration      Max. Score Residual      Max. Logit Change |
|             Elements    %  Categories   Elements    Steps |
|------------------------------------------------------------|
| BIAS   1      11.4564 -144.                1.0000          |
| BIAS   2       7.4020 -69.5                1.0000          |
| BIAS   3       3.5526 -15.8                 .9932          |
| BIAS   4        .3979  -8.0                 .1458          |
| BIAS   5        .0510  -4.2                 .0510          |
| BIAS   6       -.0217  -2.2                -.0217          |
| BIAS   7       -.0114  -1.1                -.0114          |
| BIAS   8       -.0060   -.6                -.0060          |
+------------------------------------------------------------+

Rasch Analysis of Expert Ratings 08/01/2015 15:30:27
Table 11.3  Bias/Interaction Measurement Summary.

Bias/Interaction: 2. Dhivehi English Familiarity Level, 3. Speaker

+---------------------------------------------------+
| Cat  Score  Exp.  Resd StRes|                     |
|-----------------------------+---------------------|
| 2.83  2.83  2.83   .00  .00 | Mean (Count: 2280) |
|  .97   .97   .66   .71  .97 | S.D. (Population)  |
|  .97   .97   .66   .71  .97 | S.D. (Sample)      |
+---------------------------------------------------+
Count of measurable responses          = 2280
Raw-score variance of observations     =   0.94 100.00%
Variance explained by Rasch measures   =   0.39  42.00%
Variance of residuals                  =   0.55  58.00%
Variance explained by bias/interactions =  0.04   4.32%
Variance remaining in residuals        =   0.50  53.69%

Rasch Analysis of Expert Ratings 08/01/2015 15:30:27
Table 12.3  Bias/Interaction Summary Report.

Bias/Interaction: 2. Dhivehi English Familiarity Level, 3. Speaker (higher score = higher bias measure)

Bias/Interaction Size:
          1  21  1 1 1 111 22466612 2 11             22       1
 +--------Q+---------S---------M---------S---------+Q--------+---------+
-3        -2        -1         0         1         2         3         4

Bias/Interaction Significance:
   21  1   1  12361327144 2  2   1    1  2    1
 +-Q--+----+S---+----M----+---S+----+-Q--+----+----+
-4   -3   -2   -1    0    1    2    3    4    5    6


Rasch Analysis of Expert Ratings 08/01/2015 15:30:27
Table 13.3.1  Bias/Interaction Report (arranged by mN).

Bias/Interaction: 2. Dhivehi English Familiarity Level, 3. Speaker (higher score = higher bias measure)

+------------------------------------------------------------------------------------------------------------------------------------------------------------------+
|Observd  Expctd  Observd  Obs-Exp|  Bias  Model                    |Infit Outfit|    Dhivehi English Familiarity LSpeaker                                        |
|  Score   Score    Count  Average|  Size   S.E.      t   d.f. Prob. | MnSq  MnSq | Sq N Dhivehi English Fam  measr Nu Speaker                               measr |
|---------------------------------+---------------------------------+------------+-------------------------------------------------------------------------------|
|    31      19.54     7      1.64|   3.14   .62   5.05     6 .0023 |  1.5   1.5 |  4 4 Very Familiar          .29  1 Speaker 2 Dhivehi-English Female 1    -.54 |
|    30      21.51     7      1.21|   2.29   .58   3.93     6 .0077 |  1.5   1.3 | 28 4 Very Familiar          .29  7 Speaker 8 Dhivehi-English Male 1      -.06 |
|    32      24.80     7      1.03|   2.27   .68   3.32     6 .0161 |   .7    .6 | 48 4 Very Familiar          .29 12 Speaker 13 Dhivehi-English Male 2      .75 |
|    29      20.55     7      1.21|   2.20   .56   3.96     6 .0075 |  1.9   2.1 | 12 4 Very Familiar          .29  3 Speaker 4 Dhivehi-English Female 2    -.29 |
```

3	1.84	1	1.16	2.19	1.29	1.69	1	.3399	.0	.0	7	3 Some Familiarity	-.42	2 Speaker 3 Arabic-English Female 1	-.98
51	43.51	17	.44	.78	.32	2.44	16	.0269	.6	.6	30	2 Limited Familiarity	.11	8 Speaker 9 Spanish-English Female 1	-.62
3	2.58	1	.42	.72	1.29	.55	1	.6782	.0	.0	19	3 Some Familiarity	-.42	5 Speaker 6 Arabic-English Male 1	.50
65	60.12	17	.29	.53	.34	1.58	16	.1346	.7	.8	38	2 Limited Familiarity	.11	10 Speaker 11 Spanish-English Female 2	1.09
2	1.79	1	.21	.47	1.46	.33	1	.7996	.0	.0	15	3 Some Familiarity	-.42	4 Speaker 5 Spanish-English Male 1	-1.10
60	56.90	17	.18	.32	.32	.99	16	.3359	.4	.4	46	2 Limited Familiarity	.11	12 Speaker 13 Dhivehi-English Male 2	.75
25	23.77	7	.18	.31	.50	.61	6	.5622	1.5	1.5	20	4 Very Familiar	.29	5 Speaker 6 Arabic-English Male 1	.50
57	55.46	17	.09	.16	.32	.49	16	.6297	.8	.8	42	2 Limited Familiarity	.11	11 Speaker 12 Arabic-English Female 2	.61
3	2.93	1	.07	.12	1.29	.10	1	.9391	.0	.0	35	3 Some Familiarity	-.42	9 Speaker 10 Arabic-English Male 2	1.09
3	2.93	1	.07	.12	1.29	.10	1	.9391	.0	.0	39	3 Some Familiarity	-.42	10 Speaker 11 Spanish-English Female 2	1.09
574	564.83	165	.06	.10	.10	.95	164	.3423	.7	.7	33	1 No Familiarity	.02	9 Speaker 10 Arabic-English Male 2	1.09
382	376.06	165	.04	.07	.11	.65	164	.5164	1.3	1.3	5	1 No Familiarity	.02	2 Speaker 3 Arabic-English Female 1	-.98
428	422.70	165	.03	.06	.11	.56	164	.5760	.8	.8	21	1 No Familiarity	.02	6 Speaker 7 Spanish-English Male 2	-.44
370	365.71	165	.03	.05	.11	.47	164	.6355	.7	.7	13	1 No Familiarity	.02	4 Speaker 5 Spanish-English Male 1	-1.10
523	519.68	165	.02	.04	.10	.34	164	.7326	1.0	1.0	41	1 No Familiarity	.02	11 Speaker 12 Arabic-English Female 2	.61
567	564.83	165	.01	.02	.10	.23	164	.8222	.9	.9	37	1 No Familiarity	.02	10 Speaker 11 Spanish-English Female 2	1.09
510	509.27	165	.00	.01	.10	.08	164	.9400	.8	.8	17	1 No Familiarity	.02	5 Speaker 6 Arabic-English Male 1	.50
49	48.93	17	.00	.01	.32	.02	16	.9816	1.1	1.1	26	2 Limited Familiarity	.11	7 Speaker 8 Dhivehi-English Male 1	-.06
2	2.00	1	.00	.00	1.46	.00	1	.9996	.0	.0	31	3 Some Familiarity	-.42	8 Speaker 9 Spanish-English Female 1	-.62
406	406.28	165	.00	.00	.11	-.03	164	.9758	.9	.9	29	1 No Familiarity	.02	8 Speaker 9 Spanish-English Female 1	-.62
431	435.67	165	-.03	-.05	.11	-.49	164	.6247	.7	.7	9	1 No Familiarity	.02	3 Speaker 4 Dhivehi-English Female 2	-.29
2	2.04	1	-.04	-.08	1.46	-.05	1	.9655	.0	.0	3	3 Some Familiarity	-.42	1 Speaker 2 Dhivehi-English Female 1	-.54
449	457.29	165	-.05	-.09	.10	-.86	164	.3897	1.5	1.5	25	1 No Familiarity	.02	7 Speaker 8 Dhivehi-English Male 1	-.06
405	413.20	165	-.05	-.09	.11	-.87	164	.3846	1.2	1.2	1	1 No Familiarity	.02	1 Speaker 2 Dhivehi-English Female 1	-.54
524	533.56	165	-.06	-.10	.10	-.99	164	.3258	.7	.8	45	1 No Familiarity	.02	12 Speaker 13 Dhivehi-English Male 2	.75
39	40.28	17	-.08	-.14	.34	-.43	16	.6753	1.0	1.0	6	2 Limited Familiarity	.11	2 Speaker 3 Arabic-English Female 1	-.98
2	2.09	1	-.09	-.19	1.46	-.13	1	.9195	.0	.0	23	3 Some Familiarity	-.42	6 Speaker 7 Spanish-English Male 2	-.44
58	60.12	17	-.12	-.22	.32	-.69	16	.5010	.3	.3	34	2 Limited Familiarity	.11	9 Speaker 10 Arabic-English Male 2	1.09
43	45.26	17	-.13	-.24	.33	-.73	16	.4749	1.2	1.2	22	2 Limited Familiarity	.11	6 Speaker 7 Spanish-English Male 2	-.44
52	54.38	17	-.14	-.24	.32	-.76	16	.4593	.9	.9	18	2 Limited Familiarity	.11	5 Speaker 6 Arabic-English Male 1	.50
37	39.16	17	-.13	-.25	.35	-.73	16	.4754	.4	.4	14	2 Limited Familiarity	.11	4 Speaker 5 Spanish-English Male 1	-1.10
2	2.16	1	-.16	-.33	1.46	-.23	1	.8588	.0	.0	11	3 Some Familiarity	-.42	3 Speaker 4 Dhivehi-English Female 2	-.29
41	44.25	17	-.19	-.35	.33	-1.06	16	.3065	.7	.7	2	2 Limited Familiarity	.11	1 Speaker 2 Dhivehi-English Female 1	-.54
43	46.64	17	-.21	-.38	.33	-1.17	16	.2589	.5	.5	10	2 Limited Familiarity	.11	3 Speaker 4 Dhivehi-English Female 2	-.29
2	2.28	1	-.28	-.56	1.46	-.39	1	.7652	.0	.0	27	3 Some Familiarity	-.42	7 Speaker 8 Dhivehi-English Male 1	-.06
15	17.37	7	-.34	-.66	.54	-1.22	6	.2690	1.9	2.0	16	4 Very Familiar	.29	4 Speaker 5 Spanish-English Male 1	-1.10
17	19.97	7	-.42	-.76	.52	-1.48	6	.1903	1.2	1.3	24	4 Very Familiar	.29	6 Speaker 7 Spanish-English Male 2	-.44
20	24.21	7	-.60	-1.04	.50	-2.08	6	.0827	1.3	1.3	44	4 Very Familiar	.29	11 Speaker 12 Arabic-English Female 2	.61
2	2.64	1	-.64	-1.23	1.46	-.84	1	.5547	.0	.0	43	3 Some Familiarity	-.42	11 Speaker 12 Arabic-English Female 2	.61
2	2.73	1	-.73	-1.37	1.46	-.94	1	.5190	.0	.0	47	3 Some Familiarity	-.42	12 Speaker 13 Dhivehi-English Male 2	.75
12	17.85	7	-.84	-1.75	.60	-2.90	6	.0274	1.7	1.5	8	4 Very Familiar	.29	2 Speaker 3 Arabic-English Female 1	-.98
19	26.10	7	-1.01	-1.77	.50	-3.52	6	.0126	1.3	1.4	36	4 Very Familiar	.29	9 Speaker 10 Arabic-English Male 2	1.09
19	26.10	7	-1.01	-1.77	.50	-3.52	6	.0126	.9	1.0	40	4 Very Familiar	.29	10 Speaker 11 Spanish-English Female 2	1.09
12	19.23	7	-1.03	-2.10	.60	-3.49	6	.0131	1.2	1.3	32	4 Very Familiar	.29	8 Speaker 9 Spanish-English Female 1	-.62
134.4	134.44	47.5	.00	.00	.60	-.02			.8	.8	Mean (Count: 48)				
194.2	194.14	68.1	.55	1.04	.49	1.74			.6	.6	S.D. (Population)				
196.3	196.20	68.8	.55	1.05	.50	1.76			.6	.6	S.D. (Sample)				

Fixed (all = 0) chi-square: 145.5 d.f.: 48 significance (probability): .00

Rasch Analysis of Expert Ratings 08/01/2015 15:30:27
Table 14.3.1.2 Bias/Interaction Pairwise Report (arranged by mN).

Bias/Interaction: 2. Dhivehi English Familiarity Level, 3. Speaker

Target Nu Speaker	Target Measr	S.E.	Obs-Exp Average	Context N Dhivehi English Fam	Target Measr	S.E.	Obs-Exp Average	Context N Dhivehi English Fam	Target Contrast	Joint S.E.	t	Welch d.f.	Prob.
2 Speaker 3 Arabic-English Female 1	1.21	1.29	1.16	3 Some Familiarity	-2.72	.60	-.84	4 Very Familiar	3.93	1.43	2.76	0	.0000
8 Speaker 9 Spanish-English Female 1	.15	.32	.44	2 Limited Familiarity	-2.72	.60	-1.03	4 Very Familiar	2.88	.68	4.22	12	.0012
10 Speaker 11 Spanish-English Female 2	1.62	.34	.29	2 Limited Familiarity	-.68	.50	-1.01	4 Very Familiar	2.30	.60	3.80	14	.0019
8 Speaker 9 Spanish-English Female 1	-.62	1.46	.00	3 Some Familiarity	-2.72	.60	-1.03	4 Very Familiar	2.10	1.58	1.33	0	.0000
8 Speaker 9 Spanish-English Female 1	-.63	.11	.00	1 No Familiarity	-2.72	.60	-1.03	4 Very Familiar	2.10	.61	3.43	8	.0090
9 Speaker 10 Arabic-English Male 2	1.21	1.29	.07	3 Some Familiarity	-.68	.50	-1.01	4 Very Familiar	1.90	1.39	1.37	0	.0000
10 Speaker 11 Spanish-English Female 2	1.21	1.29	.07	3 Some Familiarity	-.68	.50	-1.01	4 Very Familiar	1.90	1.39	1.37	0	.0000
9 Speaker 10 Arabic-English Male 2	1.19	.10	.06	1 No Familiarity	-.68	.50	-1.01	4 Very Familiar	1.87	.51	3.64	8	.0066
2 Speaker 3 Arabic-English Female 1	-.91	.11	.04	1 No Familiarity	-2.72	.60	-.84	4 Very Familiar	1.82	.61	2.97	8	.0179
10 Speaker 11 Spanish-English Female 2	1.11	.10	.01	1 No Familiarity	-.68	.50	-1.01	4 Very Familiar	1.79	.51	3.49	8	.0082
12 Speaker 13 Dhivehi-English Male 2	1.08	.32	.18	2 Limited Familiarity	-.62	1.46	-.73	3 Some Familiarity	1.69	1.49	1.14	0	.0000

2	Speaker 3 Arabic-English Female 1	-1.12	.34	-.08	2	Limited Familiarity	-2.72	.60	-.84	4	Very Familiar	1.60	.69	2.32	13	.0374
9	Speaker 10 Arabic-English Male 2	.87	.32	-.12	2	Limited Familiarity	-.68	.50	-1.01	4	Very Familiar	1.55	.60	2.59	13	.0222
11	Speaker 12 Arabic-English Female 2	.76	.32	.09	2	Limited Familiarity	-.62	1.46	-.64	3	Some Familiarity	1.38	1.49	.93	0	.0000
12	Speaker 13 Dhivehi-English Male 2	.65	.10	-.06	1	No Familiarity	-.62	1.46	-.73	3	Some Familiarity	1.27	1.46	.87	0	.0000
11	Speaker 12 Arabic-English Female 2	.64	.10	.02	1	No Familiarity	-.62	1.46	-.64	3	Some Familiarity	1.26	1.46	.86	0	.0000
11	Speaker 12 Arabic-English Female 2	.76	.32	.09	2	Limited Familiarity	-.43	.50	-.60	4	Very Familiar	1.20	.59	2.02	14	.0634
4	Speaker 5 Spanish-English Male 1	-.63	1.46	.21	3	Some Familiarity	-1.76	.54	-.34	4	Very Familiar	1.13	1.55	.73	0	.0000
11	Speaker 12 Arabic-English Female 2	.64	.10	.02	1	No Familiarity	-.43	.50	-.60	4	Very Familiar	1.07	.51	2.11	8	.0682
6	Speaker 7 Spanish-English Male 2	-.38	.11	.03	1	No Familiarity	-1.20	.52	-.42	4	Very Familiar	.82	.53	1.56	8	.1577
8	Speaker 9 Spanish-English Female 1	.15	.32	.44	2	Limited Familiarity	-.62	1.46	.00	3	Some Familiarity	.78	1.49	.52	0	.0000
4	Speaker 5 Spanish-English Male 1	-1.05	.11	.03	1	No Familiarity	-1.76	.54	-.34	4	Very Familiar	.71	.55	1.29	8	.2336
6	Speaker 7 Spanish-English Male 2	-.62	1.46	-.09	3	Some Familiarity	-1.20	.52	-.42	4	Very Familiar	.58	1.55	.37	0	.0000
7	Speaker 8 Dhivehi-English Male 1	-.05	.32	.00	2	Limited Familiarity	-.62	1.46	-.28	3	Some Familiarity	.57	1.49	.38	0	.0000
6	Speaker 7 Spanish-English Male 2	-.68	.33	-.13	2	Limited Familiarity	-1.20	.52	-.42	4	Very Familiar	.52	.61	.86	13	.4079
7	Speaker 8 Dhivehi-English Male 1	-.15	.10	-.05	1	No Familiarity	-.62	1.46	-.28	3	Some Familiarity	.47	1.46	.32	0	.0000
5	Speaker 6 Arabic-English Male 1	1.21	1.29	.42	3	Some Familiarity	.81	.50	.18	4	Very Familiar	.41	1.39	.29	0	.0000
4	Speaker 5 Spanish-English Male 1	-1.35	.35	-.13	2	Limited Familiarity	-1.76	.54	-.34	4	Very Familiar	.40	.64	.63	13	.5384
10	Speaker 11 Spanish-English Female 2	1.62	.34	.29	2	Limited Familiarity	1.21	1.29	.07	3	Some Familiarity	.40	1.34	.30	0	.0000
3	Speaker 4 Dhivehi-English Female 2	-.35	.11	-.03	1	No Familiarity	-.68	.33	-.21	2	Limited Familiarity	.33	.34	.96	27	.3431
9	Speaker 10 Arabic-English Male 2	1.19	.10	.06	1	No Familiarity	.87	.32	-.12	2	Limited Familiarity	.32	.34	.95	27	.3512
4	Speaker 5 Spanish-English Male 1	-1.05	.11	.03	1	No Familiarity	-1.35	.35	-.13	2	Limited Familiarity	.30	.36	.84	27	.4079
6	Speaker 7 Spanish-English Male 2	-.38	.11	.03	1	No Familiarity	-.68	.33	-.13	2	Limited Familiarity	.30	.34	.87	27	.3929
3	Speaker 4 Dhivehi-English Female 2	-.35	.11	-.03	1	No Familiarity	-.62	1.46	-.16	3	Some Familiarity	.28	1.46	.19	0	.0000
1	Speaker 2 Dhivehi-English Female 1	-.64	.11	-.05	1	No Familiarity	-.90	.33	-.19	2	Limited Familiarity	.26	.35	.74	27	.4664
5	Speaker 6 Arabic-English Male 1	.50	.10	.00	1	No Familiarity	.26	.32	-.14	2	Limited Familiarity	.25	.33	.74	27	.4628
6	Speaker 7 Spanish-English Male 2	-.38	.11	.03	1	No Familiarity	-.62	1.46	-.09	3	Some Familiarity	.24	1.46	.17	0	.0000
2	Speaker 3 Arabic-English Female 1	-.91	.11	.04	1	No Familiarity	-1.12	.34	-.08	2	Limited Familiarity	.22	.36	.61	27	.5499
8	Speaker 9 Spanish-English Female 1	-.63	.11	.00	1	No Familiarity	-.62	1.46	.00	3	Some Familiarity	.00	1.46	.00	0	.0000
1	Speaker 2 Dhivehi-English Female 1	-.64	.11	-.05	1	No Familiarity	-.62	1.46	-.04	3	Some Familiarity	-.01	1.46	-.01	0	.0000
9	Speaker 10 Arabic-English Male 2	1.19	.10	.06	1	No Familiarity	1.21	1.29	.07	3	Some Familiarity	-.02	1.30	-.02	0	.0000
6	Speaker 7 Spanish-English Male 2	-.68	.33	-.13	2	Limited Familiarity	-.62	1.46	-.09	3	Some Familiarity	-.05	1.49	-.04	0	.0000
3	Speaker 4 Dhivehi-English Female 2	-.68	.33	-.21	2	Limited Familiarity	-.62	1.46	-.16	3	Some Familiarity	-.06	1.49	-.04	0	.0000
7	Speaker 8 Dhivehi-English Male 1	-.15	.10	-.05	1	No Familiarity	-.05	.32	.00	2	Limited Familiarity	-.10	.34	-.29	27	.7743
10	Speaker 11 Spanish-English Female 2	1.11	.10	.01	1	No Familiarity	1.21	1.29	.07	3	Some Familiarity	-.10	1.30	-.08	0	.0000
11	Speaker 12 Arabic-English Female 2	.64	.10	.02	1	No Familiarity	.76	.32	.09	2	Limited Familiarity	-.12	.34	-.36	27	.7192
11	Speaker 12 Arabic-English Female 2	-.62	1.46	-.64	3	Some Familiarity	-.43	.50	-.60	4	Very Familiar	-.19	1.54	-.12	0	.0000
1	Speaker 2 Dhivehi-English Female 1	-.90	.33	-.19	2	Limited Familiarity	-.62	1.46	-.04	3	Some Familiarity	-.27	1.50	-.18	0	.0000
5	Speaker 6 Arabic-English Male 1	.50	.10	.00	1	No Familiarity	.81	.50	.18	4	Very Familiar	-.30	.51	-.59	8	.5741
9	Speaker 10 Arabic-English Male 2	.87	.32	-.12	2	Limited Familiarity	1.21	1.29	.07	3	Some Familiarity	-.35	1.33	-.26	0	.0000
4	Speaker 5 Spanish-English Male 1	-1.05	.11	.03	1	No Familiarity	-.63	1.46	.21	3	Some Familiarity	-.42	1.46	-.29	0	.0000
12	Speaker 13 Dhivehi-English Male 2	.65	.10	-.06	1	No Familiarity	1.08	.32	.18	2	Limited Familiarity	-.42	.34	-1.24	27	.2242
10	Speaker 11 Spanish-English Female 2	1.11	.10	.01	1	No Familiarity	1.62	.34	.29	2	Limited Familiarity	-.50	.35	-1.44	27	.1618
5	Speaker 6 Arabic-English Male 1	.26	.32	-.14	2	Limited Familiarity	.81	.50	.18	4	Very Familiar	-.55	.60	-.92	13	.3725
5	Speaker 6 Arabic-English Male 1	.50	.10	.00	1	No Familiarity	1.21	1.29	.42	3	Some Familiarity	-.71	1.30	-.55	0	.0000
4	Speaker 5 Spanish-English Male 1	-1.35	.35	-.13	2	Limited Familiarity	-.63	1.46	.21	3	Some Familiarity	-.73	1.50	-.49	0	.0000
8	Speaker 9 Spanish-English Female 1	-.63	.11	.00	1	No Familiarity	.15	.32	.44	2	Limited Familiarity	-.78	.34	-2.32	28	.0279
5	Speaker 6 Arabic-English Male 1	.26	.32	-.14	2	Limited Familiarity	1.21	1.29	.42	3	Some Familiarity	-.96	1.33	-.72	0	.0000
12	Speaker 13 Dhivehi-English Male 2	1.08	.32	.18	2	Limited Familiarity	3.02	.68	1.03	4	Very Familiar	-1.95	.76	-2.57	12	.0245
2	Speaker 3 Arabic-English Female 1	-.91	.11	.04	1	No Familiarity	1.21	1.29	1.16	3	Some Familiarity	-2.12	1.30	-1.63	0	.0000
3	Speaker 4 Dhivehi-English Female 2	-.35	.11	-.03	1	No Familiarity	1.91	.56	1.21	4	Very Familiar	-2.26	.57	-3.98	8	.0041
7	Speaker 8 Dhivehi-English Male 1	-.05	.32	.00	2	Limited Familiarity	2.24	.58	1.21	4	Very Familiar	-2.29	.67	-3.43	12	.0050
2	Speaker 3 Arabic-English Female 1	-1.12	.34	-.08	2	Limited Familiarity	1.21	1.29	1.16	3	Some Familiarity	-2.33	1.34	-1.74	0	.0000
12	Speaker 13 Dhivehi-English Male 2	.65	.10	-.06	1	No Familiarity	3.02	.68	1.03	4	Very Familiar	-2.37	.69	-3.43	7	.0110
7	Speaker 8 Dhivehi-English Male 1	-.15	.10	-.05	1	No Familiarity	2.24	.58	1.21	4	Very Familiar	-2.38	.59	-4.02	8	.0039
3	Speaker 4 Dhivehi-English Female 2	-.62	1.46	-.16	3	Some Familiarity	1.91	.56	1.21	4	Very Familiar	-2.53	1.56	-1.62	0	.0000
3	Speaker 4 Dhivehi-English Female 2	-.68	.33	-.21	2	Limited Familiarity	1.91	.56	1.21	4	Very Familiar	-2.59	.65	-4.00	13	.0015
7	Speaker 8 Dhivehi-English Male 1	-.62	1.46	-.28	3	Some Familiarity	2.24	.58	1.21	4	Very Familiar	-2.86	1.57	-1.82	0	.0000
1	Speaker 2 Dhivehi-English Female 1	-.62	1.46	-.04	3	Some Familiarity	2.60	.62	1.64	4	Very Familiar	-3.22	1.59	-2.03	0	.0000
1	Speaker 2 Dhivehi-English Female 1	-.64	.11	-.05	1	No Familiarity	2.60	.62	1.64	4	Very Familiar	-3.24	.63	-5.12	8	.0009
1	Speaker 2 Dhivehi-English Female 1	-.90	.33	-.19	2	Limited Familiarity	2.60	.62	1.64	4	Very Familiar	-3.49	.71	-4.95	12	.0003
12	Speaker 13 Dhivehi-English Male 2	-.62	1.46	-.73	3	Some Familiarity	3.02	.68	1.03	4	Very Familiar	-3.64	1.61	-2.26	0	.0000

Rasch Analysis of Expert Ratings 08/01/2015 15:30:27
Table 14.3.1.3 Bias/Interaction Pairwise Report (arranged by mN).

Bias/Interaction: 2. Dhivehi English Familiarity Level, 3. Speaker

Target N Dhivehi English Fam	Target Measr	S.E.	Obs-Exp Average	Context Nu Speaker	Target Measr	S.E.	Obs-Exp Average	Context Nu Speaker	Target Contrast	Joint S.E.	t	Welch d.f.	Prob.

4 Very Familiar	3.43	.62	1.64	1 Speaker 2 Dhivehi-English Female 1	-1.81	.60	-1.03	8 Speaker 9 Spanish-English Female 1	5.24	.87	6.05	11	.0001
4 Very Familiar	3.43	.62	1.64	1 Speaker 2 Dhivehi-English Female 1	-1.48	.50	-1.01	9 Speaker 10 Arabic-English Male 2	4.91	.80	6.14	11	.0001
4 Very Familiar	3.43	.62	1.64	1 Speaker 2 Dhivehi-English Female 1	-1.48	.50	-1.01	10 Speaker 11 Spanish-English Female 2	4.91	.80	6.14	11	.0001
4 Very Familiar	3.43	.62	1.64	1 Speaker 2 Dhivehi-English Female 1	-1.46	.60	-.84	2 Speaker 3 Arabic-English Female 1	4.89	.87	5.64	11	.0002
4 Very Familiar	2.58	.58	1.21	7 Speaker 8 Dhivehi-English Male 1	-1.81	.60	-1.03	8 Speaker 9 Spanish-English Female 1	4.39	.84	5.24	11	.0003
4 Very Familiar	2.49	.56	1.21	3 Speaker 4 Dhivehi-English Female 2	-1.81	.60	-1.03	8 Speaker 9 Spanish-English Female 1	4.30	.82	5.25	11	.0003
4 Very Familiar	3.43	.62	1.64	1 Speaker 2 Dhivehi-English Female 1	-.75	.50	-.60	11 Speaker 12 Arabic-English Female 2	4.18	.80	5.24	11	.0003
4 Very Familiar	2.58	.58	1.21	7 Speaker 8 Dhivehi-English Male 1	-1.48	.50	-1.01	9 Speaker 10 Arabic-English Male 2	4.06	.77	5.27	11	.0003
4 Very Familiar	2.58	.58	1.21	7 Speaker 8 Dhivehi-English Male 1	-1.48	.50	-1.01	10 Speaker 11 Spanish-English Female 2	4.06	.77	5.27	11	.0003
4 Very Familiar	2.49	.56	1.21	3 Speaker 4 Dhivehi-English Female 2	-1.48	.50	-1.01	9 Speaker 10 Arabic-English Male 2	3.98	.75	5.30	11	.0003
4 Very Familiar	2.49	.56	1.21	3 Speaker 4 Dhivehi-English Female 2	-1.48	.50	-1.01	10 Speaker 11 Spanish-English Female 2	3.98	.75	5.30	11	.0003
4 Very Familiar	3.43	.62	1.64	1 Speaker 2 Dhivehi-English Female 1	-.48	.52	-.42	6 Speaker 7 Spanish-English Male 2	3.91	.81	4.82	11	.0005
4 Very Familiar	3.43	.62	1.64	1 Speaker 2 Dhivehi-English Female 1	-.37	.54	-.34	4 Speaker 5 Spanish-English Male 1	3.80	.82	4.61	11	.0007
4 Very Familiar	2.58	.58	1.21	7 Speaker 8 Dhivehi-English Male 1	-.75	.50	-.60	11 Speaker 12 Arabic-English Female 2	3.33	.77	4.34	11	.0012
4 Very Familiar	2.49	.56	1.21	3 Speaker 4 Dhivehi-English Female 2	-.75	.50	-.60	11 Speaker 12 Arabic-English Female 2	3.24	.75	4.34	11	.0012
4 Very Familiar	2.49	.56	1.21	3 Speaker 4 Dhivehi-English Female 2	-.48	.52	-.42	6 Speaker 7 Spanish-English Male 2	2.97	.76	3.90	11	.0025
4 Very Familiar	2.49	.56	1.21	3 Speaker 4 Dhivehi-English Female 2	-.37	.54	-.34	4 Speaker 5 Spanish-English Male 1	2.86	.78	3.69	11	.0036
4 Very Familiar	3.43	.62	1.64	1 Speaker 2 Dhivehi-English Female 1	.60	.50	.18	5 Speaker 6 Arabic-English Male 1	2.83	.80	3.54	11	.0047
4 Very Familiar	.60	.50	.18	5 Speaker 6 Arabic-English Male 1	-1.81	.60	-1.03	8 Speaker 9 Spanish-English Female 1	2.41	.79	3.07	11	.0107
4 Very Familiar	.60	.50	.18	5 Speaker 6 Arabic-English Male 1	-1.48	.50	-1.01	9 Speaker 10 Arabic-English Male 2	2.08	.71	2.92	11	.0139
4 Very Familiar	.60	.50	.18	5 Speaker 6 Arabic-English Male 1	-1.48	.50	-1.01	10 Speaker 11 Spanish-English Female 2	2.08	.71	2.92	11	.0139
4 Very Familiar	2.49	.56	1.21	3 Speaker 4 Dhivehi-English Female 2	.60	.50	.18	5 Speaker 6 Arabic-English Male 1	1.90	.75	2.52	11	.0283
4 Very Familiar	-.37	.54	-.34	4 Speaker 5 Spanish-English Male 1	-1.81	.60	-1.03	8 Speaker 9 Spanish-English Female 1	1.44	.81	1.78	11	.1021
4 Very Familiar	.60	.50	.18	5 Speaker 6 Arabic-English Male 1	-.75	.50	-.60	11 Speaker 12 Arabic-English Female 2	1.35	.71	1.90	11	.0840
4 Very Familiar	-.48	.52	-.42	6 Speaker 7 Spanish-English Male 2	-1.81	.60	-1.03	8 Speaker 9 Spanish-English Female 1	1.33	.79	1.68	11	.1210
4 Very Familiar	-.37	.54	-.34	4 Speaker 5 Spanish-English Male 1	-1.48	.50	-1.01	9 Speaker 10 Arabic-English Male 2	1.11	.74	1.51	11	.1593
4 Very Familiar	-.37	.54	-.34	4 Speaker 5 Spanish-English Male 1	-1.48	.50	-1.01	10 Speaker 11 Spanish-English Female 2	1.11	.74	1.51	11	.1593
4 Very Familiar	.60	.50	.18	5 Speaker 6 Arabic-English Male 1	-.48	.52	-.42	6 Speaker 7 Spanish-English Male 2	1.07	.72	1.49	11	.1655
4 Very Familiar	-.48	.52	-.42	6 Speaker 7 Spanish-English Male 2	-1.48	.50	-1.01	9 Speaker 10 Arabic-English Male 2	1.01	.72	1.39	11	.1908
4 Very Familiar	-.48	.52	-.42	6 Speaker 7 Spanish-English Male 2	-1.48	.50	-1.01	10 Speaker 11 Spanish-English Female 2	1.01	.72	1.39	11	.1908
2 Limited Familiarity	.89	.32	.44	8 Speaker 9 Spanish-English Female 1	-.11	.32	-.12	9 Speaker 10 Arabic-English Male 2	1.00	.45	2.20	31	.0350
4 Very Familiar	3.43	.62	1.64	1 Speaker 2 Dhivehi-English Female 1	2.49	.56	1.21	3 Speaker 4 Dhivehi-English Female 2	.94	.84	1.12	11	.2854
4 Very Familiar	3.43	.62	1.64	1 Speaker 2 Dhivehi-English Female 1	2.55	.68	1.03	12 Speaker 13 Dhivehi-English Male 2	.88	.92	.95	11	.3638
4 Very Familiar	3.43	.62	1.64	1 Speaker 2 Dhivehi-English Female 1	2.58	.58	1.21	7 Speaker 8 Dhivehi-English Male 1	.85	.85	1.00	11	.3410
2 Limited Familiarity	.89	.32	.44	8 Speaker 9 Spanish-English Female 1	.27	.32	.09	11 Speaker 12 Arabic-English Female 2	.62	.45	1.37	31	.1805
2 Limited Familiarity	.89	.32	.44	8 Speaker 9 Spanish-English Female 1	.43	.32	.18	12 Speaker 13 Dhivehi-English Male 2	.46	.45	1.00	31	.3242
4 Very Familiar	-.37	.54	-.34	4 Speaker 5 Spanish-English Male 1	-.75	.50	-.60	11 Speaker 12 Arabic-English Female 2	.38	.73	.52	11	.6143
2 Limited Familiarity	.64	.34	.29	10 Speaker 11 Spanish-English Female 2	.27	.32	.09	11 Speaker 12 Arabic-English Female 2	.37	.46	.80	31	.4303
4 Very Familiar	-1.46	.60	-.84	2 Speaker 3 Arabic-English Female 1	-1.81	.60	-1.03	8 Speaker 9 Spanish-English Female 1	.35	.85	.42	11	.6861
4 Very Familiar	-.48	.52	-.42	6 Speaker 7 Spanish-English Male 2	-.75	.50	-.60	11 Speaker 12 Arabic-English Female 2	.27	.72	.38	11	.7106
2 Limited Familiarity	.89	.32	.44	8 Speaker 9 Spanish-English Female 1	.64	.34	.29	10 Speaker 11 Spanish-English Female 2	.25	.46	.54	31	.5945
2 Limited Familiarity	-.03	.34	-.08	2 Speaker 3 Arabic-English Female 1	-.27	.33	-.21	3 Speaker 4 Dhivehi-English Female 2	.24	.47	.51	31	.6146
2 Limited Familiarity	.12	.32	.00	7 Speaker 8 Dhivehi-English Male 1	-.11	.32	-.12	9 Speaker 10 Arabic-English Male 2	.23	.45	.50	31	.6175
2 Limited Familiarity	.64	.34	.29	10 Speaker 11 Spanish-English Female 2	.43	.32	.18	12 Speaker 13 Dhivehi-English Male 2	.21	.47	.44	31	.6606
1 No Familiarity	.12	.10	.06	9 Speaker 10 Arabic-English Male 2	-.08	.10	-.06	12 Speaker 13 Dhivehi-English Male 2	.20	.15	1.37	327	.1716
1 No Familiarity	.09	.11	.04	2 Speaker 3 Arabic-English Female 1	-.08	.10	-.06	12 Speaker 13 Dhivehi-English Male 2	.17	.15	1.15	327	.2512
1 No Familiarity	.09	.11	.04	2 Speaker 3 Arabic-English Female 1	-.07	.10	-.05	7 Speaker 8 Dhivehi-English Male 1	.16	.15	1.07	327	.2871
1 No Familiarity	.08	.11	.03	6 Speaker 7 Spanish-English Male 2	-.08	.10	-.06	12 Speaker 13 Dhivehi-English Male 2	.16	.15	1.09	327	.2767
1 No Familiarity	.07	.11	.03	4 Speaker 5 Spanish-English Male 1	-.08	.10	-.06	12 Speaker 13 Dhivehi-English Male 2	.15	.15	1.02	327	.3085
1 No Familiarity	.08	.11	.03	6 Speaker 7 Spanish-English Male 2	-.07	.10	-.05	7 Speaker 8 Dhivehi-English Male 1	.15	.15	1.01	327	.3156
1 No Familiarity	.07	.11	.03	4 Speaker 5 Spanish-English Male 1	-.07	.10	-.05	7 Speaker 8 Dhivehi-English Male 1	.14	.15	.94	327	.3490
1 No Familiarity	.05	.10	.02	11 Speaker 12 Arabic-English Female 2	-.08	.10	-.06	12 Speaker 13 Dhivehi-English Male 2	.14	.15	.94	328	.3484
1 No Familiarity	.04	.10	.01	10 Speaker 11 Spanish-English Female 2	-.08	.10	-.06	12 Speaker 13 Dhivehi-English Male 2	.12	.15	.85	327	.3940
1 No Familiarity	.09	.11	.04	2 Speaker 3 Arabic-English Female 1	-.03	.11	-.03	3 Speaker 4 Dhivehi-English Female 2	.12	.15	.81	327	.4195
1 No Familiarity	.02	.10	.00	5 Speaker 6 Arabic-English Male 1	-.08	.10	-.06	12 Speaker 13 Dhivehi-English Male 2	.11	.15	.75	327	.4536
2 Limited Familiarity	-.03	.34	-.08	2 Speaker 3 Arabic-English Female 1	-.14	.35	-.13	4 Speaker 5 Spanish-English Male 1	.11	.48	.22	31	.8247
4 Very Familiar	-.37	.54	-.34	4 Speaker 5 Spanish-English Male 1	-.48	.52	-.42	6 Speaker 7 Spanish-English Male 2	.11	.75	.14	11	.8885
1 No Familiarity	.01	.11	.00	8 Speaker 9 Spanish-English Female 1	-.08	.10	-.06	12 Speaker 13 Dhivehi-English Male 2	.10	.15	.66	327	.5085
1 No Familiarity	.02	.10	.00	5 Speaker 6 Arabic-English Male 1	-.07	.10	-.05	7 Speaker 8 Dhivehi-English Male 1	.10	.15	.67	327	.5053
2 Limited Familiarity	-.03	.34	-.08	2 Speaker 3 Arabic-English Female 1	-.13	.32	-.14	5 Speaker 6 Arabic-English Male 1	.10	.46	.21	31	.8355
2 Limited Familiarity	-.03	.34	-.08	2 Speaker 3 Arabic-English Female 1	-.13	.33	-.13	6 Speaker 7 Spanish-English Male 2	.10	.47	.20	31	.8403
2 Limited Familiarity	-.03	.34	-.08	2 Speaker 3 Arabic-English Female 1	-.11	.32	-.12	9 Speaker 10 Arabic-English Male 2	.08	.47	.16	31	.8701
1 No Familiarity	.12	.10	.06	9 Speaker 10 Arabic-English Male 2	.04	.10	.01	10 Speaker 11 Spanish-English Female 2	.08	.15	.52	327	.6066
1 No Familiarity	.09	.11	.04	2 Speaker 3 Arabic-English Female 1	.01	.11	.00	8 Speaker 9 Spanish-English Female 1	.07	.15	.49	327	.6277
1 No Familiarity	.12	.10	.06	9 Speaker 10 Arabic-English Male 2	.05	.10	.02	11 Speaker 12 Arabic-English Female 2	.06	.15	.44	327	.6623
1 No Familiarity	.09	.11	.04	2 Speaker 3 Arabic-English Female 1	.02	.10	.00	5 Speaker 6 Arabic-English Male 1	.06	.15	.42	327	.6741
1 No Familiarity	.08	.11	.03	6 Speaker 7 Spanish-English Male 2	.01	.11	.00	8 Speaker 9 Spanish-English Female 1	.06	.15	.42	327	.6783
1 No Familiarity	.07	.11	.03	4 Speaker 5 Spanish-English Male 1	.01	.11	.00	8 Speaker 9 Spanish-English Female 1	.06	.15	.36	327	.7176
1 No Familiarity	-.03	.11	-.03	3 Speaker 4 Dhivehi-English Female 2	-.08	.10	-.06	12 Speaker 13 Dhivehi-English Male 2	.05	.15	.34	327	.7349
1 No Familiarity	.09	.11	.04	2 Speaker 3 Arabic-English Female 1	.04	.10	.01	10 Speaker 11 Spanish-English Female 2	.05	.15	.32	327	.7529
1 No Familiarity	.07	.11	.03	4 Speaker 5 Spanish-English Male 1	.02	.10	.00	5 Speaker 6 Arabic-English Male 1	.04	.15	.30	327	.7678
1 No Familiarity	-.03	.11	-.03	3 Speaker 4 Dhivehi-English Female 2	-.07	.10	-.05	7 Speaker 8 Dhivehi-English Male 1	.04	.15	.26	327	.7957

1 No Familiarity	.08	.11	.03	6 Speaker 7 Spanish-English Male 2	.04	.10	.01	10 Speaker 11 Spanish-English Female 2	.04	.15	.24	327	.8098
1 No Familiarity	.09	.11	.04	2 Speaker 3 Arabic-English Female 1	.05	.10	.02	11 Speaker 12 Arabic-English Female 2	.04	.15	.24	327	.8124
2 Limited Familiarity	-.24	.33	-.19	1 Speaker 2 Dhivehi-English Female 1	-.27	.33	-.21	3 Speaker 4 Dhivehi-English Female 2	.03	.47	.07	31	.9448
1 No Familiarity	.07	.11	.03	4 Speaker 5 Spanish-English Male 1	.04	.10	.01	10 Speaker 11 Spanish-English Female 2	.03	.15	.19	327	.8487
4 Very Familiar	2.58	.58	1.21	7 Speaker 8 Dhivehi-English Male 1	2.55	.68	1.03	12 Speaker 13 Dhivehi-English Male 2	.03	.90	.03	11	.9770
4 Very Familiar	-1.46	.60	-.84	2 Speaker 3 Arabic-English Female 1	-1.48	.50	-1.01	9 Speaker 10 Arabic-English Male 2	.03	.79	.03	11	.9745
4 Very Familiar	-1.46	.60	-.84	2 Speaker 3 Arabic-English Female 1	-1.48	.50	-1.01	10 Speaker 11 Spanish-English Female 2	.03	.79	.03	11	.9745
1 No Familiarity	.08	.11	.03	6 Speaker 7 Spanish-English Male 2	.05	.10	.02	11 Speaker 12 Arabic-English Female 2	.02	.15	.16	327	.8716
1 No Familiarity	.09	.11	.04	2 Speaker 3 Arabic-English Female 1	.07	.11	.03	4 Speaker 5 Spanish-English Male 1	.02	.15	.12	327	.9049
1 No Familiarity	.07	.11	.03	4 Speaker 5 Spanish-English Male 1	.05	.10	.02	11 Speaker 12 Arabic-English Female 2	.02	.15	.11	327	.9099
1 No Familiarity	.09	.11	.04	2 Speaker 3 Arabic-English Female 1	.08	.11	.03	6 Speaker 7 Spanish-English Male 2	.01	.15	.08	327	.9381
1 No Familiarity	-.07	.10	-.05	7 Speaker 8 Dhivehi-English Male 1	-.08	.10	-.06	12 Speaker 13 Dhivehi-English Male 2	.01	.15	.08	327	.9375
1 No Familiarity	.02	.10	.00	5 Speaker 6 Arabic-English Male 1	.01	.11	.00	8 Speaker 9 Spanish-English Female 1	.01	.15	.07	327	.9409
1 No Familiarity	-.08	.11	-.05	1 Speaker 2 Dhivehi-English Female 1	-.08	.10	-.06	12 Speaker 13 Dhivehi-English Male 2	.01	.15	.06	327	.9559
4 Very Familiar	-1.48	.50	-1.01	9 Speaker 10 Arabic-English Male 2	-1.48	.50	-1.01	10 Speaker 11 Spanish-English Female 2	.00	.71	.00	12	1.000
2 Limited Familiarity	-.13	.32	-.14	5 Speaker 6 Arabic-English Male 1	-.13	.33	-.13	6 Speaker 7 Spanish-English Male 2	.00	.46	.00	31	.9973
1 No Familiarity	-.08	.11	-.05	1 Speaker 2 Dhivehi-English Female 1	-.07	.10	-.05	7 Speaker 8 Dhivehi-English Male 1	.00	.15	-.02	327	.9825
1 No Familiarity	.07	.11	.03	4 Speaker 5 Spanish-English Male 1	.08	.11	.03	6 Speaker 7 Spanish-English Male 2	-.01	.15	-.04	327	.9647
2 Limited Familiarity	-.14	.35	-.13	4 Speaker 5 Spanish-English Male 1	-.13	.32	-.14	5 Speaker 6 Arabic-English Male 1	-.01	.47	-.02	31	.9821
1 No Familiarity	.04	.10	.01	10 Speaker 11 Spanish-English Female 2	.05	.10	.02	11 Speaker 12 Arabic-English Female 2	-.01	.15	-.08	327	.9356
2 Limited Familiarity	-.14	.35	-.13	4 Speaker 5 Spanish-English Male 1	-.13	.33	-.13	6 Speaker 7 Spanish-English Male 2	-.01	.48	-.03	31	.9797
1 No Familiarity	.02	.10	.00	5 Speaker 6 Arabic-English Male 1	.04	.10	.01	10 Speaker 11 Spanish-English Female 2	-.02	.15	-.11	327	.9149
2 Limited Familiarity	-.13	.33	-.13	6 Speaker 7 Spanish-English Male 2	-.11	.32	-.12	9 Speaker 10 Arabic-English Male 2	-.02	.46	-.04	31	.9676
2 Limited Familiarity	-.13	.32	-.14	5 Speaker 6 Arabic-English Male 1	-.11	.32	-.12	9 Speaker 10 Arabic-English Male 2	-.02	.45	-.05	31	.9644
1 No Familiarity	.01	.11	.00	8 Speaker 9 Spanish-English Female 1	.04	.10	.01	10 Speaker 11 Spanish-English Female 2	-.03	.15	-.18	327	.8582
1 No Familiarity	.02	.10	.00	5 Speaker 6 Arabic-English Male 1	.05	.10	.02	11 Speaker 12 Arabic-English Female 2	-.03	.15	-.19	328	.8504
1 No Familiarity	.09	.11	.04	2 Speaker 3 Arabic-English Female 1	.12	.10	.06	9 Speaker 10 Arabic-English Male 2	-.03	.15	-.19	327	.8506
2 Limited Familiarity	-.14	.35	-.13	4 Speaker 5 Spanish-English Male 1	-.11	.32	-.12	9 Speaker 10 Arabic-English Male 2	-.03	.47	-.07	31	.9480
1 No Familiarity	.01	.11	.00	8 Speaker 9 Spanish-English Female 1	.05	.10	.02	11 Speaker 12 Arabic-English Female 2	-.04	.15	-.26	327	.7955
1 No Familiarity	.08	.11	.03	6 Speaker 7 Spanish-English Male 2	.12	.10	.06	9 Speaker 10 Arabic-English Male 2	-.04	.15	-.27	327	.7864
1 No Familiarity	-.08	.11	-.05	1 Speaker 2 Dhivehi-English Female 1	-.03	.11	-.03	3 Speaker 4 Dhivehi-English Female 2	-.04	.15	-.28	327	.7814
1 No Familiarity	.07	.11	.03	4 Speaker 5 Spanish-English Male 1	.12	.10	.06	9 Speaker 10 Arabic-English Male 2	-.05	.15	-.31	327	.7571
1 No Familiarity	-.03	.11	-.03	3 Speaker 4 Dhivehi-English Female 2	.01	.11	.00	8 Speaker 9 Spanish-English Female 1	-.05	.15	-.32	327	.7475
1 No Familiarity	.02	.10	.00	5 Speaker 6 Arabic-English Male 1	.08	.11	.03	6 Speaker 7 Spanish-English Male 2	-.05	.15	-.35	327	.7279
1 No Familiarity	-.03	.11	-.03	3 Speaker 4 Dhivehi-English Female 2	.02	.10	.00	5 Speaker 6 Arabic-English Male 1	-.06	.15	-.40	327	.6871
4 Very Familiar	2.49	.56	1.21	3 Speaker 4 Dhivehi-English Female 2	2.55	.68	1.03	12 Speaker 13 Dhivehi-English Male 2	-.06	.88	-.07	11	.9451
1 No Familiarity	-.03	.11	-.03	3 Speaker 4 Dhivehi-English Female 2	.04	.10	.01	10 Speaker 11 Spanish-English Female 2	-.07	.15	-.51	327	.6126
1 No Familiarity	-.07	.10	-.05	7 Speaker 8 Dhivehi-English Male 1	.01	.11	.00	8 Speaker 9 Spanish-English Female 1	-.09	.15	-.58	327	.5619
1 No Familiarity	-.03	.11	-.03	3 Speaker 4 Dhivehi-English Female 2	.05	.10	.02	11 Speaker 12 Arabic-English Female 2	-.09	.15	-.59	327	.5558
4 Very Familiar	2.49	.56	1.21	3 Speaker 4 Dhivehi-English Female 2	2.58	.58	1.21	7 Speaker 8 Dhivehi-English Male 1	-.09	.81	-.11	11	.9146
1 No Familiarity	-.08	.11	-.05	1 Speaker 2 Dhivehi-English Female 1	.01	.11	.00	8 Speaker 9 Spanish-English Female 1	-.09	.15	-.60	327	.5521
1 No Familiarity	.02	.10	.00	5 Speaker 6 Arabic-English Male 1	.12	.10	.06	9 Speaker 10 Arabic-English Male 2	-.09	.15	-.62	327	.5326
2 Limited Familiarity	-.24	.33	-.19	1 Speaker 2 Dhivehi-English Female 1	-.14	.35	-.13	4 Speaker 5 Spanish-English Male 1	-.10	.48	-.21	31	.8376
1 No Familiarity	-.08	.11	-.05	1 Speaker 2 Dhivehi-English Female 1	.02	.10	.00	5 Speaker 6 Arabic-English Male 1	-.10	.15	-.68	327	.4966
1 No Familiarity	.01	.11	.00	8 Speaker 9 Spanish-English Female 1	.12	.10	.06	9 Speaker 10 Arabic-English Male 2	-.10	.15	-.69	327	.4928
1 No Familiarity	-.03	.11	-.03	3 Speaker 4 Dhivehi-English Female 2	.07	.11	.03	4 Speaker 5 Spanish-English Male 1	-.10	.15	-.68	327	.4958
2 Limited Familiarity	-.24	.33	-.19	1 Speaker 2 Dhivehi-English Female 1	-.13	.32	-.14	5 Speaker 6 Arabic-English Male 1	-.11	.46	-.24	31	.8133
1 No Familiarity	-.03	.11	-.03	3 Speaker 4 Dhivehi-English Female 2	.08	.11	.03	6 Speaker 7 Spanish-English Male 2	-.11	.15	-.74	327	.4581
2 Limited Familiarity	-.24	.33	-.19	1 Speaker 2 Dhivehi-English Female 1	-.13	.33	-.13	6 Speaker 7 Spanish-English Male 2	-.11	.47	-.24	31	.8133
1 No Familiarity	-.07	.10	-.05	7 Speaker 8 Dhivehi-English Male 1	.04	.10	.01	10 Speaker 11 Spanish-English Female 2	-.11	.15	-.77	327	.4419
1 No Familiarity	-.08	.11	-.05	1 Speaker 2 Dhivehi-English Female 1	.04	.10	.01	10 Speaker 11 Spanish-English Female 2	-.12	.15	-.78	327	.4347
1 No Familiarity	-.07	.10	-.05	7 Speaker 8 Dhivehi-English Male 1	.05	.10	.02	11 Speaker 12 Arabic-English Female 2	-.13	.15	-.85	327	.3936
1 No Familiarity	-.08	.11	-.05	1 Speaker 2 Dhivehi-English Female 1	.05	.10	.02	11 Speaker 12 Arabic-English Female 2	-.13	.15	-.87	327	.3874
2 Limited Familiarity	-.24	.33	-.19	1 Speaker 2 Dhivehi-English Female 1	-.11	.32	-.12	9 Speaker 10 Arabic-English Male 2	-.13	.46	-.28	31	.7804
2 Limited Familiarity	-.27	.33	-.21	3 Speaker 4 Dhivehi-English Female 2	-.14	.35	-.13	4 Speaker 5 Spanish-English Male 1	-.13	.48	-.28	31	.7839
2 Limited Familiarity	-.27	.33	-.21	3 Speaker 4 Dhivehi-English Female 2	-.13	.32	-.14	5 Speaker 6 Arabic-English Male 1	-.14	.46	-.31	31	.7577
2 Limited Familiarity	-.27	.33	-.21	3 Speaker 4 Dhivehi-English Female 2	-.13	.33	-.13	6 Speaker 7 Spanish-English Male 2	-.14	.46	-.31	32	.7584
1 No Familiarity	-.08	.11	-.05	1 Speaker 2 Dhivehi-English Female 1	.07	.11	.03	4 Speaker 5 Spanish-English Male 1	-.15	.15	-.95	327	.3438
2 Limited Familiarity	.12	.32	.00	7 Speaker 8 Dhivehi-English Male 1	.27	.32	.09	11 Speaker 12 Arabic-English Female 2	-.15	.45	-.33	31	.7426
1 No Familiarity	-.03	.11	-.03	3 Speaker 4 Dhivehi-English Female 2	.12	.10	.06	9 Speaker 10 Arabic-English Male 2	-.15	.15	-1.02	327	.3092
2 Limited Familiarity	-.03	.34	-.08	2 Speaker 3 Arabic-English Female 1	.12	.32	.00	7 Speaker 8 Dhivehi-English Male 1	-.15	.47	-.33	31	.7466
1 No Familiarity	-.08	.11	-.05	1 Speaker 2 Dhivehi-English Female 1	.08	.11	.03	6 Speaker 7 Spanish-English Male 2	-.15	.15	-1.01	327	.3112
2 Limited Familiarity	-.27	.33	-.21	3 Speaker 4 Dhivehi-English Female 2	-.11	.32	-.12	9 Speaker 10 Arabic-English Male 2	-.16	.46	-.35	31	.7255
2 Limited Familiarity	.27	.32	.09	11 Speaker 12 Arabic-English Female 2	.43	.32	.18	12 Speaker 13 Dhivehi-English Male 2	-.16	.46	-.36	31	.7215
1 No Familiarity	-.08	.11	-.05	1 Speaker 2 Dhivehi-English Female 1	.09	.11	.04	2 Speaker 3 Arabic-English Female 1	-.16	.15	-1.07	327	.2832
1 No Familiarity	-.07	.10	-.05	7 Speaker 8 Dhivehi-English Male 1	.12	.10	.06	9 Speaker 10 Arabic-English Male 2	-.19	.15	-1.28	327	.2003
1 No Familiarity	-.08	.11	-.05	1 Speaker 2 Dhivehi-English Female 1	.12	.10	.06	9 Speaker 10 Arabic-English Male 2	-.19	.15	-1.29	327	.1983
2 Limited Familiarity	-.24	.33	-.19	1 Speaker 2 Dhivehi-English Female 1	-.03	.34	-.08	2 Speaker 3 Arabic-English Female 1	-.21	.47	-.44	31	.6656
2 Limited Familiarity	-.13	.33	-.13	6 Speaker 7 Spanish-English Male 2	.12	.32	.00	7 Speaker 8 Dhivehi-English Male 1	-.25	.46	-.54	31	.5929
2 Limited Familiarity	-.13	.32	-.14	5 Speaker 6 Arabic-English Male 1	.12	.32	.00	7 Speaker 8 Dhivehi-English Male 1	-.25	.45	-.55	31	.5851
2 Limited Familiarity	-.14	.35	-.13	4 Speaker 5 Spanish-English Male 1	.12	.32	.00	7 Speaker 8 Dhivehi-English Male 1	-.26	.47	-.55	31	.5850
2 Limited Familiarity	-.03	.34	-.08	2 Speaker 3 Arabic-English Female 1	.27	.32	.09	11 Speaker 12 Arabic-English Female 2	-.30	.47	-.65	31	.5219

2 Limited Familiarity	.12	.32	.00	7 Speaker 8 Dhivehi-English Male 1	.43	.32	.18	12 Speaker 13 Dhivehi-English Male 2	-.31	.46	-.69	31	.4956
4 Very Familiar	-1.81	.60	-1.03	8 Speaker 9 Spanish-English Female 1	-1.48	.50	-1.01	9 Speaker 10 Arabic-English Male 2	-.33	.79	-.42	11	.6843
4 Very Familiar	-1.81	.60	-1.03	8 Speaker 9 Spanish-English Female 1	-1.48	.50	-1.01	10 Speaker 11 Spanish-English Female 2	-.33	.79	-.42	11	.6843
2 Limited Familiarity	-.24	.33	-.19	1 Speaker 2 Dhivehi-English Female 1	.12	.32	.00	7 Speaker 8 Dhivehi-English Male 1	-.36	.46	-.78	31	.4427
2 Limited Familiarity	-.11	.32	-.12	9 Speaker 10 Arabic-English Male 2	.27	.32	.09	11 Speaker 12 Arabic-English Female 2	-.38	.45	-.83	31	.4103
2 Limited Familiarity	-.27	.33	-.21	3 Speaker 4 Dhivehi-English Female 2	.12	.32	.00	7 Speaker 8 Dhivehi-English Male 1	-.39	.46	-.85	31	.3995
2 Limited Familiarity	-.13	.33	-.13	6 Speaker 7 Spanish-English Male 2	.27	.32	.09	11 Speaker 12 Arabic-English Female 2	-.40	.46	-.87	31	.3926
2 Limited Familiarity	-.13	.32	-.14	5 Speaker 6 Arabic-English Male 1	.27	.32	.09	11 Speaker 12 Arabic-English Female 2	-.40	.45	-.88	31	.3839
2 Limited Familiarity	-.14	.35	-.13	4 Speaker 5 Spanish-English Male 1	.27	.32	.09	11 Speaker 12 Arabic-English Female 2	-.41	.47	-.87	31	.3909
2 Limited Familiarity	-.03	.34	-.08	2 Speaker 3 Arabic-English Female 1	.43	.32	.18	12 Speaker 13 Dhivehi-English Male 2	-.47	.47	-.99	31	.3277
2 Limited Familiarity	-.24	.33	-.19	1 Speaker 2 Dhivehi-English Female 1	.27	.32	.09	11 Speaker 12 Arabic-English Female 2	-.51	.46	-1.10	31	.2790
2 Limited Familiarity	.12	.32	.00	7 Speaker 8 Dhivehi-English Male 1	.64	.34	.29	10 Speaker 11 Spanish-English Female 2	-.52	.46	-1.12	31	.2698
2 Limited Familiarity	-.27	.33	-.21	3 Speaker 4 Dhivehi-English Female 2	.27	.32	.09	11 Speaker 12 Arabic-English Female 2	-.54	.46	-1.18	31	.2467
2 Limited Familiarity	-.11	.32	-.12	9 Speaker 10 Arabic-English Male 2	.43	.32	.18	12 Speaker 13 Dhivehi-English Male 2	-.54	.46	-1.19	31	.2434
2 Limited Familiarity	-.13	.33	-.13	6 Speaker 7 Spanish-English Male 2	.43	.32	.18	12 Speaker 13 Dhivehi-English Male 2	-.56	.46	-1.22	31	.2325
2 Limited Familiarity	-.13	.32	-.14	5 Speaker 6 Arabic-English Male 1	.43	.32	.18	12 Speaker 13 Dhivehi-English Male 2	-.56	.45	-1.24	31	.2246
2 Limited Familiarity	-.14	.35	-.13	4 Speaker 5 Spanish-English Male 1	.43	.32	.18	12 Speaker 13 Dhivehi-English Male 2	-.57	.47	-1.21	31	.2347
2 Limited Familiarity	-.03	.34	-.08	2 Speaker 3 Arabic-English Female 1	.64	.34	.29	10 Speaker 11 Spanish-English Female 2	-.67	.48	-1.41	31	.1679
2 Limited Familiarity	-.24	.33	-.19	1 Speaker 2 Dhivehi-English Female 1	.43	.32	.18	12 Speaker 13 Dhivehi-English Male 2	-.67	.46	-1.45	31	.1574
2 Limited Familiarity	-.27	.33	-.21	3 Speaker 4 Dhivehi-English Female 2	.43	.32	.18	12 Speaker 13 Dhivehi-English Male 2	-.71	.46	-1.53	31	.1361
4 Very Familiar	-1.46	.60	-.84	2 Speaker 3 Arabic-English Female 1	-.75	.50	-.60	11 Speaker 12 Arabic-English Female 2	-.71	.78	-.90	11	.3852
4 Very Familiar	-1.48	.50	-1.01	9 Speaker 10 Arabic-English Male 2	-.75	.50	-.60	11 Speaker 12 Arabic-English Female 2	-.73	.71	-1.03	11	.3233
4 Very Familiar	-1.48	.50	-1.01	10 Speaker 11 Spanish-English Female 2	-.75	.50	-.60	11 Speaker 12 Arabic-English Female 2	-.73	.71	-1.03	11	.3233
2 Limited Familiarity	-.11	.32	-.12	9 Speaker 10 Arabic-English Male 2	.64	.34	.29	10 Speaker 11 Spanish-English Female 2	-.75	.46	-1.61	31	.1167
2 Limited Familiarity	-.13	.33	-.13	6 Speaker 7 Spanish-English Male 2	.64	.34	.29	10 Speaker 11 Spanish-English Female 2	-.77	.47	-1.64	31	.1115
2 Limited Familiarity	.12	.32	.00	7 Speaker 8 Dhivehi-English Male 1	.89	.32	.44	8 Speaker 9 Spanish-English Female 1	-.77	.45	-1.70	31	.0985
2 Limited Familiarity	-.13	.32	-.14	5 Speaker 6 Arabic-English Male 1	.64	.34	.29	10 Speaker 11 Spanish-English Female 2	-.77	.46	-1.66	31	.1061
2 Limited Familiarity	-.14	.35	-.13	4 Speaker 5 Spanish-English Male 1	.64	.34	.29	10 Speaker 11 Spanish-English Female 2	-.78	.48	-1.62	31	.1149
2 Limited Familiarity	-.24	.33	-.19	1 Speaker 2 Dhivehi-English Female 1	.64	.34	.29	10 Speaker 11 Spanish-English Female 2	-.88	.47	-1.86	31	.0720
2 Limited Familiarity	-.27	.33	-.21	3 Speaker 4 Dhivehi-English Female 2	.64	.34	.29	10 Speaker 11 Spanish-English Female 2	-.91	.47	-1.94	31	.0609
2 Limited Familiarity	-.03	.34	-.08	2 Speaker 3 Arabic-English Female 1	.89	.32	.44	8 Speaker 9 Spanish-English Female 1	-.92	.47	-1.98	31	.0565
4 Very Familiar	-.37	.54	-.34	4 Speaker 5 Spanish-English Male 1	.60	.50	.18	5 Speaker 6 Arabic-English Male 1	-.97	.74	-1.31	11	.2173
4 Very Familiar	-1.46	.60	-.84	2 Speaker 3 Arabic-English Female 1	-.48	.52	-.42	6 Speaker 7 Spanish-English Male 2	-.98	.79	-1.24	11	.2424
2 Limited Familiarity	-.13	.33	-.13	6 Speaker 7 Spanish-English Male 2	.89	.32	.44	8 Speaker 9 Spanish-English Female 1	-1.02	.46	-2.22	31	.0337
2 Limited Familiarity	-.13	.32	-.14	5 Speaker 6 Arabic-English Male 1	.89	.32	.44	8 Speaker 9 Spanish-English Female 1	-1.02	.45	-2.26	31	.0310
2 Limited Familiarity	-.14	.35	-.13	4 Speaker 5 Spanish-English Male 1	.89	.32	.44	8 Speaker 9 Spanish-English Female 1	-1.03	.47	-2.19	31	.0362
4 Very Familiar	-1.81	.60	-1.03	8 Speaker 9 Spanish-English Female 1	-.75	.50	-.60	11 Speaker 12 Arabic-English Female 2	-1.06	.78	-1.36	11	.2022
4 Very Familiar	-1.46	.60	-.84	2 Speaker 3 Arabic-English Female 1	-.37	.54	-.34	4 Speaker 5 Spanish-English Male 1	-1.09	.81	-1.35	11	.2054
2 Limited Familiarity	-.24	.33	-.19	1 Speaker 2 Dhivehi-English Female 1	.89	.32	.44	8 Speaker 9 Spanish-English Female 1	-1.13	.46	-2.45	31	.0202
2 Limited Familiarity	-.27	.33	-.21	3 Speaker 4 Dhivehi-English Female 2	.89	.32	.44	8 Speaker 9 Spanish-English Female 1	-1.16	.46	-2.54	31	.0164
4 Very Familiar	.60	.50	.18	5 Speaker 6 Arabic-English Male 1	2.55	.68	1.03	12 Speaker 13 Dhivehi-English Male 2	-1.96	.85	-2.31	11	.0416
4 Very Familiar	.60	.50	.18	5 Speaker 6 Arabic-English Male 1	2.58	.58	1.21	7 Speaker 8 Dhivehi-English Male 1	-1.98	.77	-2.57	11	.0260
4 Very Familiar	-1.46	.60	-.84	2 Speaker 3 Arabic-English Female 1	.60	.50	.18	5 Speaker 6 Arabic-English Male 1	-2.05	.79	-2.62	11	.0240
4 Very Familiar	-.37	.54	-.34	4 Speaker 5 Spanish-English Male 1	2.55	.68	1.03	12 Speaker 13 Dhivehi-English Male 2	-2.92	.87	-3.36	11	.0064
4 Very Familiar	-.37	.54	-.34	4 Speaker 5 Spanish-English Male 1	2.58	.58	1.21	7 Speaker 8 Dhivehi-English Male 1	-2.95	.80	-3.71	11	.0034
4 Very Familiar	-.48	.52	-.42	6 Speaker 7 Spanish-English Male 2	2.55	.68	1.03	12 Speaker 13 Dhivehi-English Male 2	-3.03	.86	-3.54	11	.0047
4 Very Familiar	-.48	.52	-.42	6 Speaker 7 Spanish-English Male 2	2.58	.58	1.21	7 Speaker 8 Dhivehi-English Male 1	-3.06	.78	-3.92	11	.0024
4 Very Familiar	-.75	.50	-.60	11 Speaker 12 Arabic-English Female 2	2.55	.68	1.03	12 Speaker 13 Dhivehi-English Male 2	-3.31	.85	-3.91	11	.0025
4 Very Familiar	-1.46	.60	-.84	2 Speaker 3 Arabic-English Female 1	2.49	.56	1.21	3 Speaker 4 Dhivehi-English Female 2	-3.95	.82	-4.82	11	.0005
4 Very Familiar	-1.46	.60	-.84	2 Speaker 3 Arabic-English Female 1	2.55	.68	1.03	12 Speaker 13 Dhivehi-English Male 2	-4.01	.91	-4.40	11	.0011
4 Very Familiar	-1.48	.50	-1.01	9 Speaker 10 Arabic-English Male 2	2.55	.68	1.03	12 Speaker 13 Dhivehi-English Male 2	-4.04	.85	-4.76	11	.0006
4 Very Familiar	-1.48	.50	-1.01	10 Speaker 11 Spanish-English Female 2	2.55	.68	1.03	12 Speaker 13 Dhivehi-English Male 2	-4.04	.85	-4.76	11	.0006
4 Very Familiar	-1.46	.60	-.84	2 Speaker 3 Arabic-English Female 1	2.58	.58	1.21	7 Speaker 8 Dhivehi-English Male 1	-4.04	.84	-4.81	11	.0005
4 Very Familiar	-1.81	.60	-1.03	8 Speaker 9 Spanish-English Female 1	2.55	.68	1.03	12 Speaker 13 Dhivehi-English Male 2	-4.37	.91	-4.79	11	.0006

The Facets analyses output of the Spanish-English intelligibility data

```
Facets (Many-Facet Rasch Measurement) Version No. 3.71.1  Copyright �(c) 1987-2013, John M. Linacre. All rights reserved.
24/02/2016 17:09:26

Rasch Analysis of Expert Ratings 24/02/2016 17:09:26
Table 1. Specifications from file "C:\Users\kevin\Desktop\Main Study Intelligibility Operating Files\Spanish Intelligibility SE items ONLY Operating File.txt".

Title = Rasch Analysis of Expert Ratings 24/02/2016 17:09:26
Data file = "C:\Users\kevin\Desktop\Intelligibility Data Files\Intelligibility - Spanish Fam & English Items ONLY.xlsx"
Output file = C:\Users\kevin\Dropbox\C-Leicester Stuff\Dissertation Data\For FACETS\New Intelligibility outcomes\Spanish Intelligibility SE items ONLY Operating
File.out.txt

; Data specification
Facets = 3
Delements = N
Non-centered = 1
Positive =  1, 2, 3
Labels =
 1,rater ; (elements = 189)
 2,Spanish English Familiarity Level, G  ; (elements = 4)
 3,Item ; (elements = 24)
Model = ?,?,?,D,1

; Output description
Arrange tables in order = MN
Bias/Interaction direction = plus ; ability, easiness, leniency: higher score = positive logit
Fair score = Mean
Pt-biserial = Measure
Heading lines in output data files = Y
Inter-rater coefficients reported for facet = 1
Barchart = Yes
Total score for elements = Yes
T3onscreen show only one line on screen iteration report = Y
T4MAX maximum number of unexpected observations reported in Table 4 = 100
T8NBC show table 8 numbers-barcharts-curves = NBC
Unexpected observations reported if standardized residual >= 3
Usort unexpected observations sort order = u
WHexact - Wilson-Hilferty standardization = Y

; Convergence control
Convergence = .1, .01
Iterations (maximum) = 0 ; unlimited
Xtreme scores adjusted by = .3, .5  ;(estimation, bias)

Rasch Analysis of Expert Ratings 24/02/2016 17:09:26
Table 2. Data Summary Report.

Assigning models to Data= "C:\Users\kevin\Desktop\Intelligibility Data Files\Intelligibility - Spanish Fam & English Items ONLY.xlsx"
Check (2)? Invalid datum location: 103,3,11,9 in line 106. Datum "9" is too big or not a positive integer, treated as missing.

In data: 129  3            1-24a        0            1            1            1            0            1            1            0            1            0
              1            1            1            1            1            1            0            0            1            1            1            1
              1            1
Processed: 129, 3, 1-24a, 0
Check (1)? Unspecified element: facet: 1 element: 129 129 in line 132 - see Table 2. Delements = N

Total lines in data file = 194
Total data lines = 190
Responses matched to model: ?,?,?,D,1 = 4524
    Total non-blank responses found = 4548
Responses with unspecified elements = 24
             Number of blank lines = 3
Number of missing-null observations = 12
Number of invalid observations treated as missing = 4
Valid responses used for estimation = 4520

List of unspecified elements. Please copy and paste into your specification file, where needed
Labels=Nobuild ; to suppress this list
 1, rater,  ; facet 1
 129 = 129
 *

Rasch Analysis of Expert Ratings 24/02/2016 17:09:26
Table 3. Iteration Report.

+----------------------------------------------------------+
| Iteration      Max. Score Residual       Max. Logit Change |
|              Elements    %  Categories   Elements    Steps |
|----------------------------------------------------------|
| PROX   1     Recount required             -6.7752          |
| PROX   2                                   1.7256          |
| JMLE   3    -84.9900  20.3      .0000      -.4425    .0000 |
| JMLE   4     32.0276   9.4      .0000      -.3400    .0000 |
| JMLE   5     16.7999   5.2      .0000      -.2059    .0000 |
| JMLE   6     10.7120   3.1      .0000      -.1168    .0000 |
| JMLE   7      7.1526   2.0      .0000      -.0666    .0000 |
| JMLE   8      5.5110   1.3      .0000      -.0434    .0000 |
| JMLE   9      4.3958   -.9      .0000      -.0307    .0000 |
| JMLE  10      3.4949   -.7      .0000      -.0222    .0000 |
| JMLE  11      2.7814   -.5      .0000      -.0162    .0000 |
| JMLE  12      2.2173    .4      .0000      -.0120    .0000 |
| JMLE  13      1.7700    .3      .0000      -.0090    .0000 |
| JMLE  14      1.4143    .2      .0000      -.0068    .0000 |
| JMLE  15      1.1306    .2      .0000      -.0052    .0000 |
| JMLE  16       .9039    .1      .0000      -.0040    .0000 |
| JMLE  17       .7225    .1      .0000      -.0031    .0000 |
| JMLE  18       .5775    .1      .0000      -.0024    .0000 |
| JMLE  19       .4614    .1      .0000      -.0019    .0000 |
| JMLE  20       .3687    .1      .0000      -.0015    .0000 |
| JMLE  21       .2944    .0      .0000      -.0012    .0000 |
| JMLE  22       .2352    .0      .0000      -.0010    .0000 |
| JMLE  23       .1879    .0      .0000      -.0008    .0000 |
| JMLE  24       .1500    .0      .0000      -.0006    .0000 |
| JMLE  25       .1199    .0      .0000      -.0005    .0000 |
| JMLE  26       .0957    .0      .0000      -.0004    .0000 |
+----------------------------------------------------------+
Warning (6)! There may be 4 disjoint subsets


Rasch Analysis of Expert Ratings 24/02/2016 17:09:26
Table 4. Unexpected Responses - appears after Table 8.


Rasch Analysis of Expert Ratings 24/02/2016 17:09:26
Table 5. Measurable Data Summary.

+--------------------------------------------------+
| Cat  Score  Exp.  Resd StRes|                    |
|-----------------------------+--------------------|
|  .64   .64   .64   .00  .01 | Mean (Count: 4311) |
|  .48   .48   .33   .35  .98 | S.D. (Population)  |
|  .48   .48   .33   .35  .98 | S.D. (Sample)      |
+--------------------------------------------------+
Data log-likelihood chi-square = 3347.3843
Approximate degrees of freedom = 4098
Chi-square significance prob.  = 1.0000
                                         Count   Mean   S.D.   Params
```

```
Responses in two extreme scores    =         1   0.00   0.00        0
Responses used for estimation      =      4311   0.64   0.48      213
Responses in one extreme score     =       208   0.00   0.00        2
All Responses                      =      4520   0.61   0.49      215
Count of measurable responses      =      4520
Raw-score variance of observations   =   0.23 100.00%
Variance explained by Rasch measures =   0.11  46.95%
Variance of residuals                =   0.12  53.05%

Rasch Analysis of Expert Ratings 24/02/2016 17:09:26
Table 6.0  All Facet Vertical "Rulers".

Vertical = (1*,2A,3*,S) Yardstick (columns lines low high extreme)= 0,5,-4,5,End
+------------------------------------------------------------------+
|Measr|+rater     |+Spanish English Familiarity Level        |+Item|
|-----+-----------+------------------------------------------+-----|
|   5 +           +                                          +     |
|     |           |                                          | *   |
|     |           |                                          |     |
|     |           |                                          |     |
|     |           |                                          |     |
|   4 +           +                                          +     |
|     |           |                                          |     |
|     | .         |                                          |     |
|     | .         |                                          | *   |
|     | *         |                                          |     |
|   3 +           +                                          +     |
|     | .         |                                          |     |
|     | **        |                                          |     |
|     | *         |                                          | *   |
|     | **.       |                                          |     |
|   2 + ***       +                                          + *   |
|     | ******    |                                          |     |
|     | ****.     |                                          | *   |
|     | **.       |                                          | *   |
|     | *****.    |                                          | *   |
|   1 + ***.      +                                          + *   |
|     | ********. |                                          | **  |
|     | *         |                                          |     |
|     | *****.    | Very Familiar                            |     |
|     | ***.      |                                          | **  |
*   0 * **.       * Limited Familiarity  Some Familiarity    * **  *
|     | ***.      |                                          |     |
|     | *.        | No Familiarity                           | *   |
|     | .         |                                          |     |
|     | *.        |                                          |     |
|  -1 +           +                                          +     |
|     |           |                                          | *   |
|     | .         |                                          |     |
|     |           |                                          |     |
|     | .         |                                          | *   |
|  -2 +           +                                          + *   |
|     |           |                                          | *   |
|     |           |                                          | *   |
|     |           |                                          | *   |
|     |           |                                          |     |
|  -3 + .         +                                          +     |
|     |           |                                          | *   |
|     |           |                                          |     |
|     |           |                                          |     |
|     |           |                                          |     |
|  -4 + .         +                                          + **  |
|-----+-----------+------------------------------------------+-----|
|Measr| * = 3     |+Spanish English Familiarity Level        | * = |
+------------------------------------------------------------------+

S.1: Model = ?,?,?,D
There are 4 disconnected subsets identified in Table 7.

Rasch Analysis of Expert Ratings 24/02/2016 17:09:26
Table 6.1  rater Facet Summary.

Logit:
                                             1 11 211 11
              1                1     1 2  5250706361774898362 311
 +----+----+----+----+----+----+----+----+--Q-+---S+----M----+S---+Q---+----+
-9   -8   -7   -6   -5   -4   -3   -2   -1    0    1    2    3    4    5

Infit MnSq:
    243311
   979854042 1
 +Q-S-M-S-Q+----+----+----+----+----+----+----+
 0    1    2    3    4    5    6    7    8    9

Outfit MnSq:
  124222  1
  93170274243211   1       1                  1
 +S---M---S+---Q+----+----+----+----+----+----+
 0    1    2    3    4    5    6    7    8    9

Infit ZStd:
             1   111111111
        13233325626353123366543251  1
 +----+---Q+---S+----M----S----+Q---+----+----+----+----+
-4   -3   -2   -1    0    1    2    3    4    5    6    7

Outfit ZStd:
                   112222 1
                256868835785412       1
 +----+----+----Q--S-+M--S+-Q--+----+----+----+----+----+
-4   -3   -2   -1    0    1    2    3    4    5    6    7


Rasch Analysis of Expert Ratings 24/02/2016 17:09:26
Table 6.2  Spanish English Familiarity Level Facet Summary.

Logit:
                                            1 2 1
 +----+----+----+----+----+----+----+----+----+--QSMSQ--+----+----+----+----+
-9   -8   -7   -6   -5   -4   -3   -2   -1    0    1    2    3    4    5

Infit MnSq:
      4
 +----MQ---+----+----+----+----+----+----+----+
 0    1    2    3    4    5    6    7    8    9

Outfit MnSq:
     12 1
 +--QSMS-Q-+----+----+----+----+----+----+----+
 0    1    2    3    4    5    6    7    8    9

Infit ZStd:
                1  1     11
 +----+----+-Q--+S---M----S---Q+----+----+----+----+----+
-4   -3   -2   -1    0    1    2    3    4    5    6    7

Outfit ZStd:
                1     2         1
 +----+----Q----+S---+-M--+--S-+--Q-+----+----+----+----+
-4   -3   -2   -1    0    1    2    3    4    5    6    7
```

```
Rasch Analysis of Expert Ratings 24/02/2016 17:09:26
Table 6.3  Item Facet Summary.

Logit:
      1                   1   1  11111  1   1 22  21111 1 1    1      1
 +----+----+----+-Q--+----+----S----+----+--M-+----+----+S---+----+----Q
-9   -8   -7   -6   -5   -4   -3   -2   -1    0    1    2    3    4    5

Infit MnSq:
      1
     572
 +---SMS---+----+----+----+----+----+----+----+
 0    1    2    3    4    5    6    7    8    9

Outfit MnSq:
 11 15744             1
 +-S--M---S+-Q--+----+----+----+----+----+----+
 0    1    2    3    4    5    6    7    8    9

Infit ZStd:
  1        1 1   113141412 2  1
 +----+--Q-+---S+----M----S----+Q---+----+----+----+----+
-4   -3   -2   -1    0    1    2    3    4    5    6    7

Outfit ZStd:
      1     1 11214  2121111 3                      1
 +---Q+----+-S--+----+M---+---S+----+--Q-+----+----+----+
-4   -3   -2   -1    0    1    2    3    4    5    6    7
```

Rasch Analysis of Expert Ratings 24/02/2016 17:09:26
Table 7.1.1 rater Measurement Report (arranged by MN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	Exact Agree. Obs %	Exact Agree. Exp %	Num rater	
0	21	.00	.00	(-6.34	1.94)	Minimum					.00	.00	.0	.0	188 Limited Familiarity	in subset: 1
3	21	.14	.04	-3.07	.77	2.03	1.7	9.00	3.3	-.34	.04	.44	35.5	45.0	189 Limited Familiarity	in subset: 1
5	23	.22	.14	-1.78	.64	1.21	.6	.90	.4	.80	.45	.51	58.0	58.9	186 No Familiarity	in subset: 4
7	24	.29	.20	-1.38	.58	1.18	.6	.86	.1	.82	.51	.56	60.3	59.9	93 Limited Familiarity	in subset: 1
7	24	.29	.20	-1.38	.58	1.05	.2	.76	.0	.98	.54	.56	60.0	59.9	138 Limited Familiarity	in subset: 1
9	24	.38	.30	-.85	.55	.89	-.3	.60	-.3	1.23	.62	.60	65.5	63.8	13 Some Familiarity	in subset: 3
9	24	.38	.30	-.85	.55	.78	-.8	.52	-.4	1.38	.65	.60	67.2	63.8	60 Some Familiarity	in subset: 3
10	24	.42	.32	-.77	.55	1.22	.8	1.28	.5	.64	.54	.61	60.1	63.7	183 Very Familiar	in subset: 2
9	24	.38	.32	-.75	.55	1.11	.4	1.09	.3	.82	.55	.60	60.9	64.5	61 Limited Familiarity	in subset: 1
8	24	.33	.33	-.73	.56	1.04	.2	.68	-.1	1.03	.57	.58	65.9	65.8	72 No Familiarity	in subset: 4
10	24	.42	.37	-.54	.55	1.66	2.2	1.85	1.1	.01	.42	.61	58.9	66.1	118 Some Familiarity	in subset: 3
10	24	.42	.37	-.54	.55	.81	-.7	1.27	.5	1.13	.62	.61	66.1	66.1	131 Some Familiarity	in subset: 3
11	24	.46	.38	-.47	.55	.89	-.3	.71	-.2	1.17	.64	.63	68.4	66.3	158 Very Familiar	in subset: 2
10	24	.42	.39	-.44	.55	1.21	.8	.82	.0	.83	.57	.61	66.3	66.5	126 Limited Familiarity	in subset: 1
10	24	.42	.39	-.44	.55	1.13	.5	.91	.1	.86	.57	.61	66.3	66.5	147 Limited Familiarity	in subset: 1
9	24	.38	.40	-.41	.56	.82	-.6	.56	-.3	1.31	.64	.60	69.6	67.3	40 No Familiarity	in subset: 4
9	23	.39	.40	-.39	.56	1.21	.8	1.97	1.2	.48	.50	.60	63.4	67.2	108 No Familiarity	in subset: 4
11	24	.46	.44	-.24	.55	1.19	.7	1.30	.6	.70	.56	.63	66.2	68.2	14 Some Familiarity	in subset: 3
11	24	.46	.44	-.24	.55	.82	-.6	.56	-.4	1.31	.67	.63	71.7	68.2	28 Some Familiarity	in subset: 3
11	24	.46	.44	-.24	.55	1.18	.7	1.71	1.0	.58	.54	.63	64.1	68.2	47 Some Familiarity	in subset: 3
12	24	.50	.46	-.16	.56	.74	-.9	.50	-.5	1.39	.70	.64	68.3	68.6	90 Very Familiar	in subset: 2
12	24	.50	.46	-.16	.56	.98	.0	.80	.0	1.05	.63	.64	71.1	68.6	155 Very Familiar	in subset: 2
11	24	.46	.47	-.14	.55	.69	-1.2	.48	-.6	1.46	.70	.63	71.4	68.2	3 Limited Familiarity	in subset: 1
11	24	.46	.47	-.14	.55	.94	-.1	.85	.0	1.09	.63	.63	67.6	68.2	64 Limited Familiarity	in subset: 1
11	24	.46	.47	-.14	.55	1.67	2.1	1.89	1.2	.02	.43	.63	59.7	68.2	69 Limited Familiarity	in subset: 1
11	24	.46	.47	-.14	.55	.75	-.9	.52	-.5	1.39	.68	.63	71.1	68.2	185 Limited Familiarity	in subset: 1
10	24	.42	.47	-.11	.55	.97	.0	1.06	.3	.99	.60	.61	66.4	68.5	175 No Familiarity	in subset: 4
10	21	.48	.49	-.04	.57	.87	-.4	.62	-.2	1.28	.62	.60	70.8	67.7	190 No Familiarity	in subset: 4
12	24	.50	.52	.07	.56	.62	-1.4	.43	-.7	1.51	.72	.64	74.7	70.0	12 Some Familiarity	in subset: 3
12	24	.50	.52	.07	.56	.99	.0	.85	.0	1.03	.63	.64	70.9	70.0	33 Some Familiarity	in subset: 3
12	24	.50	.52	.07	.56	1.33	1.1	1.11	.3	.63	.55	.64	66.2	70.0	35 Some Familiarity	in subset: 3
12	24	.50	.52	.07	.56	.96	.0	.87	.0	1.04	.63	.64	71.5	70.0	53 Some Familiarity	in subset: 3
12	24	.50	.52	.07	.56	1.05	.2	.79	.0	1.00	.62	.64	68.8	70.0	97 Some Familiarity	in subset: 3
12	24	.50	.52	.07	.56	.51	-2.0	.34	-.9	1.65	.75	.64	75.7	70.0	149 Some Familiarity	in subset: 3
13	24	.54	.54	.15	.57	1.83	2.2	2.11	1.3	-.07	.42	.65	65.1	70.7	36 Very Familiar	in subset: 2
12	24	.50	.54	.17	.56	1.29	1.0	2.45	1.6	.40	.51	.64	64.9	69.7	8 Limited Familiarity	in subset: 1
12	24	.50	.54	.17	.56	1.32	1.1	1.27	.5	.59	.54	.64	67.6	69.7	88 Limited Familiarity	in subset: 1
12	24	.50	.54	.17	.56	.81	-.6	.79	.0	1.22	.67	.64	71.4	69.7	102 Limited Familiarity	in subset: 1
12	24	.50	.54	.17	.56	.90	-.2	.92	.1	1.10	.64	.64	69.2	69.7	132 Limited Familiarity	in subset: 1
12	24	.50	.54	.17	.56	.83	-.5	.57	-.4	1.28	.68	.64	72.5	69.7	142 Limited Familiarity	in subset: 1
12	24	.50	.54	.17	.56	.82	-.5	.56	-.4	1.29	.68	.64	71.7	69.7	182 Limited Familiarity	in subset: 1
11	24	.46	.55	.20	.55	.74	-.9	.59	-.4	1.36	.68	.63	70.5	69.5	45 No Familiarity	in subset: 4
11	24	.46	.55	.20	.55	.79	-.7	.52	-.5	1.36	.68	.63	70.5	69.5	162 No Familiarity	in subset: 4
12	22	.55	.56	.24	.58	1.20	.7	1.61	.9	.62	.54	.63	66.0	69.9	187 Limited Familiarity	in subset: 1
13	24	.54	.59	.38	.57	.50	-1.9	.33	-.8	1.62	.76	.65	78.2	71.4	9 Some Familiarity	in subset: 3
13	24	.54	.59	.38	.57	.58	-1.5	.41	-.6	1.52	.74	.65	76.9	71.4	38 Some Familiarity	in subset: 3
13	24	.54	.59	.38	.57	.94	.0	1.12	.4	1.01	.63	.65	71.6	71.4	51 Some Familiarity	in subset: 3
13	24	.54	.59	.38	.57	.49	-1.9	.33	-.8	1.62	.76	.65	78.0	71.4	54 Some Familiarity	in subset: 3
13	24	.54	.59	.38	.57	.86	-.3	.77	.0	1.17	.67	.65	74.3	71.4	58 Some Familiarity	in subset: 3
13	24	.54	.59	.38	.57	1.68	1.9	1.35	.6	.27	.49	.65	63.0	71.4	75 Some Familiarity	in subset: 3
13	24	.54	.59	.38	.57	1.05	.2	.87	.1	.98	.63	.65	70.6	71.4	79 Some Familiarity	in subset: 3
13	24	.54	.59	.38	.57	1.28	.9	1.07	.3	.70	.57	.65	68.2	71.4	114 Some Familiarity	in subset: 3
13	24	.54	.59	.38	.57	.46	-2.1	.31	-.9	1.65	.77	.65	77.8	71.4	127 Some Familiarity	in subset: 3
13	24	.54	.59	.38	.57	.58	-1.5	.38	-.7	1.54	.74	.65	76.7	71.4	128 Some Familiarity	in subset: 3
13	24	.54	.59	.38	.57	.87	-.3	.68	-.1	1.18	.67	.65	74.1	71.4	135 Some Familiarity	in subset: 3
13	24	.54	.59	.38	.57	1.21	.7	1.24	.5	.71	.57	.65	69.6	71.4	151 Some Familiarity	in subset: 3
14	24	.58	.62	.48	.58	1.09	.3	.85	.1	.95	.63	.65	67.5	72.4	46 Very Familiar	in subset: 2
14	24	.58	.62	.48	.58	1.64	1.7	2.25	1.3	.17	.46	.65	70.6	72.4	145 Very Familiar	in subset: 2
13	24	.54	.62	.49	.57	1.35	1.1	1.23	.5	.60	.55	.65	66.8	70.8	73 Limited Familiarity	in subset: 1
13	24	.54	.62	.49	.57	1.19	.6	1.00	.2	.81	.59	.65	68.2	70.8	140 Limited Familiarity	in subset: 1
12	24	.50	.62	.50	.56	1.20	.7	1.24	.5	.71	.57	.64	66.8	70.1	10 No Familiarity	in subset: 4
12	24	.50	.62	.50	.56	1.07	.3	.78	.0	.99	.62	.64	65.4	70.1	24 No Familiarity	in subset: 4
12	24	.50	.62	.50	.56	1.33	1.1	1.13	.4	.63	.55	.64	65.2	70.1	177 No Familiarity	in subset: 4
14	24	.58	.67	.71	.58	1.50	1.4	1.26	.5	.49	.53	.65	67.4	72.5	21 Some Familiarity	in subset: 3
14	24	.58	.67	.71	.58	.61	-1.2	.40	-.5	1.47	.74	.65	78.1	72.5	34 Some Familiarity	in subset: 3
14	24	.58	.67	.71	.58	.71	-.8	.50	-.3	1.36	.72	.65	76.6	72.5	49 Some Familiarity	in subset: 3
14	24	.58	.67	.71	.58	.88	-.2	.78	.0	1.14	.67	.65	73.7	72.5	65 Some Familiarity	in subset: 3
14	24	.58	.67	.71	.58	1.09	.3	.84	.1	.95	.63	.65	72.0	72.5	96 Some Familiarity	in subset: 3
14	24	.58	.67	.71	.58	.78	-.6	.60	-.2	1.27	.70	.65	75.2	72.5	109 Some Familiarity	in subset: 3
14	24	.58	.67	.71	.58	.74	-.7	.56	-.2	1.32	.71	.65	75.8	72.5	110 Some Familiarity	in subset: 3
14	24	.58	.67	.71	.58	1.53	1.5	1.23	.5	.48	.53	.65	66.3	72.5	141 Some Familiarity	in subset: 3
14	24	.58	.67	.71	.58	.79	-.5	.68	.0	1.24	.69	.65	75.6	72.5	169 Some Familiarity	in subset: 3
14	24	.58	.67	.71	.58	1.15	.5	1.25	.5	.80	.60	.65	71.4	72.5	176 Some Familiarity	in subset: 3
14	24	.58	.69	.81	.58	.74	-.7	.56	-.2	1.32	.71	.65	74.6	71.6	2 Limited Familiarity	in subset: 1
14	24	.58	.69	.81	.58	.77	-.6	.58	-.2	1.29	.70	.65	74.6	71.6	19 Limited Familiarity	in subset: 1
14	24	.58	.69	.81	.58	.84	-.4	1.46	.7	1.06	.65	.65	73.6	71.6	20 Limited Familiarity	in subset: 1
14	24	.58	.69	.81	.58	.95	.0	.81	.0	1.06	.65	.65	73.0	71.6	37 Limited Familiarity	in subset: 1
14	24	.58	.69	.81	.58	1.14	.5	1.09	.3	.84	.61	.65	69.7	71.6	121 Limited Familiarity	in subset: 1
14	24	.58	.69	.81	.58	.93	-.1	1.40	.6	.98	.64	.65	72.5	71.6	144 Limited Familiarity	in subset: 1
14	24	.58	.69	.81	.58	1.47	1.3	1.82	1.0	.42	.52	.65	65.5	71.6	157 Limited Familiarity	in subset: 1
14	24	.58	.69	.81	.58	.79	-.5	.81	.0	1.21	.68	.65	72.8	71.6	178 Limited Familiarity	in subset: 1
14	24	.58	.69	.81	.58	1.99	2.4	2.32	1.3	-.15	.39	.65	60.0	71.6	184 Limited Familiarity	in subset: 1
15	24	.63	.69	.81	.59	.34	-2.5	.23	-.7	1.70	.79	.66	78.9	73.8	27 Very Familiar	in subset: 2
15	24	.63	.69	.81	.59	.88	-.2	.56	-.1	1.20	.69	.66	78.9	73.8	66 Very Familiar	in subset: 2
15	24	.63	.69	.81	.59	.95	.0	.65	.0	1.12	.67	.66	75.6	73.8	107 Very Familiar	in subset: 2
15	24	.63	.69	.81	.59	1.15	.5	1.23	.5	.77	.59	.66	70.7	73.8	163 Very Familiar	in subset: 2
15	24	.63	.69	.81	.59	2.36	3.0	3.63	1.8	-.68	.28	.66	61.8	73.8	165 Very Familiar	in subset: 2
15	24	.63	.69	.81	.59	.88	-.2	.56	-.1	1.20	.69	.66	78.9	73.8	174 Very Familiar	in subset: 2
13	24	.54	.69	.82	.57	1.19	.6	1.25	.5	.73	.58	.65	69.1	70.4	112 No Familiarity	in subset: 4
15	24	.63	.74	1.04	.59	.53	-1.5	.36	-.4	1.52	.75	.66	79.0	73.3	6 Some Familiarity	in subset: 3
15	24	.63	.74	1.04	.59	1.21	.7	1.75	.9	.60	.56	.66	70.2	73.3	15 Some Familiarity	in subset: 3

15	24	.63	.74	1.04	.59	.96	.0	.64	.0	1.12	.67	.66	72.9	73.3	17 Some Familiarity	in subset: 3
15	24	.63	.74	1.04	.59	1.40	1.1	1.26	.5	.58	.55	.66	69.5	73.3	29 Some Familiarity	in subset: 3
15	24	.63	.74	1.04	.59	.55	-1.5	.38	-.4	1.50	.75	.66	78.8	73.3	50 Some Familiarity	in subset: 3
15	24	.63	.74	1.04	.59	.81	-.4	.53	-.1	1.26	.70	.66	77.2	73.3	63 Some Familiarity	in subset: 3
15	24	.63	.74	1.04	.59	1.01	.1	.71	.0	1.05	.65	.66	73.0	73.3	74 Some Familiarity	in subset: 3
15	24	.63	.74	1.04	.59	.73	-.7	.63	.0	1.29	.70	.66	76.8	73.3	83 Some Familiarity	in subset: 3
15	24	.63	.74	1.04	.59	1.11	.4	.92	.2	.91	.62	.66	71.9	73.3	91 Some Familiarity	in subset: 3
15	24	.63	.74	1.04	.59	1.12	.4	1.42	.7	.77	.59	.66	72.0	73.3	123 Some Familiarity	in subset: 3
15	24	.63	.74	1.04	.59	.85	-.3	.93	.2	1.11	.66	.66	74.2	73.3	130 Some Familiarity	in subset: 3
14	24	.58	.76	1.14	.58	.98	.0	.82	.1	1.03	.65	.65	70.5	70.3	76 No Familiarity	in subset: 4
14	24	.58	.76	1.14	.58	1.24	.8	1.12	.4	.74	.58	.65	70.5	70.3	85 No Familiarity	in subset: 4
14	24	.58	.76	1.14	.58	1.74	1.9	1.79	1.0	.15	.46	.65	63.1	70.3	181 No Familiarity	in subset: 4
15	24	.63	.76	1.15	.59	.62	-1.1	.41	-.3	1.44	.74	.66	75.5	72.1	42 Limited Familiarity	in subset: 1
15	24	.63	.76	1.15	.59	1.01	.1	.69	.0	1.06	.66	.66	71.1	72.1	77 Limited Familiarity	in subset: 1
15	24	.63	.76	1.15	.59	.53	-1.5	.36	-.4	1.52	.75	.66	77.3	72.1	81 Limited Familiarity	in subset: 1
15	24	.63	.76	1.15	.59	1.17	.5	.97	.3	.84	.61	.66	71.4	72.1	92 Limited Familiarity	in subset: 1
15	24	.63	.76	1.15	.59	.55	-1.5	.38	-.4	1.50	.75	.66	77.4	72.1	104 Limited Familiarity	in subset: 1
15	24	.63	.76	1.15	.59	.80	-.5	.67	.0	1.22	.69	.66	75.4	72.1	167 Limited Familiarity	in subset: 1
15	24	.63	.76	1.15	.59	.55	-1.5	.38	-.4	1.50	.75	.66	77.8	72.1	168 Limited Familiarity	in subset: 1
15	24	.63	.76	1.15	.59	.80	-.5	.70	.0	1.21	.69	.66	74.4	72.1	170 Limited Familiarity	in subset: 1
15	24	.63	.76	1.15	.59	1.47	1.3	1.28	.5	.53	.54	.66	65.5	72.1	172 Limited Familiarity	in subset: 1
16	24	.67	.76	1.17	.60	.74	-.7	.44	-.1	1.33	.71	.65	80.6	74.7	59 Very Familiar	in subset: 2
16	24	.67	.76	1.17	.60	1.35	1.0	2.87	1.4	.46	.52	.65	68.8	74.7	87 Very Familiar	in subset: 2
16	24	.67	.76	1.17	.60	.84	-.3	.64	.1	1.18	.68	.65	75.2	74.7	143 Very Familiar	in subset: 2
16	24	.67	.76	1.17	.60	1.26	.8	1.10	.4	.74	.58	.65	75.4	74.7	146 Very Familiar	in subset: 2
14	23	.61	.78	1.24	.60	1.61	1.6	1.39	.6	.40	.52	.66	65.8	70.6	105 No Familiarity	in subset: 4
16	24	.67	.80	1.40	.60	1.06	.2	.77	.2	.98	.64	.65	73.8	73.6	18 Some Familiarity	in subset: 3
16	24	.67	.80	1.40	.60	1.84	2.1	1.80	.9	.09	.44	.65	64.6	73.6	26 Some Familiarity	in subset: 3
16	24	.67	.80	1.40	.60	1.56	1.5	1.04	.4	.53	.54	.65	68.3	73.6	67 Some Familiarity	in subset: 3
16	24	.67	.80	1.40	.60	.77	-.6	.53	.0	1.28	.70	.65	76.5	73.6	68 Some Familiarity	in subset: 3
16	24	.67	.80	1.40	.60	1.41	1.1	2.05	1.0	.41	.51	.65	69.3	73.6	94 Some Familiarity	in subset: 3
16	24	.67	.80	1.40	.60	.72	-.7	.50	.0	1.32	.71	.65	77.6	73.6	100 Some Familiarity	in subset: 3
16	24	.67	.80	1.40	.60	.37	-2.4	.23	-.4	1.67	.79	.65	80.7	73.6	150 Some Familiarity	in subset: 3
16	24	.67	.82	1.50	.60	.76	-.6	.74	.2	1.22	.69	.65	74.4	72.0	4 Limited Familiarity	in subset: 1
16	24	.67	.82	1.50	.60	.78	-.5	.75	.2	1.20	.68	.65	74.4	72.0	80 Limited Familiarity	in subset: 1
16	24	.67	.82	1.50	.60	.37	-2.4	.23	-.4	1.67	.79	.65	79.2	72.0	99 Limited Familiarity	in subset: 1
16	24	.67	.82	1.50	.60	.37	-2.4	.23	-.4	1.67	.79	.65	79.2	72.0	101 Limited Familiarity	in subset: 1
16	24	.67	.82	1.50	.60	1.33	.9	1.43	.7	.59	.55	.65	68.4	72.0	113 Limited Familiarity	in subset: 1
16	24	.67	.82	1.50	.60	1.18	.6	1.01	.4	.82	.60	.65	71.2	72.0	134 Limited Familiarity	in subset: 1
16	24	.67	.82	1.50	.60	.77	-.6	.54	.0	1.27	.70	.65	76.2	72.0	136 Limited Familiarity	in subset: 1
17	24	.71	.82	1.54	.62	1.32	.9	1.62	.8	.57	.54	.65	68.7	75.2	5 Very Familiar	in subset: 2
17	24	.71	.82	1.54	.62	.75	-.7	.53	.1	1.28	.69	.65	80.4	75.2	84 Very Familiar	in subset: 2
17	24	.71	.82	1.54	.62	.75	-.7	.53	.1	1.28	.69	.65	80.4	75.2	139 Very Familiar	in subset: 2
17	24	.71	.82	1.54	.62	1.10	.4	.78	.3	.94	.62	.65	76.3	75.2	153 Very Familiar	in subset: 2
17	24	.71	.82	1.54	.62	1.12	.4	.77	.3	.93	.62	.65	77.4	75.2	164 Very Familiar	in subset: 2
17	24	.71	.82	1.54	.62	1.20	.6	1.99	1.0	.65	.56	.65	75.6	75.2	179 Very Familiar	in subset: 2
16	23	.70	.84	1.67	.64	.34	-2.2	.21	-.3	1.64	.79	.66	79.4	72.8	106 Limited Familiarity	in subset: 1
17	24	.71	.85	1.77	.62	.98	.0	.65	.2	1.08	.65	.65	74.3	73.4	11 Some Familiarity	in subset: 3
17	24	.71	.85	1.77	.62	.91	-.1	.65	.2	1.13	.66	.65	75.6	73.4	32 Some Familiarity	in subset: 3
17	24	.71	.85	1.77	.62	1.11	.4	.64	.2	.99	.63	.65	74.2	73.4	41 Some Familiarity	in subset: 3
17	24	.71	.85	1.77	.62	1.61	1.6	1.89	.9	.24	.47	.65	64.9	73.4	43 Some Familiarity	in subset: 3
17	24	.71	.85	1.77	.62	1.09	.3	.91	.4	.90	.61	.65	73.7	73.4	44 Some Familiarity	in subset: 3
17	24	.71	.85	1.77	.62	.92	-.1	.62	.2	1.12	.66	.65	75.1	73.4	62 Some Familiarity	in subset: 3
17	24	.71	.85	1.77	.62	1.25	.8	1.24	.6	.69	.57	.65	70.2	73.4	86 Some Familiarity	in subset: 3
17	24	.71	.85	1.77	.62	.93	-.1	.67	.2	1.11	.66	.65	75.9	73.4	119 Some Familiarity	in subset: 3
17	24	.71	.85	1.77	.62	.51	-1.6	.28	-.1	1.55	.75	.65	79.5	73.4	148 Some Familiarity	in subset: 3
17	24	.71	.85	1.77	.62	1.84	2.1	1.43	.7	.17	.46	.65	65.7	73.4	154 Some Familiarity	in subset: 3
17	24	.71	.85	1.77	.62	.51	-1.6	.28	-.1	1.55	.75	.65	79.5	73.4	161 Some Familiarity	in subset: 3
17	24	.71	.85	1.77	.62	1.51	1.4	1.14	.6	.52	.53	.65	69.5	73.4	180 Some Familiarity	in subset: 3
16	24	.67	.86	1.83	.60	.85	-.3	1.06	.4	1.08	.66	.65	67.7	68.9	55 No Familiarity	in subset: 4
17	24	.71	.87	1.87	.62	1.35	1.0	2.51	1.2	.40	.51	.65	67.0	71.6	1 Limited Familiarity	in subset: 1
17	24	.71	.87	1.87	.62	.56	-1.4	.31	-.1	1.50	.74	.65	77.6	71.6	22 Limited Familiarity	in subset: 1
17	24	.71	.87	1.87	.62	1.27	.8	1.94	.9	.54	.53	.65	67.6	71.6	52 Limited Familiarity	in subset: 1
17	24	.71	.87	1.87	.62	1.06	.2	1.02	.5	.91	.62	.65	70.9	71.6	78 Limited Familiarity	in subset: 1
17	24	.71	.87	1.87	.62	.92	-.1	1.27	.6	.99	.63	.65	71.2	71.6	111 Limited Familiarity	in subset: 1
18	24	.75	.87	1.93	.64	.51	-1.7	.25	.0	1.55	.74	.64	82.9	75.3	30 Very Familiar	in subset: 2
18	24	.75	.87	1.93	.64	.51	-1.7	.25	.0	1.55	.74	.64	82.9	75.3	39 Very Familiar	in subset: 2
18	24	.75	.87	1.93	.64	.88	-.2	.68	.4	1.12	.65	.64	76.7	75.3	56 Very Familiar	in subset: 2
18	24	.75	.87	1.93	.64	.67	-1.0	.33	.1	1.40	.71	.64	79.8	75.3	71 Very Familiar	in subset: 2
18	24	.75	.87	1.93	.64	.67	-1.0	.32	.1	1.41	.71	.64	73.4	75.3	95 Very Familiar	in subset: 2
18	24	.75	.87	1.93	.64	1.63	1.7	1.70	.9	.27	.47	.64	72.6	75.3	137 Very Familiar	in subset: 2
18	24	.75	.87	1.93	.64	.88	-.2	.52	.3	1.18	.66	.64	78.0	75.3	152 Very Familiar	in subset: 2
18	24	.75	.87	1.93	.64	1.47	1.3	2.17	1.0	.35	.49	.64	73.7	75.3	171 Very Familiar	in subset: 2
17	23	.74	.88	1.99	.66	.45	-1.7	.23	.0	1.55	.77	.66	78.9	74.3	103 Some Familiarity	in subset: 3
18	24	.75	.90	2.16	.64	.82	-.4	.66	.4	1.17	.66	.64	74.1	72.8	7 Some Familiarity	in subset: 3
18	24	.75	.90	2.16	.64	.78	-.5	.51	.2	1.25	.68	.64	75.1	72.8	57 Some Familiarity	in subset: 3
18	24	.75	.90	2.16	.64	.89	-.2	.74	.4	1.10	.65	.64	73.0	72.8	133 Some Familiarity	in subset: 3
18	24	.75	.90	2.16	.64	.92	-.1	1.15	.6	.99	.62	.64	71.8	72.8	156 Some Familiarity	in subset: 3
17	24	.71	.90	2.20	.62	.94	.0	.65	.2	1.11	.66	.65	68.9	67.6	16 No Familiarity	in subset: 4
17	24	.71	.90	2.20	.62	1.06	.2	1.86	.9	.81	.59	.65	64.7	67.6	23 No Familiarity	in subset: 4
17	24	.71	.90	2.20	.62	.51	-1.6	.28	-.1	1.55	.75	.65	71.5	67.6	117 No Familiarity	in subset: 4
18	24	.75	.91	2.27	.64	1.13	.4	5.12	1.7	.53	.53	.64	67.4	70.8	48 Limited Familiarity	in subset: 1
19	24	.79	.91	2.37	.68	.95	.0	1.77	1.0	.86	.58	.63	74.8	74.9	122 Very Familiar	in subset: 2
19	24	.79	.91	2.37	.68	.63	-1.1	.28	.3	1.42	.70	.63	81.2	74.9	166 Very Familiar	in subset: 2
19	24	.79	.91	2.37	.68	.84	-.3	.38	.4	1.24	.66	.63	76.6	74.9	173 Very Familiar	in subset: 2
19	24	.79	.93	2.60	.68	.54	-1.5	.24	.2	1.50	.72	.63	75.7	71.9	70 Some Familiarity	in subset: 3
19	24	.79	.93	2.60	.68	1.17	.5	1.04	.7	.80	.57	.63	69.9	71.9	115 Some Familiarity	in subset: 3
19	24	.79	.93	2.60	.68	.54	-1.5	.24	.2	1.50	.72	.63	75.7	71.9	116 Some Familiarity	in subset: 3
19	24	.79	.93	2.60	.68	.54	-1.5	.24	.2	1.50	.72	.63	75.7	71.9	125 Some Familiarity	in subset: 3
19	24	.79	.93	2.60	.68	1.08	.3	1.90	1.0	.75	.56	.63	69.3	71.9	159 Some Familiarity	in subset: 3
19	24	.79	.93	2.60	.68	1.01	.1	1.61	.9	.83	.58	.63	69.1	71.9	160 Some Familiarity	in subset: 3
19	24	.79	.94	2.70	.68	1.05	.2	.92	.7	.92	.60	.63	67.1	69.5	31 Limited Familiarity	in subset: 1
19	24	.79	.94	2.70	.68	.92	-.1	1.57	.9	.91	.59	.63	68.2	69.5	82 Limited Familiarity	in subset: 1
20	24	.83	.96	3.19	.73	.57	-1.1	.22	.6	1.43	.69	.61	70.8	67.9	25 Limited Familiarity	in subset: 1
20	24	.83	.96	3.19	.73	.67	-.8	.27	.6	1.34	.68	.61	70.5	67.9	89 Limited Familiarity	in subset: 1
20	24	.83	.96	3.19	.73	.61	-1.0	.24	.6	1.40	.69	.61	69.7	67.9	124 Limited Familiarity	in subset: 1
21	24	.88	.97	3.47	.84	1.29	.6	.50	1.1	.88	.56	.59	75.6	72.9	98 Very Familiar	in subset: 2
21	24	.88	.98	3.70	.84	.74	-.4	.25	1.0	1.25	.65	.59	70.5	68.7	120 Some Familiarity	in subset: 3
14.5	23.9	.61	.70	.96	.61	.99	-.1	.99	.3		.62				Mean (Count: 189)	
3.2	.4	.13	.19	1.14	.11	.36	1.1	.88	.6		.11				S.D. (Population)	
3.2	.4	.13	.19	1.14	.11	.36	1.1	.88	.6		.11				S.D. (Sample)	

With extremes, Model, Populn: RMSE .62 Adj (True) S.D. .96 Separation 1.56 Strata 2.41 Reliability (not inter-rater) .71
With extremes, Model, Sample: RMSE .62 Adj (True) S.D. .96 Separation 1.57 Strata 2.42 Reliability (not inter-rater) .71
Without extremes, Model, Populn: RMSE .60 Adj (True) S.D. .81 Separation 1.35 Strata 2.14 Reliability (not inter-rater) .65
Without extremes, Model, Sample: RMSE .60 Adj (True) S.D. .82 Separation 1.36 Strata 2.15 Reliability (not inter-rater) .65
With extremes, Model, Fixed (all same) chi-square: 498.2 d.f.: 188 significance (probability): .00
With extremes, Model, Random (normal) chi-square: 132.6 d.f.: 187 significance (probability): 1.00
Inter-Rater agreement opportunities: 120260 Exact agreements: 86297 = 71.8% Expected: 85801.2 = 71.3%

Rasch Analysis of Expert Ratings 24/02/2016 17:09:26
Table 7.2.1 Spanish English Familiarity Level Measurement Report (arranged by MN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	N Spanish English Familiarity Level	
245	474	.52	.66	-.36	.13	1.07	1.0	1.01	.1	.92	.63	.65	1 No Familiarity	in subset: 4
790	1359	.58	.73	-.03	.08	.98	-.4	1.43	2.1	1.00	.66	.66	2 Limited Familiarity	in subset: 1
1144	1847	.62	.75	.08	.07	.96	-1.0	.84	-1.0	1.05	.66	.65	3 Some Familiarity	in subset: 3
564	840	.67	.79	.31	.10	1.05	.8	1.02	.1	.94	.64	.65	4 Very Familiar	in subset: 2
685.8	1130.0	.60	.73	.00	.09	1.01	.1	1.08	.4		.65		Mean (Count: 4)	
327.9	519.8	.06	.05	.24	.02	.04	.9	.21	1.1		.01		S.D. (Population)	
378.6	600.3	.07	.06	.28	.03	.05	1.0	.25	1.3		.02		S.D. (Sample)	

Model, Populn: RMSE .10 Adj (True) S.D. .22 Separation 2.25 Strata 3.33 Reliability .83
Model, Sample: RMSE .10 Adj (True) S.D. .26 Separation 2.66 Strata 3.88 Reliability .88

Model, Fixed (all same) chi-square: 17.3 d.f.: 3 significance (probability): .00
Model, Random (normal) chi-square: 2.5 d.f.: 2 significance (probability): .28

Rasch Analysis of Expert Ratings 24/02/2016 17:09:26
Table 7.3.1 Item Measurement Report (arranged by MN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	Nu Item
0	189	.00	.00	(-8.00	1.83)	Minimum					.00	.00	1 S5 jug
15	189	.08	.05	-3.90	.28	1.09	.4	1.00	.1	.94	.16	.23	5 S5 function
27	189	.14	.10	-3.17	.22	1.10	.7	3.88	6.1	.84	.16	.29	17 S9 transferring
41	185	.22	.17	-2.58	.19	1.13	1.2	1.32	1.6	.79	.22	.36	8 S7 awkwardly
45	189	.24	.19	-2.45	.19	1.19	1.8	1.32	1.6	.70	.18	.34	18 S9 along
54	184	.29	.24	-2.15	.18	1.02	.2	1.04	.3	.96	.36	.38	13 S9 change
58	189	.31	.26	-2.04	.17	.85	-2.0	.82	-1.3	1.29	.48	.37	10 S7 they're
66	182	.36	.32	-1.76	.17	.94	-.8	.92	-.6	1.15	.46	.40	11 S7 staying
89	189	.47	.45	-1.19	.16	.79	-3.7	.73	-3.0	1.65	.56	.40	23 S11 suffered
118	189	.62	.64	-.42	.17	1.00	.0	.94	-.5	1.03	.42	.41	22 S11 patient
130	189	.69	.72	-.08	.17	.87	-1.6	.77	-1.8	1.27	.51	.41	24 S11 seizure
133	189	.70	.73	.02	.18	1.09	1.1	1.17	1.2	.81	.33	.41	2 S5 stood
136	189	.72	.75	.11	.18	.94	-.6	.92	-.5	1.10	.45	.41	21 S11 attack
137	189	.72	.76	.14	.18	.97	-.3	.88	-.7	1.08	.44	.41	16 S9 agency
154	189	.81	.85	.76	.21	1.07	.6	1.38	1.6	.88	.30	.39	3 S5 shelf
157	189	.83	.87	.89	.21	1.09	.7	1.19	.8	.90	.31	.39	19 S11 soldiers
158	189	.84	.87	.94	.22	1.04	.3	.97	.0	.99	.38	.39	9 S7 noisy
164	189	.87	.90	1.24	.23	1.05	.3	1.14	.5	.95	.33	.38	15 S9 strength
166	189	.88	.91	1.36	.24	.93	-.3	.72	-.9	1.07	.44	.37	4 S5 washing
171	189	.90	.94	1.69	.27	.97	-.1	1.13	.4	1.00	.36	.37	14 S9 demanded
174	189	.92	.95	1.93	.30	.97	.0	.56	-1.1	1.06	.42	.36	20 S11 swept
179	189	.95	.97	2.45	.36	1.09	.4	1.46	.9	.94	.28	.35	6 S5 broke
184	189	.97	.99	3.37	.52	.76	-.4	.26	-.9	1.12	.50	.34	12 S7 improvement
187	189	.99	1.00	4.84	1.02	.76	.0	.04	-.5	1.12	.50	.35	7 S7 machine
114.3	188.3	.61	.61	-.33	.33	.99	-.1	1.07	.1		.36		Mean (Count: 24)
59.0	1.8	.31	.34	2.63	.36	.12	1.2	.69	1.7		.13		S.D. (Population)
60.3	1.9	.32	.35	2.69	.37	.12	1.2	.70	1.8		.14		S.D. (Sample)

With extremes, Model, Populn: RMSE .49 Adj (True) S.D. 2.58 Separation 5.32 Strata 7.43 Reliability .97
With extremes, Model, Sample: RMSE .49 Adj (True) S.D. 2.64 Separation 5.44 Strata 7.58 Reliability .97
Without extremes, Model, Populn: RMSE .32 Adj (True) S.D. 2.11 Separation 6.65 Strata 9.19 Reliability .98
Without extremes, Model, Sample: RMSE .32 Adj (True) S.D. 2.16 Separation 6.80 Strata 9.40 Reliability .98
With extremes, Model, Fixed (all same) chi-square: 1333.8 d.f.: 23 significance (probability): .00
With extremes, Model, Random (normal) chi-square: 19.9 d.f.: 22 significance (probability): .59

Rasch Analysis of Expert Ratings 24/02/2016 17:09:26
Table 8.1 Category Statistics.

Model = ?,?,?,D

DATA Score	Category Counts Total	Used	%	Cum. %	QUALITY CONTROL Avge Meas	Exp. Meas	OUTFIT MnSq
0	1777	1568	36%	36%	-.95	-.95	.9
1	2743	2743	64%	100%	2.21	2.21	1.4

Rasch Analysis of Expert Ratings 24/02/2016 17:09:26
Table 4.1 Unexpected Responses (78 residuals sorted by u).

Cat	Score	Exp.	Resd	StRes	Num	rater	N	Spanish English Fam	Nu	Item
0	0	1.0	-1.0	-9.0	48	Limited Familiarity	2	Limited Familiarity	6	S5 broke
1	1	.0	1.0	9.0	189	Limited Familiarity	2	Limited Familiarity	17	S9 transferring
0	0	1.0	-1.0	-7.1	87	Very Familiar	4	Very Familiar	6	S5 broke
0	0	1.0	-1.0	-6.1	159	Some Familiarity	3	Some Familiarity	9	S7 noisy
0	0	1.0	-1.0	-6.0	122	Very Familiar	4	Very Familiar	19	S11 soldiers
0	0	1.0	-1.0	-6.0	165	Very Familiar	4	Very Familiar	6	S5 broke
0	0	1.0	-1.0	-5.8	1	Limited Familiarity	2	Limited Familiarity	14	S9 demanded
0	0	1.0	-1.0	-5.8	8	Limited Familiarity	2	Limited Familiarity	12	S7 improvement
0	0	1.0	-1.0	-5.8	23	No Familiarity	1	No Familiarity	14	S9 demanded
0	0	1.0	-1.0	-5.8	179	Very Familiar	4	Very Familiar	14	S9 demanded
0	0	1.0	-1.0	-5.7	171	Very Familiar	4	Very Familiar	15	S9 strength
0	0	1.0	-1.0	-5.6	82	Limited Familiarity	2	Limited Familiarity	3	S5 shelf
0	0	1.0	-1.0	-5.6	160	Some Familiarity	3	Some Familiarity	3	S5 shelf
1	1	.0	1.0	5.3	108	No Familiarity	1	No Familiarity	8	S7 awkwardly
0	0	1.0	-1.0	-5.0	20	Limited Familiarity	2	Limited Familiarity	6	S5 broke
0	0	1.0	-1.0	-5.0	145	Very Familiar	4	Very Familiar	6	S5 broke
0	0	1.0	-1.0	-4.9	94	Some Familiarity	3	Some Familiarity	14	S9 demanded
1	1	.0	1.0	4.8	144	Limited Familiarity	2	Limited Familiarity	5	S5 function
1	1	.0	1.0	4.8	157	Limited Familiarity	2	Limited Familiarity	5	S5 function
1	1	.0	1.0	4.8	184	Limited Familiarity	2	Limited Familiarity	5	S5 function
0	0	1.0	-1.0	-4.7	5	Very Familiar	4	Very Familiar	15	S9 strength
0	0	1.0	-1.0	-4.7	52	Limited Familiarity	2	Limited Familiarity	15	S9 strength
0	0	1.0	-1.0	-4.7	111	Limited Familiarity	2	Limited Familiarity	15	S9 strength
1	1	.0	1.0	4.6	131	Some Familiarity	3	Some Familiarity	8	S7 awkwardly
0	0	1.0	-1.0	-4.5	137	Very Familiar	4	Very Familiar	3	S5 shelf
0	0	1.0	-1.0	-4.5	156	Some Familiarity	3	Some Familiarity	3	S5 shelf
1	1	.0	1.0	4.4	187	Limited Familiarity	2	Limited Familiarity	17	S9 transferring
1	1	.1	.9	4.3	118	Some Familiarity	3	Some Familiarity	18	S9 along
0	0	.9	-.9	-4.1	15	Some Familiarity	3	Some Familiarity	14	S9 demanded
0	0	.9	-.9	-4.1	55	No Familiarity	1	No Familiarity	4	S5 washing
0	0	.9	-.9	-4.0	115	Some Familiarity	3	Some Familiarity	21	S11 attack
1	1	.1	.9	4.0	123	Some Familiarity	3	Some Familiarity	5	S5 function
0	0	.9	-.9	-3.9	1	Limited Familiarity	2	Limited Familiarity	19	S11 soldiers
1	1	.1	.9	3.9	36	Very Familiar	4	Very Familiar	17	S9 transferring
0	0	.9	-.9	-3.9	43	Some Familiarity	3	Some Familiarity	19	S11 soldiers
1	1	.1	.9	3.9	47	Some Familiarity	3	Some Familiarity	8	S7 awkwardly
1	1	.1	.9	3.9	51	Some Familiarity	3	Some Familiarity	17	S9 transferring
0	0	.9	-.9	-3.9	52	Limited Familiarity	2	Limited Familiarity	19	S11 soldiers
1	1	.1	.9	3.9	69	Limited Familiarity	2	Limited Familiarity	8	S7 awkwardly
0	0	.9	-.9	-3.9	176	Some Familiarity	3	Some Familiarity	20	S11 swept
0	0	.9	-.9	-3.8	31	Limited Familiarity	2	Limited Familiarity	2	S5 stood
1	1	.1	.9	3.7	14	Some Familiarity	3	Some Familiarity	18	S9 along
0	0	.9	-.9	-3.7	43	Some Familiarity	3	Some Familiarity	3	S5 shelf
1	1	.1	.9	3.7	47	Some Familiarity	3	Some Familiarity	18	S9 along
0	0	.9	-.9	-3.7	78	Limited Familiarity	2	Limited Familiarity	3	S5 shelf
0	0	.9	-.9	-3.7	86	Some Familiarity	3	Some Familiarity	3	S5 shelf
1	1	.1	.9	3.7	175	No Familiarity	1	No Familiarity	13	S9 change
1	1	.1	.9	3.7	183	Very Familiar	4	Very Familiar	13	S9 change
1	1	.1	.9	3.6	61	Limited Familiarity	2	Limited Familiarity	11	S7 staying
0	0	.9	-.9	-3.5	130	Some Familiarity	3	Some Familiarity	4	S5 washing
0	0	.9	-.9	-3.5	165	Very Familiar	4	Very Familiar	4	S5 washing
1	1	.1	.9	3.4	26	Some Familiarity	3	Some Familiarity	5	S5 function
1	1	.1	.9	3.4	94	Some Familiarity	3	Some Familiarity	5	S5 function
1	1	.1	.9	3.4	113	Limited Familiarity	2	Limited Familiarity	5	S5 function
1	1	.1	.9	3.4	132	Limited Familiarity	2	Limited Familiarity	8	S7 awkwardly
1	1	.1	.9	3.4	146	Very Familiar	4	Very Familiar	5	S5 function
0	0	.9	-.9	-3.4	181	No Familiarity	1	No Familiarity	14	S9 demanded
0	0	.9	-.9	-3.3	4	Limited Familiarity	2	Limited Familiarity	19	S11 soldiers
0	0	.9	-.9	-3.3	15	Some Familiarity	3	Some Familiarity	15	S9 strength
0	0	.9	-.9	-3.3	26	Some Familiarity	3	Some Familiarity	19	S11 soldiers
0	0	.9	-.9	-3.3	36	Very Familiar	4	Very Familiar	20	S11 swept
0	0	.9	-.9	-3.3	80	Limited Familiarity	2	Limited Familiarity	19	S11 soldiers
0	0	.9	-.9	-3.3	112	No Familiarity	1	No Familiarity	20	S11 swept

```
|  0     0      .9   -.9 -3.3 | 113 Limited Familiarity 2 Limited Familiarity  9 S7 noisy        |
|  1     1      .1    .9  3.3 | 121 Limited Familiarity 2 Limited Familiarity 17 S9 transferring |
|  0     0      .9   -.9 -3.3 | 133 Some Familiarity    3 Some Familiarity    16 S9 agency       |
|  0     0      .9   -.9 -3.3 | 172 Limited Familiarity 2 Limited Familiarity 15 S9 strength     |
|  1     1      .1    .9  3.3 | 178 Limited Familiarity 2 Limited Familiarity 17 S9 transferring |
|  1     1      .1    .9  3.2 |   8 Limited Familiarity 2 Limited Familiarity 18 S9 along        |
|  1     1      .1    .9  3.2 |  10 No Familiarity      1 No Familiarity      18 S9 along        |
|  1     1      .1    .9  3.2 |  88 Limited Familiarity 2 Limited Familiarity 18 S9 along        |
|  1     1      .1    .9  3.2 | 102 Limited Familiarity 2 Limited Familiarity 18 S9 along        |
|  0     0      .9   -.9 -3.1 |   7 Some Familiarity    3 Some Familiarity     2 S5 stood        |
|  0     0      .9   -.9 -3.1 |  56 Very Familiar       4 Very Familiar        2 S5 stood        |
|  1     1      .1    .9  3.1 | 105 No Familiarity      1 No Familiarity      17 S9 transferring |
|  0     0      .9   -.9 -3.1 | 134 Limited Familiarity 2 Limited Familiarity  3 S5 shelf        |
|  1     1      .1    .9  3.0 |  64 Limited Familiarity 2 Limited Familiarity 10 S7 they're      |
|  1     1      .1    .9  3.0 |  69 Limited Familiarity 2 Limited Familiarity 10 S7 they're      |
|-----------------------------+------------------------------------------------------------------|
| Cat  Score  Exp.  Resd StRes| Num rater               N Spanish English Fam Nu Item            |
+------------------------------------------------------------------------------------------------+
```

The Facets analyses output of the Arabic-English intelligibility data

```
Facets (Many-Facet Rasch Measurement) Version No. 3.71.1  Copyright �(c) 1987-2013, John M. Linacre. All rights reserved.
24/02/2016 15:40:38

Rasch Analysis of Expert Ratings 24/02/2016 15:40:38
Table 1. Specifications from file "C:\Users\kevin\Desktop\Main Study Intelligibility Operating Files\Arabic Intelligibility AE items ONLY Operating File.txt".

Title = Rasch Analysis of Expert Ratings 24/02/2016 15:40:38
Data file = "C:\Users\kevin\Desktop\Main Study Intelligibility Operating Files\Intelligibility - Arabic Fam & English Items ONLY.xlsx"
Output file = C:\Users\kevin\Desktop\Main Study Intelligibility Operating Files\Arabic Intelligibility AE items ONLY Operating File.out.txt

; Data specification
Facets = 3
Delements = N
Non-centered = 1
Positive =  1, 2, 3
Labels =
 1,rater ; (elements = 190)
 2,Arabic English Familiarity Level, G  ; (elements = 4)
 3,Item ; (elements = 25)
Model = ?,?,?,D,1

; Output description
Arrange tables in order = MN
Bias/Interaction direction = plus ; ability, easiness, leniency: higher score = positive logit
Fair score = Mean
Pt-biserial = Measure
Heading lines in output data files = Y
Inter-rater coefficients reported for facet = 1
Barchart = Yes
Total score for elements = Yes
T3onscreen show only one line on screen iteration report = Y
T4MAX maximum number of unexpected observations reported in Table 4 = 100
T8NBC show table 8 numbers-barcharts-curves = NBC
Unexpected observations reported if standardized residual >= 3
Usort unexpected observations sort order = u
WHexact - Wilson-Hilferty standardization = Y

; Convergence control
Convergence = .1, .01
Iterations (maximum) = 0 ; unlimited
Xtreme scores adjusted by = .3, .5  ;(estimation, bias)

Rasch Analysis of Expert Ratings 24/02/2016 15:40:38
Table 2. Data Summary Report.

Assigning models to Data= "C:\Users\kevin\Desktop\Main Study Intelligibility Operating Files\Intelligibility - Arabic Fam & English Items ONLY.xlsx"
Total lines in data file = 194
Total data lines = 190
Responses matched to model: ?,?,?,D,1 = 4736
    Total non-blank responses found = 4736
              Number of blank lines = 3
Number of missing-null observations = 14
Valid responses used for estimation = 4736

Rasch Analysis of Expert Ratings 24/02/2016 15:40:38
Table 3. Iteration Report.

+-------------------------------------------------------------+
| Iteration      Max. Score Residual      Max. Logit Change |
|            Elements    %  Categories   Elements    Steps |
|-------------------------------------------------------------|
| PROX   1     Recount required            4.7855           |
| PROX   2                                 1.2505           |
| JMLE   3     80.0001  -9.0      .0000      .2848    .0000 |
| JMLE   4     34.4489   4.1      .0000      .1442    .0000 |
| JMLE   5     17.2841   2.7      .0000      .0969    .0000 |
| JMLE   6      9.9111   2.0      .0000      .0688    .0000 |
| JMLE   7      6.4167   1.5      .0000      .0513    .0000 |
| JMLE   8      4.5726   1.2      .0000      .0396    .0000 |
| JMLE   9      3.4777    .9      .0000      .0312    .0000 |
| JMLE  10      2.9257    .7      .0000      .0250    .0000 |
| JMLE  11      2.4531    .6      .0000      .0202    .0000 |
| JMLE  12      2.0424    .5      .0000      .0164    .0000 |
| JMLE  13      1.6929    .4      .0000      .0133    .0000 |
| JMLE  14      1.3990    .3      .0000      .0109    .0000 |
| JMLE  15      1.1536    .3      .0000      .0089    .0000 |
| JMLE  16       .9493    .2      .0000      .0072    .0000 |
| JMLE  17       .7803    .2      .0000      .0059    .0000 |
| JMLE  18       .6404    .1      .0000      .0048    .0000 |
| JMLE  19       .5251    .1      .0000      .0040    .0000 |
| JMLE  20       .4303    .1      .0000      .0032    .0000 |
| JMLE  21       .3523    .1      .0000      .0026    .0000 |
| JMLE  22       .2883    .1      .0000      .0022    .0000 |
| JMLE  23       .2360    .1      .0000      .0018    .0000 |
| JMLE  24       .1930    .0      .0000      .0014    .0000 |
| JMLE  25       .1577    .0      .0000      .0012    .0000 |
| JMLE  26       .1290    .0      .0000      .0010    .0000 |
| JMLE  27       .1055    .0      .0000      .0008    .0000 |
| JMLE  28       .0862    .0      .0000      .0006    .0000 |
+-------------------------------------------------------------+
Warning (6)! There may be 4 disjoint subsets


Rasch Analysis of Expert Ratings 24/02/2016 15:40:38
Table 4. Unexpected Responses - appears after Table 8.


Rasch Analysis of Expert Ratings 24/02/2016 15:40:38
Table 5. Measurable Data Summary.

+---------------------------------------------------+
| Cat  Score  Exp.  Resd StRes|                      |
|-----------------------------+----------------------|
|  .70   .70   .70   .00  .03 | Mean (Count: 4546)   |
|  .46   .46   .31   .34 1.00 | S.D. (Population)    |
|  .46   .46   .31   .34 1.00 | S.D. (Sample)        |
+---------------------------------------------------+
Data log-likelihood chi-square = 3271.4163
Approximate degrees of freedom = 4330
Chi-square significance prob.  = 1.0000
                                         Count   Mean   S.D.   Params
Responses used for estimation      =      4546   0.70   0.46      216
Responses in one extreme score     =       190   1.00   0.00        1
All Responses                      =      4736   0.71   0.45      217
Count of measurable responses      =      4736
Raw-score variance of observations   =   0.21 100.00%
Variance explained by Rasch measures =   0.10  46.66%
Variance of residuals                =   0.11  53.34%

Rasch Analysis of Expert Ratings 24/02/2016 15:40:38
Table 6.0  All Facet Vertical "Rulers".
```

```
Vertical = (1*,2A,3*,S) Yardstick (columns lines low high extreme)= 0,5,-5,5,End
+---------------------------------------------------------------------------------------+
|Measr|+rater     |+Arabic English Familiarity Level                               |+Item|
|-----+----------+----------------------------------------------------------------+-----|
|   5 +          +                                                                + *   |
|     | .        |                                                                |     |
|     |          |                                                                |     |
|     |          |                                                                |     |
|     |          |                                                                |     |
|   4 + *        +                                                                +     |
|     | .        |                                                                |     |
|     |          |                                                                |     |
|     |          |                                                                |     |
|     | ****     |                                                                | *** |
|   3 +          +                                                                +     |
|     | **.      |                                                                |     |
|     | *.       |                                                                |     |
|     | **.      |                                                                |     |
|     | *****.   |                                                                |     |
|   2 + ******.  +                                                                +     |
|     | ***.     |                                                                | *   |
|     | ***      |                                                                | *   |
|     | ****.    |                                                                | **  |
|     | *******. |                                                                |     |
|   1 + **.      +                                                                + *   |
|     | *.       |                                                                |     |
|     | ****     |                                                                | **  |
|     | *.       |                                                                | *   |
|     | *.       | Very Familiar                                                  | **  |
*   0 * **.      * Limited Familiarity  No Familiarity        Some Familiarity    *     *
|     | **.      |                                                                | *** |
|     | .        |                                                                | *   |
|     | .        |                                                                |     |
|     | **       |                                                                | *   |
|  -1 +          +                                                                +     |
|     | *        |                                                                | *   |
|     | .        |                                                                |     |
|     |          |                                                                | *   |
|     |          |                                                                |     |
|  -2 + .        +                                                                +     |
|     |          |                                                                |     |
|     |          |                                                                | *   |
|     |          |                                                                |     |
|     |          |                                                                |     |
|  -3 +          +                                                                + *   |
|     |          |                                                                |     |
|     |          |                                                                |     |
|     |          |                                                                |     |
|     |          |                                                                |     |
|  -4 +          +                                                                + *   |
|     |          |                                                                |     |
|     |          |                                                                | *   |
|     |          |                                                                |     |
|     |          |                                                                |     |
|  -5 +          +                                                                +     |
|-----+----------+----------------------------------------------------------------+-----|
|Measr| * = 3    |+Arabic English Familiarity Level                               | * = |
+---------------------------------------------------------------------------------------+

S.1: Model = ?,?,?,D
There are 4 disconnected subsets identified in Table 7.

Rasch Analysis of Expert Ratings 24/02/2016 15:40:38
Table 6.1  rater Facet Summary.

Logit:
                             1  21 121    1
                1  23 6127844247239107757 2  23   1
 +----+----+----+----Q----+S---+-M--+--S-+---Q+----+----+
-5   -4   -3   -2   -1    0    1    2    3    4    5    6

Infit MnSq:
   124332
   7690138132
 +Q-S-M-S-Q+----+----+----+----+----+----+----+
 0    1    2    3    4    5    6    7    8    9

Outfit MnSq:
  323211
  57437286263411 12 1112 1   1              1
 S----M----+S---+Q---+----+----+----+----+----+
 0    1    2    3    4    5    6    7    8    9

Infit ZStd:
             1 1 11211  11
            35 5702152896684431322 1 1
 +----+---Q+----S----M----S----Q----+----+----+
-4   -3   -2   -1    0    1    2    3    4    5

Outfit ZStd:
                  1222211
                259752028966431 322 2 2
 +----+----+--Q-+-S--+M---S----Q----+----+----+
-4   -3   -2   -1    0    1    2    3    4    5


Rasch Analysis of Expert Ratings 24/02/2016 15:40:38
Table 6.2  Arabic English Familiarity Level Facet Summary.

Logit:
                          31
 +----+----+----+----+---QMQ---+----+----+----+----+----+
-5   -4   -3   -2   -1    0    1    2    3    4    5    6

Infit MnSq:
     121
 +---QMS---+----+----+----+----+----+----+----+
 0    1    2    3    4    5    6    7    8    9

Outfit MnSq:
     112
 +--QSMSQ--+----+----+----+----+----+----+----+
 0    1    2    3    4    5    6    7    8    9

Infit ZStd:
      1             1    1          1
 +----+----S----+----+M---+----+S---+----+-Q--+
-4   -3   -2   -1    0    1    2    3    4    5

Outfit ZStd:
            1          1   11
 +----+--Q-+----S----+-M--+--S-+----Q----+----+
-4   -3   -2   -1    0    1    2    3    4    5


Rasch Analysis of Expert Ratings 24/02/2016 15:40:38
Table 6.3  Item Facet Summary.

Logit:
    1 1    1  1   1 1 1 13 212 1 211      3            1
```

```
 +---Q+----+----S----+----+M---+----+-S--+----+---Q+----+
-5   -4   -3   -2   -1    0    1    2    3    4    5    6

Infit MnSq:
      1
    1626
 +--QSMS---+----+----+----+----+----+----+----+
 0    1    2    3    4    5    6    7    8    9

Outfit MnSq:
  116454 1  21
 Q-S--M--S-+Q---+----+----+----+----+----+----+
 0    1    2    3    4    5    6    7    8    9

Infit ZStd:
   1 1    1  1  2111132112 1211         1
 +---Q+----+-S--+----M----+--S-+----+Q---+----+
-4   -3   -2   -1    0    1    2    3    4    5

Outfit ZStd:
     1 1 1 1  3  3  23  23 1    11     1   1
 +-Q--+----+S---+----M----+---S+----+--Q-+----+
-4   -3   -2   -1    0    1    2    3    4    5
```

Rasch Analysis of Expert Ratings 24/02/2016 15:40:38
Table 7.1.1 rater Measurement Report (arranged by MN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	Exact Agree. Obs %	Exact Agree. Exp %	Num rater	
6	25	.24	.11	-2.05	.61	1.18	.6	1.12	.6	.78	.51	.58	46.9	48.1	186 Limited Familiarity	in subset: 1
8	25	.32	.20	-1.39	.55	.90	-.2	.81	.2	1.11	.61	.59	53.4	53.7	93 Limited Familiarity	in subset: 1
8	25	.32	.20	-1.39	.55	.75	-.8	.49	-.1	1.39	.67	.59	55.3	53.7	138 Limited Familiarity	in subset: 1
9	25	.36	.22	-1.28	.53	.81	-.7	.73	.0	1.27	.64	.60	57.5	54.3	24 Some Familiarity	in subset: 3
9	25	.36	.22	-1.28	.53	.83	-.6	.57	-.1	1.32	.65	.60	56.2	54.3	75 Some Familiarity	in subset: 3
8	21	.38	.24	-1.15	.59	2.19	3.3	5.64	2.3	-1.35	.20	.63	48.2	58.8	189 Limited Familiarity	in subset: 1
9	22	.41	.29	-.89	.57	1.98	3.0	4.50	2.2	-1.02	.26	.63	47.5	60.4	188 No Familiarity	in subset: 2
10	25	.40	.31	-.82	.52	1.60	2.2	3.72	2.1	-.38	.33	.60	45.9	58.9	126 No Familiarity	in subset: 2
10	25	.40	.31	-.82	.52	1.31	1.2	1.87	1.0	.37	.47	.60	52.5	58.9	162 No Familiarity	in subset: 2
10	25	.40	.31	-.82	.52	.99	.0	.70	.0	1.11	.61	.60	59.2	58.7	10 Limited Familiarity	in subset: 1
10	25	.40	.31	-.82	.52	.87	-.4	.67	-.1	1.24	.64	.60	59.2	58.7	183 Limited Familiarity	in subset: 1
11	25	.44	.33	-.73	.52	1.01	.1	2.30	1.4	.73	.55	.61	57.6	59.8	44 Some Familiarity	in subset: 3
11	25	.44	.37	-.55	.52	.97	.0	.78	.0	1.09	.61	.61	58.9	61.2	102 No Familiarity	in subset: 2
12	25	.48	.39	-.46	.51	1.19	.8	1.04	.2	.69	.54	.61	62.4	62.4	43 Some Familiarity	in subset: 3
12	25	.48	.39	-.46	.51	1.11	.5	1.99	1.3	.60	.53	.61	60.6	62.4	79 Some Familiarity	in subset: 3
12	25	.48	.43	-.29	.51	1.26	1.1	1.92	1.2	.28	.47	.61	60.8	63.4	40 No Familiarity	in subset: 2
12	25	.48	.43	-.29	.51	.95	-.1	.74	-.1	1.14	.62	.61	63.3	63.4	141 No Familiarity	in subset: 2
12	25	.48	.43	-.29	.51	.65	-1.7	.46	-.7	1.66	.71	.61	64.9	63.4	177 No Familiarity	in subset: 2
12	25	.48	.43	-.28	.51	.96	.0	1.17	.4	.97	.59	.61	61.6	63.2	69 Limited Familiarity	in subset: 1
12	25	.48	.43	-.28	.51	1.04	.2	1.28	.6	.82	.57	.61	59.9	63.2	184 Limited Familiarity	in subset: 1
13	25	.52	.44	-.24	.51	1.37	1.5	1.42	.7	.31	.48	.61	59.0	64.8	118 Very Familiar	in subset: 4
13	25	.52	.45	-.20	.51	.79	-.9	.57	-.5	1.43	.67	.61	69.4	64.8	31 Some Familiarity	in subset: 3
13	25	.52	.49	-.02	.51	1.45	1.8	4.75	3.3	-.25	.38	.61	55.2	65.2	28 Limited Familiarity	in subset: 1
13	25	.52	.49	-.02	.51	.87	-.5	.68	-.3	1.27	.65	.61	66.1	65.2	61 Limited Familiarity	in subset: 1
13	25	.52	.49	-.02	.51	.99	.0	4.08	2.9	.64	.53	.61	62.9	65.2	64 Limited Familiarity	in subset: 1
14	25	.56	.51	.03	.52	.86	-.5	.60	-.5	1.31	.66	.61	68.2	67.4	81 Very Familiar	in subset: 4
14	25	.56	.52	.06	.52	.97	.0	.69	-.3	1.15	.63	.61	70.0	67.1	17 Some Familiarity	in subset: 3
14	25	.56	.52	.06	.52	1.01	.1	.76	-.2	1.06	.61	.61	69.3	67.1	34 Some Familiarity	in subset: 3
14	25	.56	.52	.06	.52	1.35	1.4	1.59	1.0	.36	.48	.61	63.8	67.1	131 Some Familiarity	in subset: 3
14	25	.56	.52	.06	.52	1.06	.3	1.37	.7	.80	.56	.61	65.8	67.1	158 Some Familiarity	in subset: 3
14	25	.56	.56	.24	.52	1.19	.8	.91	.0	.78	.56	.61	64.0	67.4	108 No Familiarity	in subset: 2
14	25	.56	.56	.24	.52	1.69	2.5	4.12	3.0	-.74	.27	.61	60.5	67.4	173 No Familiarity	in subset: 2
14	25	.56	.56	.24	.52	.99	.0	.92	.0	1.01	.60	.61	67.2	67.1	85 Limited Familiarity	in subset: 1
14	25	.56	.56	.24	.52	1.04	.2	3.40	2.6	.56	.52	.61	64.7	67.1	144 Limited Familiarity	in subset: 1
15	25	.60	.57	.30	.53	1.40	1.5	2.88	2.2	.21	.44	.61	64.7	69.7	87 Very Familiar	in subset: 4
15	25	.60	.58	.34	.53	1.20	.8	1.08	.3	.71	.54	.61	65.3	69.2	97 Some Familiarity	in subset: 3
15	25	.60	.58	.34	.53	1.37	1.4	2.36	1.7	.25	.45	.61	64.7	69.2	151 Some Familiarity	in subset: 3
15	25	.60	.58	.34	.53	.85	-.5	.63	-.4	1.27	.66	.61	71.4	69.2	157 Some Familiarity	in subset: 3
15	25	.60	.63	.51	.53	.96	.0	.83	.0	1.07	.61	.61	70.1	69.1	76 No Familiarity	in subset: 2
15	25	.60	.63	.51	.53	1.15	.6	1.10	.3	.75	.55	.61	68.3	69.1	121 No Familiarity	in subset: 2
15	25	.60	.63	.52	.53	1.14	.6	1.06	.3	.77	.55	.61	67.9	68.9	5 Limited Familiarity	in subset: 1
15	25	.60	.63	.52	.53	.84	-.6	.60	-.5	1.31	.66	.61	70.8	68.9	8 Limited Familiarity	in subset: 1
15	25	.60	.63	.52	.53	.96	.0	.83	.0	1.07	.61	.61	69.6	68.9	60 Limited Familiarity	in subset: 1
15	25	.60	.63	.52	.53	1.69	2.4	3.34	2.5	-.30	.34	.61	59.0	68.9	142 Limited Familiarity	in subset: 1
15	25	.60	.63	.52	.53	1.22	.9	1.09	.3	.69	.54	.61	67.3	68.9	145 Limited Familiarity	in subset: 1
16	25	.64	.64	.59	.54	1.58	1.9	3.16	2.3	-.01	.37	.61	66.9	71.9	36 Very Familiar	in subset: 4
16	25	.64	.64	.59	.54	1.22	.8	.97	.1	.76	.55	.61	69.5	71.9	66 Very Familiar	in subset: 4
16	25	.64	.65	.62	.54	.71	-1.1	.47	-.7	1.45	.70	.61	76.4	71.2	32 Some Familiarity	in subset: 3
16	25	.64	.65	.62	.54	1.20	.7	.93	.1	.79	.55	.61	68.4	71.2	47 Some Familiarity	in subset: 3
16	25	.64	.65	.62	.54	.79	-.7	.53	-.6	1.34	.68	.61	75.5	71.2	143 Some Familiarity	in subset: 3
16	25	.64	.69	.80	.54	.98	.0	1.12	.4	.94	.59	.61	70.0	70.7	179 No Familiarity	in subset: 2
16	25	.64	.69	.80	.54	.72	-1.0	.65	-.3	1.36	.68	.61	74.0	70.4	68 Limited Familiarity	in subset: 1
16	25	.64	.69	.80	.54	1.11	.4	.94	.1	.88	.57	.61	69.1	70.4	91 Limited Familiarity	in subset: 1
16	25	.64	.69	.80	.54	1.17	.6	.87	.0	.85	.57	.61	68.8	70.4	111 Limited Familiarity	in subset: 1
17	25	.68	.72	.93	.56	1.35	1.1	2.10	1.4	.44	.46	.61	69.2	73.0	6 Some Familiarity	in subset: 3
17	25	.68	.72	.93	.56	1.08	.3	.75	-.1	.99	.59	.61	74.8	73.0	15 Some Familiarity	in subset: 3
17	25	.68	.72	.93	.56	.83	-.4	.70	-.2	1.20	.65	.61	76.6	73.0	54 Some Familiarity	in subset: 3
17	25	.68	.72	.93	.56	.85	-.4	.62	-.3	1.22	.65	.61	75.7	73.0	65 Some Familiarity	in subset: 3
17	25	.68	.72	.93	.56	.92	-.1	.61	-.3	1.16	.64	.61	76.5	73.0	133 Some Familiarity	in subset: 3
17	25	.68	.72	.93	.56	1.04	.2	.71	-.2	1.04	.61	.61	72.3	73.0	172 Some Familiarity	in subset: 3
17	25	.68	.72	.93	.56	.92	-.1	.76	-.1	1.11	.63	.61	73.9	73.0	187 Some Familiarity	in subset: 3
17	25	.68	.75	1.10	.56	1.25	.8	1.31	.6	.68	.52	.61	69.5	72.0	16 No Familiarity	in subset: 2
17	25	.68	.75	1.10	.56	1.65	1.9	1.97	1.2	.15	.39	.61	65.6	72.0	53 No Familiarity	in subset: 2
17	25	.68	.75	1.10	.56	1.89	2.4	2.12	1.4	-.09	.33	.61	60.3	72.0	55 No Familiarity	in subset: 2
17	25	.68	.75	1.10	.56	1.42	1.3	1.29	.6	.53	.48	.61	67.8	72.0	112 No Familiarity	in subset: 2
17	25	.68	.75	1.10	.56	1.04	.2	.86	.0	.97	.59	.61	70.3	72.0	117 No Familiarity	in subset: 2
17	25	.68	.75	1.10	.56	.94	-.1	.72	-.1	1.11	.62	.61	72.7	72.0	140 No Familiarity	in subset: 2
17	25	.68	.75	1.10	.56	2.00	2.6	2.42	1.6	-.26	.28	.61	61.4	72.0	168 No Familiarity	in subset: 2
17	25	.68	.75	1.11	.56	.88	-.3	.62	-.3	1.19	.64	.61	73.7	71.7	1 Limited Familiarity	in subset: 1
17	25	.68	.75	1.11	.56	1.21	.7	1.06	.3	.77	.54	.61	69.2	71.7	13 Limited Familiarity	in subset: 1
17	25	.68	.75	1.11	.56	.88	-.3	.58	-.4	1.21	.65	.61	72.4	71.7	39 Limited Familiarity	in subset: 1
17	25	.68	.75	1.11	.56	1.31	1.0	1.04	.2	.70	.52	.61	67.6	71.7	150 Limited Familiarity	in subset: 1
18	25	.72	.77	1.22	.59	1.15	.5	1.21	.5	.81	.54	.61	73.5	75.4	23 Very Familiar	in subset: 4
18	25	.72	.77	1.22	.59	.63	-1.1	.44	-.5	1.40	.70	.61	76.5	75.4	42 Very Familiar	in subset: 4
18	25	.72	.77	1.22	.59	1.18	.6	1.60	.8	.70	.51	.61	73.0	75.4	107 Very Familiar	in subset: 4
18	25	.72	.77	1.22	.59	.69	-.9	.41	-.6	1.37	.70	.61	76.5	75.4	122 Very Familiar	in subset: 4
18	25	.72	.77	1.22	.59	1.60	1.6	1.25	.5	.45	.44	.61	70.0	75.4	171 Very Familiar	in subset: 4
18	25	.72	.78	1.26	.59	.63	-1.1	.44	-.5	1.40	.70	.61	79.7	74.5	12 Some Familiarity	in subset: 3
18	25	.72	.78	1.26	.59	.67	-.9	.40	-.6	1.39	.70	.61	78.6	74.5	105 Some Familiarity	in subset: 3
18	25	.72	.78	1.26	.59	.83	-.4	.56	-.3	1.22	.66	.61	78.4	74.5	113 Some Familiarity	in subset: 3
18	25	.72	.78	1.26	.59	1.31	.9	1.02	.3	.73	.52	.61	74.0	74.5	115 Some Familiarity	in subset: 3
18	25	.72	.78	1.26	.59	.74	-.7	.52	-.4	1.29	.67	.61	78.5	74.5	127 Some Familiarity	in subset: 3
18	25	.72	.78	1.26	.59	.97	.0	.67	-.1	1.09	.62	.61	76.7	74.5	156 Some Familiarity	in subset: 3
18	25	.72	.81	1.43	.59	.83	-.4	.56	-.3	1.22	.66	.61	72.8	73.1	14 No Familiarity	in subset: 2
18	25	.72	.81	1.43	.59	.76	-.6	.47	-.5	1.30	.68	.61	74.0	73.1	20 No Familiarity	in subset: 2
18	25	.72	.81	1.43	.59	1.34	1.0	1.52	.8	.58	.48	.61	68.8	73.1	72 No Familiarity	in subset: 2
18	25	.72	.81	1.43	.59	.63	-1.1	.44	-.5	1.40	.70	.61	74.0	73.1	146 No Familiarity	in subset: 2
18	25	.72	.81	1.43	.59	.75	-.6	.51	-.4	1.29	.67	.61	72.7	73.1	180 No Familiarity	in subset: 2
18	25	.72	.81	1.44	.59	.92	-.1	.69	-.1	1.11	.62	.61	73.3	72.8	22 Limited Familiarity	in subset: 1
18	25	.72	.81	1.44	.59	.82	-.4	.96	.2	1.13	.63	.61	73.5	72.8	50 Limited Familiarity	in subset: 1
18	25	.72	.81	1.44	.59	.47	-1.8	.29	-.9	1.56	.75	.61	78.8	72.8	88 Limited Familiarity	in subset: 1
18	25	.72	.81	1.44	.59	.80	-.5	.49	-.4	1.27	.67	.61	74.9	72.8	128 Limited Familiarity	in subset: 1
18	25	.72	.81	1.44	.59	.47	-1.8	.29	-.9	1.56	.75	.61	78.8	72.8	147 Limited Familiarity	in subset: 1
18	25	.72	.81	1.44	.59	1.08	.3	.90	.1	.94	.58	.61	71.2	72.8	163 Limited Familiarity	in subset: 1
18	25	.72	.81	1.44	.59	1.18	.6	.78	.0	.91	.57	.61	70.9	72.8	167 Limited Familiarity	in subset: 1
18	25	.72	.81	1.44	.59	1.16	.5	1.27	.5	.81	.54	.61	70.5	72.8	170 Limited Familiarity	in subset: 1
19	25	.76	.84	1.62	.62	1.31	.8	1.30	.6	.70	.49	.60	74.5	75.7	21 Some Familiarity	in subset: 3
19	25	.76	.84	1.62	.62	.79	-.4	1.18	.4	1.10	.62	.60	78.9	75.7	46 Some Familiarity	in subset: 3

19	25	.76	.84	1.62	.62	.57	-1.2	.35	-.5	1.41	.71	.60	81.2	75.7	52 Some Familiarity	in subset: 3
19	25	.76	.84	1.62	.62	1.22	.6	.84	.1	.87	.55	.60	74.9	75.7	56 Some Familiarity	in subset: 3
19	25	.76	.84	1.62	.62	.70	-.7	.41	-.4	1.31	.68	.60	79.1	75.7	57 Some Familiarity	in subset: 3
19	25	.76	.84	1.62	.62	1.23	.6	1.37	.6	.73	.50	.60	74.3	75.7	83 Some Familiarity	in subset: 3
19	25	.76	.84	1.62	.62	.46	-1.6	.25	-.7	1.51	.74	.60	82.5	75.7	99 Some Familiarity	in subset: 3
19	25	.76	.84	1.62	.62	.58	-1.1	.37	-.5	1.39	.71	.60	80.7	75.7	129 Some Familiarity	in subset: 3
19	25	.76	.84	1.62	.62	.91	-.1	1.42	.7	.98	.58	.60	77.7	75.7	176 Some Familiarity	in subset: 3
13	18	.72	.85	1.71	.77	.36	-1.6	.19	-.5	1.55	.79	.68	80.1	76.2	190 No Familiarity	in subset: 2
19	25	.76	.86	1.80	.62	1.15	.4	1.53	.7	.81	.53	.60	71.8	73.8	95 No Familiarity	in subset: 2
19	25	.76	.86	1.80	.62	.88	-.2	.53	-.2	1.17	.64	.60	74.6	73.8	166 No Familiarity	in subset: 2
19	25	.76	.86	1.80	.62	.50	-1.5	.28	-.7	1.47	.73	.60	78.8	73.6	35 Limited Familiarity	in subset: 1
19	25	.76	.86	1.80	.62	.80	-.4	1.33	.6	1.08	.61	.60	74.5	73.6	74 Limited Familiarity	in subset: 1
19	25	.76	.86	1.80	.62	.85	-.3	.48	-.3	1.21	.65	.60	75.8	73.6	84 Limited Familiarity	in subset: 1
19	25	.76	.86	1.80	.62	1.35	.9	.85	.1	.79	.52	.60	70.9	73.6	90 Limited Familiarity	in subset: 1
19	25	.76	.86	1.80	.62	1.15	.5	.93	.2	.90	.55	.60	72.7	73.6	123 Limited Familiarity	in subset: 1
19	25	.76	.86	1.80	.62	.46	-1.6	.25	-.7	1.51	.74	.60	78.4	73.6	136 Limited Familiarity	in subset: 1
19	25	.76	.86	1.80	.62	.86	-.2	.65	.0	1.14	.63	.60	74.2	73.6	152 Limited Familiarity	in subset: 1
19	25	.76	.86	1.80	.62	.73	-.6	.39	-.5	1.31	.68	.60	77.5	73.6	178 Limited Familiarity	in subset: 1
20	25	.80	.88	2.01	.67	.61	-.9	.39	-.2	1.32	.68	.58	78.3	77.6	7 Very Familiar	in subset: 4
20	25	.80	.88	2.01	.67	.41	-1.6	.21	-.6	1.48	.73	.58	82.2	77.6	48 Very Familiar	in subset: 4
20	25	.80	.88	2.01	.67	2.12	2.1	2.22	1.1	.08	.27	.58	67.3	77.6	181 Very Familiar	in subset: 4
20	25	.80	.89	2.04	.67	.51	-1.2	.26	-.4	1.42	.71	.58	82.8	76.5	9 Some Familiarity	in subset: 3
20	25	.80	.89	2.04	.67	.51	-1.2	.26	-.4	1.42	.71	.58	82.8	76.5	18 Some Familiarity	in subset: 3
20	25	.80	.89	2.04	.67	.51	-1.2	.26	-.4	1.42	.71	.58	82.8	76.5	19 Some Familiarity	in subset: 3
20	25	.80	.89	2.04	.67	.51	-1.2	.26	-.4	1.42	.71	.58	82.8	76.5	29 Some Familiarity	in subset: 3
20	25	.80	.89	2.04	.67	.61	-.9	.39	-.2	1.32	.68	.58	81.6	76.5	38 Some Familiarity	in subset: 3
20	25	.80	.89	2.04	.67	.78	-.4	.46	-.1	1.22	.64	.58	78.9	76.5	62 Some Familiarity	in subset: 3
20	25	.80	.89	2.04	.67	.62	-.8	.42	-.2	1.30	.67	.58	81.1	76.5	63 Some Familiarity	in subset: 3
20	25	.80	.89	2.04	.67	1.19	.5	.91	.3	.87	.53	.58	77.0	76.5	98 Some Familiarity	in subset: 3
20	25	.80	.89	2.04	.67	.54	-1.1	.29	-.4	1.39	.70	.58	80.8	76.5	103 Some Familiarity	in subset: 3
20	25	.80	.89	2.04	.67	.87	-.1	.51	-.1	1.15	.62	.58	78.7	76.5	106 Some Familiarity	in subset: 3
20	25	.80	.89	2.04	.67	.41	-1.6	.21	-.6	1.48	.73	.58	83.0	76.5	109 Some Familiarity	in subset: 3
20	25	.80	.89	2.04	.67	.41	-1.6	.21	-.6	1.48	.73	.58	83.0	76.5	114 Some Familiarity	in subset: 3
20	25	.80	.89	2.04	.67	.51	-1.2	.26	-.4	1.42	.71	.58	82.8	76.5	116 Some Familiarity	in subset: 3
20	25	.80	.89	2.04	.67	.72	-.5	1.58	.8	1.08	.60	.58	79.4	76.5	130 Some Familiarity	in subset: 3
20	25	.80	.89	2.04	.67	1.53	1.2	1.36	.6	.57	.43	.58	72.0	76.5	132 Some Familiarity	in subset: 3
20	25	.80	.89	2.04	.67	.41	-1.6	.21	-.6	1.48	.73	.58	83.0	76.5	135 Some Familiarity	in subset: 3
20	25	.80	.89	2.04	.67	.89	-.1	.54	.0	1.14	.62	.58	78.2	76.5	137 Some Familiarity	in subset: 3
20	25	.80	.90	2.22	.67	1.24	.6	.71	.1	.91	.54	.58	73.4	74.2	4 No Familiarity	in subset: 2
20	25	.80	.90	2.22	.67	.79	-.3	.48	-.1	1.20	.64	.58	75.4	74.2	41 No Familiarity	in subset: 2
20	25	.80	.90	2.22	.67	.54	-1.1	.29	-.4	1.39	.70	.58	75.5	74.2	101 No Familiarity	in subset: 2
20	25	.80	.90	2.22	.67	.91	.0	.54	.0	1.12	.61	.58	75.2	74.2	164 No Familiarity	in subset: 2
20	25	.80	.90	2.22	.67	1.08	.3	.80	.2	.97	.56	.58	71.2	74.2	169 No Familiarity	in subset: 2
20	25	.80	.90	2.22	.67	.54	-1.1	.29	-.4	1.39	.70	.58	78.3	74.0	2 Limited Familiarity	in subset: 1
20	25	.80	.90	2.22	.67	1.46	1.0	2.03	1.0	.56	.43	.58	71.0	74.0	26 Limited Familiarity	in subset: 1
20	25	.80	.90	2.22	.67	.82	-.3	.40	-.2	1.22	.64	.58	75.3	74.0	45 Limited Familiarity	in subset: 1
20	25	.80	.90	2.22	.67	.73	-.5	2.11	1.1	1.04	.58	.58	74.5	74.0	58 Limited Familiarity	in subset: 1
20	25	.80	.90	2.22	.67	.41	-1.6	.21	-.6	1.48	.73	.58	78.4	74.0	73 Limited Familiarity	in subset: 1
20	25	.80	.90	2.22	.67	1.46	1.0	2.03	1.0	.56	.43	.58	71.0	74.0	92 Limited Familiarity	in subset: 1
20	25	.80	.90	2.22	.67	.51	-1.2	.26	-.4	1.42	.71	.58	77.8	74.0	94 Limited Familiarity	in subset: 1
20	25	.80	.90	2.22	.67	1.30	.7	1.36	.6	.72	.48	.58	72.5	74.0	104 Limited Familiarity	in subset: 1
20	25	.80	.90	2.22	.67	.81	-.3	.41	-.2	1.22	.64	.58	77.0	74.0	110 Limited Familiarity	in subset: 1
20	25	.80	.90	2.22	.67	1.67	1.4	9.00	3.3	-.02	.23	.58	66.4	74.0	139 Limited Familiarity	in subset: 1
20	25	.80	.90	2.22	.67	.61	-.9	.39	-.2	1.32	.68	.58	76.1	74.0	155 Limited Familiarity	in subset: 1
20	25	.80	.90	2.22	.67	.66	-.7	.54	.0	1.25	.65	.58	76.1	74.0	159 Limited Familiarity	in subset: 1
21	25	.84	.92	2.49	.73	.74	-.4	.56	.1	1.18	.61	.56	78.3	78.0	30 Very Familiar	in subset: 4
21	25	.84	.92	2.49	.73	.37	-1.6	.16	-.3	1.47	.71	.56	82.2	78.0	78 Very Familiar	in subset: 4
21	25	.84	.92	2.49	.73	1.05	.2	.54	.1	1.03	.56	.56	78.3	78.0	86 Very Familiar	in subset: 4
21	25	.84	.92	2.49	.73	.37	-1.6	.16	-.3	1.47	.71	.56	82.2	78.0	96 Very Familiar	in subset: 4
21	25	.84	.92	2.49	.73	.93	.0	.41	.0	1.13	.59	.56	79.2	78.0	119 Very Familiar	in subset: 4
21	25	.84	.92	2.49	.73	1.10	.3	.61	.2	.99	.54	.56	77.9	78.0	153 Very Familiar	in subset: 4
21	25	.84	.92	2.49	.73	1.18	.5	2.58	1.2	.73	.46	.56	74.8	78.0	175 Very Familiar	in subset: 4
21	25	.84	.93	2.53	.73	.74	-.4	.30	-.1	1.26	.64	.56	80.5	76.8	27 Some Familiarity	in subset: 3
21	25	.84	.93	2.53	.73	.74	-.4	.56	.1	1.18	.61	.56	78.7	76.8	49 Some Familiarity	in subset: 3
21	25	.84	.93	2.53	.73	.67	-.6	.36	.0	1.26	.64	.56	78.5	76.8	77 Some Familiarity	in subset: 3
21	25	.84	.93	2.53	.73	1.34	.8	1.22	.6	.75	.46	.56	74.5	76.8	125 Some Familiarity	in subset: 3
21	25	.84	.93	2.53	.73	.37	-1.6	.16	-.3	1.47	.71	.56	83.3	76.8	160 Some Familiarity	in subset: 3
21	25	.84	.94	2.70	.73	.74	-.4	.30	-.1	1.26	.64	.56	75.1	74.1	134 No Familiarity	in subset: 2
21	25	.84	.94	2.70	.73	2.07	1.9	1.67	.8	.27	.30	.56	66.6	74.1	165 No Familiarity	in subset: 2
21	25	.84	.94	2.70	.73	1.29	.7	.98	.5	.80	.48	.56	72.5	74.1	174 No Familiarity	in subset: 2
21	25	.84	.94	2.70	.73	.74	-.4	.30	-.1	1.26	.64	.56	75.1	74.1	182 No Familiarity	in subset: 2
21	25	.84	.94	2.71	.73	.81	-.2	2.44	1.1	.94	.53	.56	73.5	73.9	51 Limited Familiarity	in subset: 1
21	25	.84	.94	2.71	.73	1.17	.4	.81	.3	.90	.51	.56	73.2	73.9	89 Limited Familiarity	in subset: 1
21	25	.84	.94	2.71	.73	1.23	.6	.62	.2	.93	.52	.56	74.2	73.9	100 Limited Familiarity	in subset: 1
22	25	.88	.96	3.11	.80	.37	-1.5	.13	.0	1.44	.66	.52	81.0	76.5	33 Some Familiarity	in subset: 3
22	25	.88	.96	3.11	.80	.37	-1.5	.13	.0	1.44	.66	.52	81.0	76.5	71 Some Familiarity	in subset: 3
22	25	.88	.96	3.11	.80	.59	-.7	.20	.0	1.32	.62	.52	80.8	76.5	82 Some Familiarity	in subset: 3
22	25	.88	.96	3.11	.80	1.40	.8	.96	.6	.76	.42	.52	75.6	76.5	124 Some Familiarity	in subset: 3
22	25	.88	.96	3.11	.80	.59	-.7	.20	.0	1.32	.62	.52	80.8	76.5	154 Some Familiarity	in subset: 3
22	25	.88	.96	3.28	.80	1.34	.7	.51	.3	.90	.47	.52	71.2	73.3	11 No Familiarity	in subset: 2
22	25	.88	.96	3.28	.80	.76	-.3	.32	.2	1.21	.58	.52	74.8	73.3	80 No Familiarity	in subset: 2
22	25	.88	.96	3.28	.80	1.46	.9	4.38	1.6	.48	.33	.52	67.9	73.3	149 No Familiarity	in subset: 2
22	25	.88	.96	3.29	.80	1.31	.7	1.73	.9	.70	.40	.52	71.0	73.2	3 Limited Familiarity	in subset: 1
22	25	.88	.96	3.29	.80	.87	-.1	.28	.1	1.18	.57	.52	74.5	73.2	59 Limited Familiarity	in subset: 1
22	25	.88	.96	3.29	.80	.37	-1.5	.13	.0	1.44	.66	.52	76.6	73.2	67 Limited Familiarity	in subset: 1
22	25	.88	.96	3.29	.80	.59	-.7	.20	.0	1.32	.62	.52	74.9	73.2	148 Limited Familiarity	in subset: 1
23	25	.92	.98	3.83	.90	1.13	.4	.53	.7	.94	.42	.45	74.6	75.5	37 Some Familiarity	in subset: 3
23	25	.92	.98	3.83	.90	.78	-.2	.19	.5	1.22	.51	.45	77.5	75.5	70 Some Familiarity	in subset: 3
23	25	.92	.98	4.00	.90	.78	-.2	.19	.5	1.22	.51	.45	72.5	71.8	185 No Familiarity	in subset: 2
23	25	.92	.98	4.00	.90	1.32	.7	.82	.9	.78	.36	.45	70.4	71.8	25 Limited Familiarity	in subset: 1
23	25	.92	.98	4.00	.90	1.00	.1	.30	.6	1.08	.47	.45	71.4	71.8	161 Limited Familiarity	in subset: 1
24	25	.96	.99	4.83	1.14	1.42	.7	.59	1.5	.75	.23	.33	73.6	73.7	120 Some Familiarity	in subset: 3
17.7	24.9	.71	.76	1.43	.63	.97	.0	1.03	.3		.58				Mean (Count: 190)	
3.6	.6	.14	.20	1.21	.10	.37	1.1	1.10	.8		.12				S.D. (Population)	
3.6	.6	.14	.20	1.21	.10	.37	1.1	1.11	.8		.12				S.D. (Sample)	

Model, Populn: RMSE .64 Adj (True) S.D. 1.03 Separation 1.62 Strata 2.49 Reliability (not inter-rater) .72
Model, Sample: RMSE .64 Adj (True) S.D. 1.03 Separation 1.62 Strata 2.50 Reliability (not inter-rater) .72
Model, Fixed (all same) chi-square: 682.1 d.f.: 189 significance (probability): .00
Model, Random (normal) chi-square: 158.2 d.f.: 188 significance (probability): .94
Inter-Rater agreement opportunities: 122705 Exact agreements: 88763 = 72.3% Expected: 88080.4 = 71.8%

Rasch Analysis of Expert Ratings 24/02/2016 15:40:38
Table 7.2.1 Arabic English Familiarity Level Measurement Report (arranged by MN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	N Arabic English Familiarity Level	
1065	1546	.69	.79	-.10	.08	.99	-.2	1.23	1.4	.99	.63	.64	2 Limited Familiarity	in subset: 1
683	990	.69	.79	-.10	.09	1.16	2.9	1.25	1.2	.81	.59	.63	1 No Familiarity	in subset: 2
1238	1700	.73	.82	.08	.07	.88	-2.9	.72	-1.8	1.14	.65	.61	3 Some Familiarity	in subset: 3
371	500	.74	.82	.12	.14	1.07	.8	1.09	.4	.93	.59	.60	4 Very Familiar	in subset: 4
839.3	1184.0	.71	.81	.00	.10	1.02	.2	1.07	.3		.61		Mean (Count: 4)	
336.8	475.1	.02	.02	.10	.03	.10	2.1	.21	1.3		.03		S.D. (Population)	
388.9	548.6	.03	.02	.11	.03	.12	2.5	.25	1.5		.03		S.D. (Sample)	

Model, Populn: RMSE .10 Adj (True) S.D. .00 Separation .00 Strata .33 Reliability .00
Model, Sample: RMSE .10 Adj (True) S.D. .06 Separation .56 Strata 1.08 Reliability .24
Model, Fixed (all same) chi-square: 4.5 d.f.: 3 significance (probability): .21
Model, Random (normal) chi-square: 1.7 d.f.: 2 significance (probability): .42

Rasch Analysis of Expert Ratings 24/02/2016 15:40:38
Table 7.3.1 Item Measurement Report (arranged by MN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	Nu	Item
16	190	.08	.05	-4.33	.28	1.08	.4	2.30	2.2	.85	.13	.27	18	S10 feared
21	190	.11	.07	-3.99	.25	.92	-.4	.90	-.1	1.05	.35	.30	6	S3 learned
40	190	.21	.16	-3.09	.19	1.14	1.3	2.40	4.4	.64	.18	.38	8	S6 paid
62	190	.33	.28	-2.36	.17	1.30	3.7	1.64	3.5	.28	.17	.43	14	S10 blend
89	190	.47	.46	-1.61	.16	1.09	1.3	1.09	.8	.77	.40	.46	25	S12 patch
106	187	.57	.58	-1.12	.17	.95	-.6	.91	-.8	1.14	.51	.47	7	S3 law
115	190	.61	.63	-.89	.17	1.09	1.2	1.07	.5	.82	.41	.47	20	S12 playing
131	189	.69	.74	-.39	.18	1.16	1.7	1.18	1.1	.74	.35	.47	12	S6 quite
134	188	.71	.76	-.28	.18	1.16	1.6	1.14	.8	.77	.35	.46	19	S10 bridge
137	190	.72	.77	-.22	.18	.73	-3.2	.59	-2.8	1.42	.66	.46	2	S3 cloth
138	187	.74	.78	-.14	.19	.68	-3.6	.53	-3.2	1.46	.68	.45	4	S3 wet
146	190	.77	.82	.10	.19	.84	-1.5	.65	-1.9	1.23	.57	.44	1	S3 tea
150	190	.79	.84	.26	.20	.91	-.7	.72	-1.3	1.13	.51	.43	5	S3 jury
153	190	.81	.86	.38	.21	1.09	.7	1.17	.7	.90	.36	.43	23	S12 boy
156	187	.83	.88	.59	.22	.73	-2.2	.47	-2.4	1.29	.61	.40	3	S3 quite
159	189	.84	.89	.69	.22	1.01	.0	.78	-.7	1.03	.42	.40	13	S6 usual
164	190	.86	.91	.91	.23	.85	-1.0	.62	-1.3	1.15	.51	.39	24	S12 wore
171	190	.90	.94	1.33	.26	.81	-1.0	.53	-1.3	1.15	.48	.35	21	S12 involved
172	190	.91	.94	1.40	.27	1.15	.8	2.26	2.4	.79	.15	.34	22	S12 hands
173	189	.92	.95	1.55	.28	1.11	.5	.94	.0	.94	.27	.33	11	S6 credit
176	190	.93	.96	1.72	.30	.97	.0	.90	.0	1.03	.34	.31	10	S6 server
186	190	.98	.99	3.16	.52	.87	-.1	.26	-.7	1.09	.30	.18	15	S10 rather
186	190	.98	.99	3.16	.52	1.01	.1	1.28	.5	.99	.17	.18	16	S10 different
186	190	.98	.99	3.16	.52	1.03	.2	.54	-.2	1.01	.19	.18	17	S10 children
190	190	1.00	1.00	(5.83	1.83)	Maximum					.00	.00	9	S6 change
134.3	189.4	.71	.73	.23	.32	.99	.0	1.04	.0		.36			Mean (Count: 25)
50.6	1.0	.27	.29	2.26	.33	.16	1.6	.57	1.9		.17			S.D. (Population)
51.7	1.0	.27	.30	2.31	.33	.16	1.6	.58	1.9		.18			S.D. (Sample)

With extremes, Model, Populn: RMSE .45 Adj (True) S.D. 2.22 Separation 4.88 Strata 6.84 Reliability .96
With extremes, Model, Sample: RMSE .45 Adj (True) S.D. 2.27 Separation 4.99 Strata 6.98 Reliability .96
Without extremes, Model, Populn: RMSE .27 Adj (True) S.D. 1.98 Separation 7.19 Strata 9.93 Reliability .98
Without extremes, Model, Sample: RMSE .27 Adj (True) S.D. 2.02 Separation 7.35 Strata 10.14 Reliability .98
With extremes, Model, Fixed (all same) chi-square: 1258.4 d.f.: 24 significance (probability): .00
With extremes, Model, Random (normal) chi-square: 21.5 d.f.: 23 significance (probability): .55

Rasch Analysis of Expert Ratings 24/02/2016 15:40:38
Table 8.1 Category Statistics.

Model = ?,?,?,D

DATA					QUALITY CONTROL		
	Category Counts			Cum.	Avge	Exp.	OUTFIT
Score	Total	Used	%	%	Meas	Meas	MnSq
0	1379	1379	30%	30%	-.85	-.85	.9
1	3357	3167	70%	100%	2.41	2.41	1.4

Rasch Analysis of Expert Ratings 24/02/2016 15:40:38
Table 4.1 Unexpected Responses (81 residuals sorted by u).

Cat	Score	Exp.	Resd	StRes	Num	rater	N	Arabic English Fami	Nu	Item
1	1	.0	1.0	9.0	28	Limited Familiarity	2	Limited Familiarity	18	S10 feared
1	1	.0	1.0	9.0	64	Limited Familiarity	2	Limited Familiarity	18	S10 feared
0	0	1.0	-1.0	-9.0	139	Limited Familiarity	2	Limited Familiarity	16	S10 different
0	0	1.0	-1.0	-9.0	149	No Familiarity	1	No Familiarity	22	S12 hands
1	1	.0	1.0	8.8	189	Limited Familiarity	2	Limited Familiarity	8	S6 paid
1	1	.0	1.0	8.1	144	Limited Familiarity	2	Limited Familiarity	18	S10 feared
1	1	.0	1.0	7.7	188	No Familiarity	1	No Familiarity	8	S6 paid
0	0	1.0	-1.0	-7.4	51	Limited Familiarity	2	Limited Familiarity	22	S12 hands
1	1	.0	1.0	7.4	126	No Familiarity	1	No Familiarity	8	S6 paid
0	0	1.0	-1.0	-7.4	175	Very Familiar	4	Very Familiar	22	S12 hands
1	1	.0	1.0	7.1	87	Very Familiar	4	Very Familiar	18	S10 feared
1	1	.0	1.0	7.1	142	Limited Familiarity	2	Limited Familiarity	18	S10 feared
0	0	1.0	-1.0	-6.9	36	Very Familiar	4	Very Familiar	17	S10 children
0	0	1.0	-1.0	-6.8	58	Limited Familiarity	2	Limited Familiarity	10	S6 server
0	0	1.0	-1.0	-6.8	139	Limited Familiarity	2	Limited Familiarity	10	S6 server
1	1	.0	1.0	6.5	44	Some Familiarity	3	Some Familiarity	8	S6 paid
0	0	1.0	-1.0	-6.0	3	Limited Familiarity	2	Limited Familiarity	23	S12 boy
1	1	.0	1.0	6.0	151	Some Familiarity	3	Some Familiarity	6	S3 learned
0	0	1.0	-1.0	-5.8	26	Limited Familiarity	2	Limited Familiarity	22	S12 hands
0	0	1.0	-1.0	-5.8	92	Limited Familiarity	2	Limited Familiarity	22	S12 hands
0	0	1.0	-1.0	-5.8	130	Some Familiarity	3	Some Familiarity	22	S12 hands
1	1	.0	1.0	5.7	79	Some Familiarity	3	Some Familiarity	8	S6 paid
1	1	.0	1.0	5.3	6	Some Familiarity	3	Some Familiarity	18	S10 feared
1	1	.0	1.0	5.2	162	No Familiarity	1	No Familiarity	14	S10 blend
0	0	1.0	-1.0	-5.2	173	No Familiarity	1	No Familiarity	15	S10 rather
0	0	1.0	-1.0	-5.2	173	No Familiarity	1	No Familiarity	16	S10 different
0	0	1.0	-1.0	-5.2	173	No Familiarity	1	No Familiarity	17	S10 children
0	0	1.0	-1.0	-5.1	74	Limited Familiarity	2	Limited Familiarity	11	S6 credit
0	0	1.0	-1.0	-5.1	95	No Familiarity	1	No Familiarity	11	S6 credit
0	0	1.0	-1.0	-5.1	176	Some Familiarity	3	Some Familiarity	11	S6 credit
0	0	1.0	-1.0	-4.7	46	Some Familiarity	3	Some Familiarity	22	S12 hands
0	0	1.0	-1.0	-4.5	125	Some Familiarity	3	Some Familiarity	23	S12 boy
0	0	1.0	-1.0	-4.5	165	No Familiarity	1	No Familiarity	23	S12 boy
0	0	1.0	-1.0	-4.5	181	Very Familiar	4	Very Familiar	24	S12 wore
1	1	.0	1.0	4.4	53	No Familiarity	1	No Familiarity	6	S3 learned
1	1	.0	1.0	4.4	55	No Familiarity	1	No Familiarity	6	S3 learned
1	1	.0	1.0	4.4	131	Some Familiarity	3	Some Familiarity	8	S6 paid
1	1	.0	1.0	4.4	158	Some Familiarity	3	Some Familiarity	8	S6 paid
1	1	.0	1.0	4.4	168	No Familiarity	1	No Familiarity	6	S3 learned
0	0	.9	-.9	-4.2	170	Limited Familiarity	2	Limited Familiarity	11	S6 credit
0	0	.9	-.9	-4.1	104	Limited Familiarity	2	Limited Familiarity	13	S6 usual
0	0	.9	-.9	-4.1	124	Some Familiarity	3	Some Familiarity	12	S6 quite
0	0	.9	-.9	-4.0	25	Limited Familiarity	2	Limited Familiarity	7	S3 law
1	1	.1	.9	4.0	40	No Familiarity	1	No Familiarity	14	S10 blend
0	0	.9	-.9	-4.0	40	No Familiarity	1	No Familiarity	17	S10 children
1	1	.1	.9	4.0	69	Limited Familiarity	2	Limited Familiarity	14	S10 blend
1	1	.1	.9	4.0	184	Limited Familiarity	2	Limited Familiarity	14	S10 blend
0	0	.9	-.9	-3.9	16	No Familiarity	1	No Familiarity	10	S6 server
0	0	.9	-.9	-3.9	50	Limited Familiarity	2	Limited Familiarity	22	S12 hands
0	0	.9	-.9	-3.9	72	No Familiarity	1	No Familiarity	22	S12 hands
0	0	.9	-.9	-3.9	107	Very Familiar	4	Very Familiar	22	S12 hands
0	0	.9	-.9	-3.8	23	Very Familiar	4	Very Familiar	21	S12 involved
0	0	.9	-.9	-3.8	107	Very Familiar	4	Very Familiar	21	S12 involved
0	0	.9	-.9	-3.7	21	Some Familiarity	3	Some Familiarity	24	S12 wore
1	1	.1	.9	3.7	83	Some Familiarity	3	Some Familiarity	18	S10 feared
0	0	.9	-.9	-3.6	120	Some Familiarity	3	Some Familiarity	14	S10 blend
0	0	.9	-.9	-3.6	168	No Familiarity	1	No Familiarity	11	S6 credit
1	1	.1	.9	3.6	186	Limited Familiarity	2	Limited Familiarity	12	S6 quite
1	1	.1	.9	3.5	28	Limited Familiarity	2	Limited Familiarity	14	S10 blend
1	1	.1	.9	3.5	118	Very Familiar	4	Very Familiar	14	S10 blend
1	1	.1	.9	3.5	126	No Familiarity	1	No Familiarity	25	S12 patch
0	0	.9	-.9	-3.5	132	Some Familiarity	3	Some Familiarity	23	S12 boy
0	0	.9	-.9	-3.5	181	Very Familiar	4	Very Familiar	23	S12 boy
0	0	.9	-.9	-3.4	179	No Familiarity	1	No Familiarity	10	S6 server
1	1	.1	.9	3.3	36	Very Familiar	4	Very Familiar	8	S6 paid
0	0	.9	-.9	-3.3	83	Some Familiarity	3	Some Familiarity	13	S6 usual
0	0	.9	-.9	-3.3	112	No Familiarity	1	No Familiarity	22	S12 hands
0	0	.9	-.9	-3.3	123	Limited Familiarity	2	Limited Familiarity	13	S6 usual
0	0	.9	-.9	-3.3	169	No Familiarity	1	No Familiarity	5	S3 jury
0	0	.9	-.9	-3.2	6	Some Familiarity	3	Some Familiarity	21	S12 involved

0	0	.9	-.9	-3.2	30	Very Familiar	4	Very Familiar	19	S10 bridge
0	0	.9	-.9	-3.2	37	Some Familiarity	3	Some Familiarity	25	S12 patch
0	0	.9	-.9	-3.2	49	Some Familiarity	3	Some Familiarity	19	S10 bridge
0	0	.9	-.9	-3.2	89	Limited Familiarity	2	Limited Familiarity	19	S10 bridge
0	0	.9	-.9	-3.2	174	No Familiarity	1	No Familiarity	19	S10 bridge
1	1	.1	.9	3.0	85	Limited Familiarity	2	Limited Familiarity	14	S10 blend
0	0	.9	-.9	-3.0	98	Some Familiarity	3	Some Familiarity	1	S3 tea
1	1	.1	.9	3.0	104	Limited Familiarity	2	Limited Familiarity	18	S10 feared
1	1	.1	.9	3.0	132	Some Familiarity	3	Some Familiarity	18	S10 feared
0	0	.9	-.9	-3.0	159	Limited Familiarity	2	Limited Familiarity	1	S3 tea
0	0	.9	-.9	-3.0	174	No Familiarity	1	No Familiarity	12	S6 quite
Cat	Score	Exp.	Resd	StRes	Num	rater	N	Arabic English Fami	Nu	Item

The Facets analyses output of the Dhivehi-English intelligibility data

```
Facets (Many-Facet Rasch Measurement) Version No. 3.71.1  Copyright �(c) 1987-2013, John M. Linacre. All rights reserved.
24/02/2016 15:43:03

Rasch Analysis of Expert Ratings 24/02/2016 15:43:03
Table 1. Specifications from file "C:\Users\kevin\Desktop\Main Study Intelligibility Operating Files\Dhivehi Intelligibility DE items ONLY Operating File.txt".

Title = Rasch Analysis of Expert Ratings 24/02/2016 15:43:03
Data file = "C:\Users\kevin\Desktop\Intelligibility Data Files\Intelligibility - Dhivehi Fam & English Items ONLY.xlsx"
Output file = C:\Users\kevin\Desktop\Main Study Intelligibility Operating Files\Dhivehi Intelligibility DE items ONLY Operating File.out.txt

; Data specification
Facets = 3
Delements = N
Non-centered = 1
Positive =  1, 2, 3
Labels =
 1,rater ; (elements = 190)
 2,Dhivehi English Familiarity Level, G  ; (elements = 4)
 3,Item ; (elements = 24)
Model = ?,?,?,D,1

; Output description
Arrange tables in order = MN
Bias/Interaction direction = plus ; ability, easiness, leniency: higher score = positive logit
Fair score = Mean
Pt-biserial = Measure
Heading lines in output data files = Y
Inter-rater coefficients reported for facet = 1
Barchart = Yes
Total score for elements = Yes
T3onscreen show only one line on screen iteration report = Y
T4MAX maximum number of unexpected observations reported in Table 4 = 100
T8NBC show table 8 numbers-barcharts-curves = NBC
Unexpected observations reported if standardized residual >= 3
Usort unexpected observations sort order = u
WHexact - Wilson-Hilferty standardization = Y

; Convergence control
Convergence = .1, .01
Iterations (maximum) = 0 ; unlimited
Xtreme scores adjusted by = .3, .5  ;(estimation, bias)

Rasch Analysis of Expert Ratings 24/02/2016 15:43:03
Table 2. Data Summary Report.

Assigning models to Data= "C:\Users\kevin\Desktop\Intelligibility Data Files\Intelligibility - Dhivehi Fam & English Items ONLY.xlsx"
Total lines in data file = 194
Total data lines = 190
Responses matched to model: ?,?,?,D,1 = 4560
    Total non-blank responses found = 4560
              Number of blank lines = 3
Valid responses used for estimation = 4560

Rasch Analysis of Expert Ratings 24/02/2016 15:43:03
Table 3. Iteration Report.

+-----------------------------------------------------------+
| Iteration      Max. Score Residual      Max. Logit Change |
|             Elements    %  Categories   Elements    Steps |
|-----------------------------------------------------------|
| PROX   1     Recount required            3.8712           |
| PROX   2                                 2.3179           |
| JMLE   3  1063.8719  35.8      .0000      .7246    .0000 |
| JMLE   4   425.6169  15.3      .0000      .3478    .0000 |
| JMLE   5   199.2483   7.4      .0000      .1808    .0000 |
| JMLE   6   102.4321   3.9      .0000      .1008    .0000 |
| JMLE   7    57.1370   2.2      .0000      .0665    .0000 |
| JMLE   8    34.4678   1.4      .0000      .0459    .0000 |
| JMLE   9    22.3643    .9      .0000      .0319    .0000 |
| JMLE  10    15.4526    .6      .0000      .0227    .0000 |
| JMLE  11    11.2246    .5      .0000      .0166    .0000 |
| JMLE  12     8.4609    .4      .0000      .0130    .0000 |
| JMLE  13     6.5454    .3      .0000      .0102    .0000 |
| JMLE  14     5.1526    .2      .0000      .0082    .0000 |
| JMLE  15     4.1028    .2      .0000      .0066    .0000 |
| JMLE  16     3.2910    .1      .0000      .0053    .0000 |
| JMLE  17     2.6521    .1      .0000      .0043    .0000 |
| JMLE  18     2.1431    .1      .0000      .0035    .0000 |
| JMLE  19     1.7351    .1      .0000      .0029    .0000 |
| JMLE  20     1.4060    .1      .0000      .0023    .0000 |
| JMLE  21     1.1401    .0      .0000      .0019    .0000 |
| JMLE  22      .9248    .0      .0000      .0016    .0000 |
| JMLE  23      .7502    .0      .0000      .0013    .0000 |
| JMLE  24      .6086    .0      .0000      .0010    .0000 |
| JMLE  25      .4939    .0      .0000      .0008    .0000 |
| JMLE  26      .4006    .0      .0000      .0007    .0000 |
| JMLE  27      .3250    .0      .0000      .0006    .0000 |
| JMLE  28      .2634    .0      .0000      .0004    .0000 |
| JMLE  29      .2139    .0      .0000      .0004    .0000 |
| JMLE  30      .1733    .0      .0000      .0003    .0000 |
| JMLE  31      .1406    .0      .0000      .0002    .0000 |
| JMLE  32      .1140    .0      .0000      .0002    .0000 |
| JMLE  33      .0925    .0      .0000      .0002    .0000 |
+-----------------------------------------------------------+
Warning (6)! There may be 4 disjoint subsets


Rasch Analysis of Expert Ratings 24/02/2016 15:43:03
Table 4. Unexpected Responses - appears after Table 8.


Rasch Analysis of Expert Ratings 24/02/2016 15:43:03
Table 5. Measurable Data Summary.

+--------------------------------------------------+
| Cat  Score  Exp.  Resd StRes|                    |
|-----------------------------+--------------------|
|  .68   .68   .68   .00   .01 | Mean (Count: 4512) |
|  .46   .46   .32   .34   .99 | S.D. (Population)  |
|  .46   .46   .32   .34   .99 | S.D. (Sample)      |
+--------------------------------------------------+
Data log-likelihood chi-square = 3275.4104
Approximate degrees of freedom = 4298
Chi-square significance prob.  = 1.0000
                                       Count   Mean   S.D.   Params
Responses used for estimation      =     4512   0.68   0.46      214
Responses in one extreme score     =       48   1.00   0.00        2
All Responses                      =     4560   0.69   0.46      216
Count of measurable responses      =     4560
Raw-score variance of observations   =   0.22 100.00%
Variance explained by Rasch measures =   0.10  47.10%
Variance of residuals                =   0.11  52.90%

Rasch Analysis of Expert Ratings 24/02/2016 15:43:03
Table 6.0  All Facet Vertical "Rulers".
```

```
Vertical = (1*,2A,3*,S) Yardstick (columns lines low high extreme)= 0,4,-5,6,End
+------------------------------------------------------------------+
|Measr|+rater     |+Dhivehi English Familiarity Level        |+Item|
|-----+-----------+------------------------------------------+-----|
|   6 + .         +                                          +     |
|     |           |                                          |     |
|     |           |                                          |     |
|     | .         |                                          |     |
|   5 +           +                                          +     |
|     | .         |                                          |     |
|     | .         |                                          |     |
|     |           |                                          |     |
|   4 +           +                                          + *   |
|     | **.       |                                          |     |
|     | .         |                                          | *   |
|     | ****.     |                                          | *   |
|   3 + *.        +                                          +     |
|     | ********* |                                          | *   |
|     | .         |                                          |     |
|     | ******.   |                                          |     |
|   2 + ********* +                                          +     |
|     | *         |                                          |     |
|     | ****.     |                                          | **  |
|     | *****.    |                                          | **  |
|   1 + ****.     +                                          + *   |
|     | **.       |                                          |     |
|     | **.       |                                          |     |
|     |           | Some Familiarity     Very Familiar       | **  |
*   0 * *.        *                                          * **  *
|     | *.        | Limited Familiarity                      | *   |
|     | .         | No Familiarity                           |     |
|     | *.        |                                          | *** |
|  -1 + *         +                                          + *   |
|     |           |                                          | *   |
|     | .         |                                          |     |
|     |           |                                          | *   |
|  -2 +           +                                          +     |
|     |           |                                          | *   |
|     |           |                                          |     |
|     |           |                                          |     |
|  -3 +           +                                          +     |
|     |           |                                          |     |
|     |           |                                          | *   |
|     |           |                                          | *   |
|  -4 +           +                                          +     |
|     |           |                                          | *   |
|     |           |                                          |     |
|     |           |                                          |     |
|  -5 +           +                                          +     |
|-----+-----------+------------------------------------------+-----|
|Measr| * = 3     |+Dhivehi English Familiarity Level        | * = |
+------------------------------------------------------------------+

S.1: Model = ?,?,?,D
There are 4 disconnected subsets identified in Table 7.

Rasch Analysis of Expert Ratings 24/02/2016 15:43:03
Table 6.1  rater Facet Summary.

Logit:
                               11 12 2 2  1
                   13 5 15 567145274302741317   21   1  2
 +----+----+----+----+-Q--+--S-+---M+----S----+-Q--+----+----+
-5   -4   -3   -2   -1    0    1    2    3    4    5    6    7

Infit MnSq:
    24532
   8720316111
 +-QS-MS-Q-+----+----+----+----+----+----+----+
 0    1    2    3    4    5    6    7    8    9

Outfit MnSq:
  144221
  4106329151221 22 3111           1     1     1
 +----M----+S---+-Q--+----+----+----+----+----+
 0    1    2    3    4    5    6    7    8    9

Infit ZStd:
             1  1111111111
           2 3333321494446866412   1  1
 +----+----+Q---S----M----S---Q+----+----+----+
-4   -3   -2   -1    0    1    2    3    4    5

Outfit ZStd:
                   22223 1
                 39056719266 343112   1  1
 +----+----+---Q+--S-+-M--S---Q+----+----+----+
-4   -3   -2   -1    0    1    2    3    4    5


Rasch Analysis of Expert Ratings 24/02/2016 15:43:03
Table 6.2  Dhivehi English Familiarity Level Facet Summary.

Logit:
                        11 11
 +----+----+----+----+-QS-M-SQ-+----+----+----+----+----+----+
-5   -4   -3   -2   -1    0    1    2    3    4    5    6    7

Infit MnSq:
     12 1
 +-Q-SMS-Q-+----+----+----+----+----+----+----+
 0    1    2    3    4    5    6    7    8    9

Outfit MnSq:
     2 11
 +-Q-SMS-Q-+----+----+----+----+----+----+----+
 0    1    2    3    4    5    6    7    8    9

Infit ZStd:
                1 1 1           1
 +----+--Q-+---S+----M----+-S--+--Q-+----+----+
-4   -3   -2   -1    0    1    2    3    4    5

Outfit ZStd:
                  1     11 1
 +----+----+----Q---S+-M--S---Q+----+----+----+
-4   -3   -2   -1    0    1    2    3    4    5


Rasch Analysis of Expert Ratings 24/02/2016 15:43:03
Table 6.3  Item Facet Summary.

Logit:
    1  1 1     1 1  1 4  1211 1 121     1 11 1
 +---Q+----+---S+----+----M----+----+S---+----+Q---+----+----+
-5   -4   -3   -2   -1    0    1    2    3    4    5    6    7

Infit MnSq:
      1
     455
```

```
 +---SMS---+----+----+----+----+----+----+----+
 0    1    2    3    4    5    6    7    8    9

Outfit MnSq:
    54553     11
 Q--S-M--S-+Q---+----+----+----+----+----+----+
 0    1    2    3    4    5    6    7    8    9

Infit ZStd:
  1   1      1 1  2  2 6212112 1
 +----+-Q--+--S-+----M----+S---+--Q-+----+----+
-4   -3   -2   -1    0    1    2    3    4    5

Outfit ZStd:
     1   1   21  112112123121              1
 +----Q----+--S-+----M----+--S-+----Q----+----+
-4   -3   -2   -1    0    1    2    3    4    5
```

Rasch Analysis of Expert Ratings 24/02/2016 15:43:03
Table 7.1.1 rater Measurement Report (arranged by MN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	Exact Agree. Obs %	Exact Agree. Exp %	Num rater	
6	24	.25	.19	-1.47	.61	1.33	.9	.94	.4	.73	.49	.58	52.3	52.6	10 No Familiarity	in subset: 1
7	24	.29	.25	-1.11	.58	1.04	.2	1.17	.5	.89	.56	.60	54.4	55.3	75 No Familiarity	in subset: 1
7	24	.29	.25	-1.11	.58	1.07	.3	6.64	2.5	.51	.45	.60	49.6	55.3	138 No Familiarity	in subset: 1
7	24	.29	.25	-1.11	.58	.56	-1.5	.33	-.3	1.52	.73	.60	57.9	55.3	183 No Familiarity	in subset: 1
8	24	.33	.31	-.79	.56	1.03	.1	.90	.2	.97	.59	.61	58.7	57.9	36 No Familiarity	in subset: 1
8	24	.33	.31	-.79	.56	.51	-1.9	.32	-.5	1.62	.76	.61	61.1	57.9	40 No Familiarity	in subset: 1
8	24	.33	.31	-.79	.56	1.02	.1	.79	.1	1.01	.60	.61	58.7	57.9	61 No Familiarity	in subset: 1
8	24	.33	.31	-.79	.56	.94	-.1	.74	.0	1.10	.62	.61	57.8	57.9	121 No Familiarity	in subset: 1
8	24	.33	.31	-.79	.56	1.06	.3	.73	.0	1.00	.60	.61	58.3	57.9	145 No Familiarity	in subset: 1
9	24	.38	.38	-.48	.55	.83	-.5	3.51	2.0	.91	.59	.62	57.9	60.3	85 No Familiarity	in subset: 1
10	24	.42	.45	-.18	.54	1.16	.6	1.44	.7	.67	.54	.63	61.8	62.6	13 No Familiarity	in subset: 1
10	24	.42	.45	-.18	.54	.79	-.7	.57	-.3	1.34	.69	.63	64.2	62.6	118 No Familiarity	in subset: 1
10	24	.42	.45	-.18	.54	.86	-.4	.61	-.3	1.25	.67	.63	63.8	62.6	151 No Familiarity	in subset: 1
10	24	.42	.45	-.18	.54	1.15	.6	2.97	1.9	.54	.51	.63	58.1	62.6	162 No Familiarity	in subset: 1
11	24	.46	.47	-.13	.53	.77	-.8	.54	-.5	1.39	.70	.64	64.1	62.8	142 Limited Familiarity	in subset: 2
11	24	.46	.53	.11	.53	.96	.0	.68	-.2	1.13	.65	.64	64.8	64.6	44 No Familiarity	in subset: 1
11	24	.46	.53	.11	.53	1.03	.2	.76	-.1	1.02	.62	.64	64.5	64.6	79 No Familiarity	in subset: 1
11	24	.46	.53	.11	.53	.77	-.9	.56	-.4	1.38	.70	.64	66.8	64.6	88 No Familiarity	in subset: 1
11	24	.46	.53	.11	.53	.74	-1.0	.52	-.5	1.43	.71	.64	66.9	64.6	108 No Familiarity	in subset: 1
11	24	.46	.53	.11	.53	1.46	1.6	2.67	1.8	.11	.43	.64	57.5	64.6	184 No Familiarity	in subset: 1
12	24	.50	.60	.39	.53	.89	-.3	3.59	2.5	.83	.58	.64	64.7	66.5	14 No Familiarity	in subset: 1
12	24	.50	.60	.39	.53	1.30	1.1	.98	.2	.64	.54	.64	64.2	66.5	47 No Familiarity	in subset: 1
12	24	.50	.60	.39	.53	.86	-.5	.59	-.4	1.29	.68	.64	67.9	66.5	76 No Familiarity	in subset: 1
12	24	.50	.60	.39	.53	1.33	1.3	1.01	.2	.58	.53	.64	63.5	66.5	93 No Familiarity	in subset: 1
12	24	.50	.60	.39	.53	1.12	.5	.82	.0	.90	.60	.64	65.3	66.5	117 No Familiarity	in subset: 1
12	24	.50	.60	.39	.53	.87	-.4	.63	-.3	1.25	.67	.64	67.6	66.5	140 No Familiarity	in subset: 1
15	24	.63	.64	.56	.55	1.89	2.8	4.02	2.4	-.67	.23	.63	64.6	69.2	186 Very Familiar	in subset: 4
13	24	.54	.66	.67	.53	.94	-.1	.73	-.1	1.13	.65	.64	69.0	68.3	3 No Familiarity	in subset: 1
13	24	.54	.66	.67	.53	1.31	1.2	1.02	.2	.60	.54	.64	65.8	68.3	24 No Familiarity	in subset: 1
13	24	.54	.66	.67	.53	1.33	1.3	1.12	.4	.54	.52	.64	64.3	68.3	69 No Familiarity	in subset: 1
13	24	.54	.66	.67	.53	1.11	.5	.91	.0	.86	.59	.64	67.2	68.3	105 No Familiarity	in subset: 1
13	24	.54	.66	.67	.53	.86	-.5	.64	-.3	1.26	.68	.64	69.9	68.3	141 No Familiarity	in subset: 1
13	24	.54	.66	.67	.53	1.17	.7	1.04	.2	.74	.57	.64	65.5	68.3	177 No Familiarity	in subset: 1
14	24	.58	.67	.72	.54	.81	-.7	.55	-.5	1.34	.70	.64	71.1	70.0	43 Limited Familiarity	in subset: 2
14	24	.58	.72	.96	.54	.83	-.6	.59	-.4	1.30	.69	.64	71.7	69.8	19 No Familiarity	in subset: 1
14	24	.58	.72	.96	.54	.89	-.3	.59	-.4	1.25	.68	.64	70.9	69.8	28 No Familiarity	in subset: 1
14	24	.58	.72	.96	.54	1.18	.7	1.78	1.1	.58	.53	.64	68.2	69.8	48 No Familiarity	in subset: 1
14	24	.58	.72	.96	.54	1.19	.8	1.05	.3	.73	.56	.64	68.3	69.8	97 No Familiarity	in subset: 1
14	24	.58	.72	.96	.54	1.42	1.5	1.79	1.1	.23	.45	.64	63.7	69.8	104 No Familiarity	in subset: 1
14	24	.58	.72	.96	.54	1.22	.8	1.06	.3	.69	.55	.64	67.2	69.8	111 No Familiarity	in subset: 1
14	24	.58	.72	.96	.54	1.13	.6	1.07	.3	.76	.57	.64	68.3	69.8	115 No Familiarity	in subset: 1
14	24	.58	.72	.96	.54	.77	-.9	.51	-.5	1.40	.71	.64	72.2	69.8	122 No Familiarity	in subset: 1
14	24	.58	.72	.96	.54	1.09	.4	.78	.0	.95	.61	.64	68.3	69.8	126 No Familiarity	in subset: 1
14	24	.58	.72	.96	.54	1.02	.1	.79	.0	1.01	.62	.64	69.5	69.8	128 No Familiarity	in subset: 1
14	24	.58	.72	.96	.54	.63	-1.6	.42	-.7	1.58	.75	.64	73.6	69.8	165 No Familiarity	in subset: 1
14	24	.58	.72	.96	.54	1.06	.3	.69	-.2	1.03	.63	.64	69.3	69.8	171 No Familiarity	in subset: 1
14	24	.58	.72	.96	.54	1.19	.8	1.45	.7	.62	.54	.64	66.5	69.8	182 No Familiarity	in subset: 1
15	24	.63	.73	1.01	.55	.54	-2.0	.34	-.8	1.67	.77	.63	77.1	72.0	87 Limited Familiarity	in subset: 2
15	24	.63	.78	1.25	.55	1.36	1.3	3.11	2.0	.23	.44	.63	67.2	71.1	15 No Familiarity	in subset: 1
15	24	.63	.78	1.25	.55	1.03	.2	.86	.0	.97	.61	.63	70.7	71.1	33 No Familiarity	in subset: 1
15	24	.63	.78	1.25	.55	.98	.0	1.02	.2	.98	.61	.63	70.3	71.1	53 No Familiarity	in subset: 1
15	24	.63	.78	1.25	.55	.88	-.3	.65	-.2	1.20	.66	.63	72.5	71.1	60 No Familiarity	in subset: 1
15	24	.63	.78	1.25	.55	.80	-.7	.53	-.4	1.33	.69	.63	73.2	71.1	68 No Familiarity	in subset: 1
15	24	.63	.78	1.25	.55	1.24	.9	1.56	.8	.56	.52	.63	67.6	71.1	80 No Familiarity	in subset: 1
15	24	.63	.78	1.25	.55	1.56	1.9	3.58	2.2	-.14	.36	.63	64.2	71.1	90 No Familiarity	in subset: 1
15	24	.63	.78	1.25	.55	1.22	.8	.97	.2	.75	.56	.63	68.7	71.1	98 No Familiarity	in subset: 1
15	24	.63	.78	1.25	.55	.92	-.2	.62	-.2	1.19	.66	.63	71.8	71.1	123 No Familiarity	in subset: 1
15	24	.63	.78	1.25	.55	.92	-.2	.59	-.3	1.20	.66	.63	71.9	71.1	131 No Familiarity	in subset: 1
15	24	.63	.78	1.25	.55	.66	-1.4	.42	-.6	1.52	.74	.63	74.8	71.1	132 No Familiarity	in subset: 1
15	24	.63	.78	1.25	.55	.61	-1.6	.38	-.7	1.58	.75	.63	75.3	71.1	144 No Familiarity	in subset: 1
15	24	.63	.78	1.25	.55	.66	-1.3	.43	-.6	1.51	.73	.63	74.7	71.1	155 No Familiarity	in subset: 1
15	24	.63	.78	1.25	.55	2.15	3.4	7.75	3.9	-1.33	.08	.63	56.5	71.1	158 No Familiarity	in subset: 1
15	24	.63	.78	1.25	.55	1.28	1.0	1.07	.3	.65	.54	.63	68.3	71.1	180 No Familiarity	in subset: 1
16	24	.67	.79	1.32	.56	1.67	2.0	1.82	1.0	.07	.38	.63	65.9	73.7	99 Limited Familiarity	in subset: 2
16	24	.67	.79	1.32	.56	.93	-.1	.60	-.2	1.16	.65	.63	74.7	73.7	147 Limited Familiarity	in subset: 2
16	24	.67	.83	1.56	.56	1.49	1.6	1.89	1.0	.28	.43	.63	66.3	72.3	9 No Familiarity	in subset: 1
16	24	.67	.83	1.56	.56	.98	.0	.86	.1	1.03	.62	.63	71.9	72.3	12 No Familiarity	in subset: 1
16	24	.67	.83	1.56	.56	.88	-.3	.56	-.2	1.22	.67	.63	73.5	72.3	54 No Familiarity	in subset: 1
16	24	.67	.83	1.56	.56	1.16	.6	.73	.0	.92	.59	.63	70.8	72.3	55 No Familiarity	in subset: 1
16	24	.67	.83	1.56	.56	1.21	.7	1.12	.4	.73	.55	.63	70.2	72.3	65 No Familiarity	in subset: 1
16	24	.67	.83	1.56	.56	.91	-.2	.63	-.1	1.17	.65	.63	73.3	72.3	70 No Familiarity	in subset: 1
16	24	.67	.83	1.56	.56	.72	-1.0	.48	-.4	1.39	.71	.63	75.1	72.3	74 No Familiarity	in subset: 1
16	24	.67	.83	1.56	.56	1.22	.8	2.94	1.7	.42	.47	.63	69.0	72.3	77 No Familiarity	in subset: 1
16	24	.67	.83	1.56	.56	1.07	.3	1.05	.3	.89	.59	.63	71.6	72.3	110 No Familiarity	in subset: 1
16	24	.67	.83	1.56	.56	.97	.0	.67	-.1	1.10	.64	.63	72.5	72.3	134 No Familiarity	in subset: 1
16	24	.67	.83	1.56	.56	1.21	.8	1.14	.4	.71	.54	.63	69.4	72.3	136 No Familiarity	in subset: 1
16	24	.67	.83	1.56	.56	.75	-.8	.48	-.4	1.36	.70	.63	74.9	72.3	146 No Familiarity	in subset: 1
16	24	.67	.83	1.56	.56	.79	-.7	.54	-.3	1.31	.69	.63	74.6	72.3	174 No Familiarity	in subset: 1
18	24	.75	.83	1.56	.62	.87	-.2	.53	.0	1.19	.65	.61	71.9	77.2	143 Very Familiar	in subset: 4
17	24	.71	.84	1.65	.59	.56	-1.5	.33	-.5	1.52	.75	.62	81.3	75.2	49 Limited Familiarity	in subset: 2
17	24	.71	.84	1.65	.59	.94	.0	.68	.0	1.10	.63	.62	76.0	75.2	120 Limited Familiarity	in subset: 2
17	24	.71	.84	1.65	.59	1.40	1.2	1.46	.7	.48	.46	.62	71.4	75.2	150 Limited Familiarity	in subset: 2
17	24	.71	.87	1.89	.59	1.36	1.1	1.46	.7	.54	.48	.62	69.2	73.1	5 No Familiarity	in subset: 1
17	24	.71	.87	1.89	.59	.69	-1.0	.50	-.2	1.36	.70	.62	76.1	73.1	11 No Familiarity	in subset: 1
17	24	.71	.87	1.89	.59	1.49	1.4	1.77	.9	.36	.43	.62	67.7	73.1	16 No Familiarity	in subset: 1
17	24	.71	.87	1.89	.59	1.13	.5	1.24	.5	.82	.56	.62	71.7	73.1	18 No Familiarity	in subset: 1
17	24	.71	.87	1.89	.59	.56	-1.5	.33	-.5	1.52	.75	.62	77.6	73.1	34 No Familiarity	in subset: 1
17	24	.71	.87	1.89	.59	.79	-.6	.73	.0	1.21	.66	.62	75.1	73.1	35 No Familiarity	in subset: 1
17	24	.71	.87	1.89	.59	1.10	.4	.78	.1	.94	.59	.62	72.1	73.1	38 No Familiarity	in subset: 1
17	24	.71	.87	1.89	.59	1.09	.3	1.00	.3	.88	.57	.62	72.0	73.1	39 No Familiarity	in subset: 1
17	24	.71	.87	1.89	.59	.63	-1.2	.39	-.4	1.45	.73	.62	76.9	73.1	51 No Familiarity	in subset: 1
17	24	.71	.87	1.89	.59	1.24	.8	1.13	.4	.71	.53	.62	70.5	73.1	52 No Familiarity	in subset: 1
17	24	.71	.87	1.89	.59	1.23	.7	1.00	.3	.76	.54	.62	70.6	73.1	64 No Familiarity	in subset: 1
17	24	.71	.87	1.89	.59	.78	-.6	.47	-.2	1.31	.69	.62	75.5	73.1	66 No Familiarity	in subset: 1
17	24	.71	.87	1.89	.59	1.24	.8	.87	.2	.81	.55	.62	70.9	73.1	91 No Familiarity	in subset: 1
17	24	.71	.87	1.89	.59	.76	-.7	1.10	.4	1.16	.65	.62	75.1	73.1	96 No Familiarity	in subset: 1
17	24	.71	.87	1.89	.59	.73	-.8	.48	-.2	1.34	.70	.62	75.9	73.1	101 No Familiarity	in subset: 1
17	24	.71	.87	1.89	.59	1.31	.9	1.33	.6	.64	.51	.62	69.8	73.1	127 No Familiarity	in subset: 1
17	24	.71	.87	1.89	.59	.93	-.1	.53	-.2	1.17	.65	.62	74.1	73.1	153 No Familiarity	in subset: 1
17	24	.71	.87	1.89	.59	.66	-1.1	.44	-.3	1.40	.72	.62	76.5	73.1	154 No Familiarity	in subset: 1
17	24	.71	.87	1.89	.59	.81	-.5	.51	-.2	1.27	.68	.62	75.1	73.1	164 No Familiarity	in subset: 1
17	24	.71	.87	1.89	.59	1.08	.3	3.16	1.6	.68	.52	.62	71.6	73.1	166 No Familiarity	in subset: 1
17	24	.71	.87	1.89	.59	.77	-.7	.44	-.3	1.33	.70	.62	75.7	73.1	168 No Familiarity	in subset: 1
17	24	.71	.87	1.89	.59	.77	-.6	.56	-.1	1.27	.68	.62	75.4	73.1	173 No Familiarity	in subset: 1

17	24	.71	.87	1.89	.59	1.18	.6	1.37	.6	.73	.53	.62	71.1	73.1	176 No Familiarity	in subset: 1
17	24	.71	.87	1.89	.59	1.20	.6	.81	.1	.86	.57	.62	71.3	73.1	181 No Familiarity	in subset: 1
18	24	.75	.88	2.01	.62	.56	-1.3	.31	-.3	1.46	.73	.61	81.0	76.3	58 Limited Familiarity	in subset: 2
18	24	.75	.88	2.01	.62	.66	-.9	.48	-.1	1.33	.69	.61	80.5	76.3	137 Limited Familiarity	in subset: 2
18	24	.75	.88	2.01	.62	1.15	.5	.95	.3	.87	.55	.61	75.8	76.3	159 Limited Familiarity	in subset: 2
18	24	.75	.90	2.25	.62	1.44	1.1	.95	.3	.67	.49	.61	70.4	73.8	1 No Familiarity	in subset: 1
18	24	.75	.90	2.25	.62	.66	-.9	.37	-.2	1.38	.71	.61	77.0	73.8	8 No Familiarity	in subset: 1
18	24	.75	.90	2.25	.62	.50	-1.6	.27	-.4	1.52	.75	.61	78.6	73.8	21 No Familiarity	in subset: 1
18	24	.75	.90	2.25	.62	1.10	.4	1.05	.4	.88	.56	.61	72.6	73.8	23 No Familiarity	in subset: 1
18	24	.75	.90	2.25	.62	.66	-.9	.37	-.2	1.38	.71	.61	77.0	73.8	29 No Familiarity	in subset: 1
18	24	.75	.90	2.25	.62	.82	-.4	1.23	.5	1.07	.62	.61	74.7	73.8	32 No Familiarity	in subset: 1
18	24	.75	.90	2.25	.62	1.38	1.0	1.01	.4	.68	.50	.61	70.6	73.8	37 No Familiarity	in subset: 1
18	24	.75	.90	2.25	.62	1.13	.4	1.44	.7	.79	.53	.61	71.8	73.8	59 No Familiarity	in subset: 1
18	24	.75	.90	2.25	.62	1.22	.6	.92	.3	.82	.54	.61	71.9	73.8	73 No Familiarity	in subset: 1
18	24	.75	.90	2.25	.62	.83	-.3	.55	.0	1.20	.65	.61	75.4	73.8	82 No Familiarity	in subset: 1
18	24	.75	.90	2.25	.62	.99	.0	.85	.2	1.01	.60	.61	74.1	73.8	92 No Familiarity	in subset: 1
18	24	.75	.90	2.25	.62	.99	.0	1.47	.7	.91	.57	.61	73.3	73.8	107 No Familiarity	in subset: 1
18	24	.75	.90	2.25	.62	.50	-1.6	.27	-.4	1.52	.75	.61	78.6	73.8	116 No Familiarity	in subset: 1
18	24	.75	.90	2.25	.62	1.58	1.5	2.11	1.0	.30	.38	.61	67.7	73.8	119 No Familiarity	in subset: 1
18	24	.75	.90	2.25	.62	1.05	.2	.64	.0	1.03	.60	.61	73.4	73.8	133 No Familiarity	in subset: 1
18	24	.75	.90	2.25	.62	.66	-.9	.37	-.2	1.37	.71	.61	77.0	73.8	135 No Familiarity	in subset: 1
18	24	.75	.90	2.25	.62	.74	-.6	.43	-.1	1.30	.69	.61	76.3	73.8	139 No Familiarity	in subset: 1
18	24	.75	.90	2.25	.62	1.11	.4	.88	.3	.90	.56	.61	72.6	73.8	148 No Familiarity	in subset: 1
18	24	.75	.90	2.25	.62	.69	-.8	.59	.0	1.28	.68	.61	76.4	73.8	152 No Familiarity	in subset: 1
18	24	.75	.90	2.25	.62	1.06	.2	.70	.1	1.00	.59	.61	73.2	73.8	178 No Familiarity	in subset: 1
20	24	.83	.92	2.38	.71	.71	-.5	.76	.5	1.15	.61	.56	.0	.0	125 Some Familiarity	in subset: 3
19	24	.79	.92	2.42	.66	.94	.0	.73	.3	1.05	.59	.59	75.5	77.1	163 Limited Familiarity	in subset: 2
19	24	.79	.93	2.66	.66	.60	-1.0	.34	.0	1.36	.70	.59	77.5	74.0	2 No Familiarity	in subset: 1
19	24	.79	.93	2.66	.66	1.00	.1	.85	.4	.98	.57	.59	73.3	74.0	6 No Familiarity	in subset: 1
19	24	.79	.93	2.66	.66	1.57	1.3	1.39	.7	.49	.41	.59	69.1	74.0	17 No Familiarity	in subset: 1
19	24	.79	.93	2.66	.66	1.41	1.0	2.40	1.1	.52	.42	.59	70.1	74.0	22 No Familiarity	in subset: 1
19	24	.79	.93	2.66	.66	1.31	.8	2.22	1.1	.60	.44	.59	71.0	74.0	26 No Familiarity	in subset: 1
19	24	.79	.93	2.66	.66	.87	-.1	.66	.2	1.11	.62	.59	74.7	74.0	27 No Familiarity	in subset: 1
19	24	.79	.93	2.66	.66	.45	-1.5	.23	-.2	1.49	.74	.59	78.8	74.0	30 No Familiarity	in subset: 1
19	24	.79	.93	2.66	.66	1.40	1.0	3.80	1.6	.32	.35	.59	68.5	74.0	31 No Familiarity	in subset: 1
19	24	.79	.93	2.66	.66	.45	-1.5	.23	-.2	1.49	.74	.59	78.8	74.0	41 No Familiarity	in subset: 1
19	24	.79	.93	2.66	.66	1.36	.9	.82	.3	.78	.50	.59	71.7	74.0	46 No Familiarity	in subset: 1
19	24	.79	.93	2.66	.66	1.51	1.2	1.29	.6	.56	.43	.59	69.1	74.0	57 No Familiarity	in subset: 1
19	24	.79	.93	2.66	.66	.92	.0	.64	.2	1.09	.61	.59	74.4	74.0	67 No Familiarity	in subset: 1
19	24	.79	.93	2.66	.66	.92	.0	.64	.2	1.08	.60	.59	74.4	74.0	71 No Familiarity	in subset: 1
19	24	.79	.93	2.66	.66	.60	-1.0	.34	.0	1.36	.70	.59	77.5	74.0	81 No Familiarity	in subset: 1
19	24	.79	.93	2.66	.66	.45	-1.5	.23	-.2	1.49	.74	.59	78.8	74.0	83 No Familiarity	in subset: 1
19	24	.79	.93	2.66	.66	.78	-.4	.43	.0	1.23	.66	.59	76.0	74.0	89 No Familiarity	in subset: 1
19	24	.79	.93	2.66	.66	.54	-1.2	.28	-.1	1.42	.72	.59	78.0	74.0	95 No Familiarity	in subset: 1
19	24	.79	.93	2.66	.66	1.18	.5	.73	.3	.91	.55	.59	73.1	74.0	103 No Familiarity	in subset: 1
19	24	.79	.93	2.66	.66	.70	-.6	.38	.0	1.30	.68	.59	76.9	74.0	106 No Familiarity	in subset: 1
19	24	.79	.93	2.66	.66	1.09	.3	2.37	1.1	.72	.48	.59	71.5	74.0	109 No Familiarity	in subset: 1
19	24	.79	.93	2.66	.66	.94	.0	.50	.0	1.13	.62	.59	75.1	74.0	112 No Familiarity	in subset: 1
19	24	.79	.93	2.66	.66	.94	.0	.72	.3	1.05	.60	.59	74.2	74.0	113 No Familiarity	in subset: 1
19	24	.79	.93	2.66	.66	.79	-.4	.44	.0	1.23	.65	.59	75.8	74.0	129 No Familiarity	in subset: 1
19	24	.79	.93	2.66	.66	.95	.0	.74	.3	1.04	.59	.59	74.0	74.0	157 No Familiarity	in subset: 1
19	24	.79	.93	2.66	.66	.79	-.4	.44	.0	1.23	.65	.59	75.8	74.0	170 No Familiarity	in subset: 1
19	24	.79	.93	2.66	.66	.70	-.6	.39	.0	1.29	.67	.59	76.7	74.0	172 No Familiarity	in subset: 1
19	24	.79	.93	2.66	.66	.89	-.1	.83	.3	1.06	.60	.59	74.3	74.0	179 No Familiarity	in subset: 1
20	24	.83	.95	2.89	.71	.90	.0	.40	.2	1.16	.61	.56	79.9	77.3	42 Limited Familiarity	in subset: 2
20	24	.83	.95	2.89	.71	1.30	.7	1.25	.7	.73	.46	.56	74.7	77.3	78 Limited Familiarity	in subset: 2
20	24	.83	.95	2.89	.71	1.37	.8	1.91	1.0	.56	.39	.56	72.1	77.3	100 Limited Familiarity	in subset: 2
20	24	.83	.95	2.89	.71	.41	-1.5	.18	.0	1.47	.72	.56	82.6	77.3	161 Limited Familiarity	in subset: 2
21	24	.88	.95	3.00	.79	1.68	1.3	1.07	.8	.55	.34	.52	77.1	81.3	189 Very Familiar	in subset: 4
20	24	.83	.96	3.13	.71	.64	-.7	.40	.2	1.27	.65	.56	76.3	73.8	45 No Familiarity	in subset: 1
20	24	.83	.96	3.13	.71	.98	.1	.47	.2	1.09	.58	.56	74.8	73.8	50 No Familiarity	in subset: 1
20	24	.83	.96	3.13	.71	.64	-.7	.40	.2	1.27	.65	.56	76.3	73.8	56 No Familiarity	in subset: 1
20	24	.83	.96	3.13	.71	.41	-1.5	.18	.0	1.47	.72	.56	78.3	73.8	84 No Familiarity	in subset: 1
20	24	.83	.96	3.13	.71	1.25	.6	9.00	3.3	.33	.31	.56	70.5	73.8	86 No Familiarity	in subset: 1
20	24	.83	.96	3.13	.71	.98	.1	.47	.2	1.09	.58	.56	74.8	73.8	94 No Familiarity	in subset: 1
20	24	.83	.96	3.13	.71	1.05	.2	.59	.3	1.02	.56	.56	74.1	73.8	124 No Familiarity	in subset: 1
20	24	.83	.96	3.13	.71	.65	-.7	.43	.2	1.26	.64	.56	76.1	73.8	130 No Familiarity	in subset: 1
20	24	.83	.96	3.13	.71	.90	.0	.40	.2	1.16	.61	.56	75.7	73.8	149 No Familiarity	in subset: 1
20	24	.83	.96	3.13	.71	.59	-.9	.30	.1	1.34	.67	.56	76.8	73.8	160 No Familiarity	in subset: 1
20	24	.83	.96	3.13	.71	.82	-.2	.35	.1	1.22	.63	.56	75.9	73.8	167 No Familiarity	in subset: 1
20	24	.83	.96	3.13	.71	1.08	.3	.52	.3	1.02	.56	.56	74.2	73.8	169 No Familiarity	in subset: 1
20	24	.83	.96	3.13	.71	1.94	1.7	1.18	.7	.37	.33	.56	68.1	73.8	185 No Familiarity	in subset: 1
21	24	.88	.97	3.45	.79	.82	-.2	.29	.4	1.21	.58	.52	79.7	76.9	7 Limited Familiarity	in subset: 2
21	24	.88	.98	3.69	.79	.91	.0	.64	.6	1.05	.53	.52	73.1	73.1	4 No Familiarity	in subset: 1
21	24	.88	.98	3.69	.79	.82	-.2	.29	.4	1.21	.58	.52	75.1	73.1	62 No Familiarity	in subset: 1
21	24	.88	.98	3.69	.79	.79	-.2	.33	.4	1.20	.58	.52	74.7	73.1	63 No Familiarity	in subset: 1
21	24	.88	.98	3.69	.79	1.11	.3	4.24	1.5	.65	.38	.52	70.7	73.1	72 No Familiarity	in subset: 1
21	24	.88	.98	3.69	.79	.91	.0	.64	.6	1.05	.53	.52	73.1	73.1	102 No Familiarity	in subset: 1
21	24	.88	.98	3.69	.79	.34	-1.6	.13	.2	1.48	.68	.52	77.1	73.1	156 No Familiarity	in subset: 1
22	24	.92	.98	3.69	.89	1.36	.8	1.17	1.2	.69	.33	.45	82.3	81.3	188 Very Familiar	in subset: 4
22	24	.92	.99	4.38	.89	.50	-1.0	.13	.7	1.37	.57	.45	74.1	71.7	25 No Familiarity	in subset: 1
22	24	.92	.99	4.38	.89	1.20	.5	.40	.9	.96	.42	.45	71.6	71.7	114 No Familiarity	in subset: 1
23	24	.96	.99	4.68	1.13	.93	.1	.17	1.7	1.13	.37	.33	83.3	80.4	190 Very Familiar	in subset: 4
23	24	.96	1.00	5.37	1.13	.93	.1	.17	1.7	1.13	.37	.33	70.4	69.7	20 No Familiarity	in subset: 1
24	24	1.00	1.00	(6.07	1.90)	Maximum					.00	.00	.0	.0	175 Very Familiar	in subset: 4
24	24	1.00	1.00	(6.07	1.90)	Maximum					.00	.00	.0	.0	187 Very Familiar	in subset: 4
16.5	24.0	.69	.81	1.83	.63	.98	.0	1.03	.3		.58				Mean (Count: 190)	
3.5	.0	.15	.18	1.25	.16	.31	.9	1.16	.7		.13				S.D. (Population)	
3.5	.0	.15	.18	1.26	.16	.31	.9	1.16	.7		.13				S.D. (Sample)	

With extremes, Model, Populn: RMSE .65 Adj (True) S.D. 1.07 Separation 1.65 Strata 2.53 Reliability (not inter-rater) .73
With extremes, Model, Sample: RMSE .65 Adj (True) S.D. 1.08 Separation 1.66 Strata 2.54 Reliability (not inter-rater) .73
Without extremes, Model, Populn: RMSE .62 Adj (True) S.D. 1.00 Separation 1.61 Strata 2.48 Reliability (not inter-rater) .72
Without extremes, Model, Sample: RMSE .62 Adj (True) S.D. 1.01 Separation 1.62 Strata 2.49 Reliability (not inter-rater) .72
With extremes, Model, Fixed (all same) chi-square: 649.8 d.f.: 189 significance (probability): .00
With extremes, Model, Random (normal) chi-square: 148.9 d.f.: 188 significance (probability): .98
Inter-Rater agreement opportunities: 328224 Exact agreements: 232688 = 70.9% Expected: 232992.1 = 71.0%

Rasch Analysis of Expert Ratings 24/02/2016 15:43:03
Table 7.2.1 Dhivehi English Familiarity Level Measurement Report (arranged by MN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	N Dhivehi English Familiarity Level	
2670	3960	.67	.80	-.42	.05	.99	-.2	1.12	1.1	1.00	.64	.64	1 No Familiarity	in subset: 1
297	408	.73	.83	-.18	.15	.92	-.9	.75	-.6	1.09	.64	.62	2 Limited Familiarity	in subset: 2
147	168	.88	.89	.27	.32	1.43	2.1	1.39	.7	.57	.47	.54	4 Very Familiar	in subset: 4
20	24	.83	.89	.33	.71	.71	-.5	.76	.5	1.15	.61	.56	3 Some Familiarity	in subset: 3
783.5	1140.0	.78	.85	.00	.31	1.01	.1	1.01	.4		.59		Mean (Count: 4)	
1093.6	1633.9	.08	.04	.31	.25	.26	1.2	.27	.7		.07		S.D. (Population)	
1262.8	1886.7	.09	.05	.36	.29	.30	1.4	.31	.8		.08		S.D. (Sample)	

Model, Populn: RMSE .40 Adj (True) S.D. .00 Separation .00 Strata .33 Reliability .00
Model, Sample: RMSE .40 Adj (True) S.D. .00 Separation .00 Strata .33 Reliability .00
Model, Fixed (all same) chi-square: 7.6 d.f.: 3 significance (probability): .06
Model, Random (normal) chi-square: 1.9 d.f.: 2 significance (probability): .38

Rasch Analysis of Expert Ratings 24/02/2016 15:43:03
Table 7.3.1 Item Measurement Report (arranged by MN).

Total Score	Total Count	Obsvd Average	Fair(M) Average	Measure	Model S.E.	Infit MnSq	Infit ZStd	Outfit MnSq	Outfit ZStd	Estim. Discrm	Correlation PtMea	Correlation PtExp	Nu Item
18	190	.09	.07	-4.30	.28	.96	-.1	1.32	.8	.98	.35	.36	4 S2 cat
25	190	.13	.11	-3.84	.24	1.14	.8	.91	-.1	.92	.33	.38	12 S4 bath
33	190	.17	.16	-3.43	.21	1.10	.7	2.74	4.3	.74	.24	.40	10 S4 passenger
70	190	.37	.41	-2.15	.17	1.02	.3	.95	-.3	.98	.44	.45	21 S13 chief
86	190	.45	.52	-1.71	.16	.91	-1.3	.83	-1.5	1.26	.52	.46	1 S2 judge
103	190	.54	.63	-1.25	.17	1.09	1.3	1.11	.9	.77	.40	.46	6 S2 stick
116	190	.61	.71	-.89	.17	.76	-3.7	.65	-3.2	1.56	.62	.46	17 S8 girl
118	190	.62	.72	-.83	.17	1.01	.1	.94	-.4	1.01	.46	.46	23 S13 managed
120	190	.63	.73	-.77	.17	1.07	.9	1.16	1.2	.83	.40	.45	15 S8 theatre
121	190	.64	.74	-.74	.17	1.14	1.8	1.16	1.2	.71	.36	.45	7 S4 creature
136	190	.72	.82	-.27	.18	.85	-1.7	.75	-1.5	1.24	.54	.44	14 S8 challenged
144	190	.76	.86	.00	.19	.72	-2.9	.58	-2.4	1.37	.61	.43	13 S8 hag
144	190	.76	.86	.00	.19	1.15	1.4	1.11	.6	.82	.34	.43	24 S13 bulls
151	190	.79	.89	.27	.20	.90	-.8	.85	-.6	1.11	.48	.41	2 S2 laughed
153	190	.81	.90	.36	.21	1.05	.4	1.07	.3	.93	.37	.41	5 S2 parted
164	190	.86	.94	.88	.23	1.03	.2	.76	-.7	1.02	.38	.37	16 S8 huge
169	190	.89	.95	1.18	.25	.86	-.8	.52	-1.4	1.14	.46	.35	18 S8 shouting
171	190	.90	.96	1.31	.26	.95	-.1	1.21	.6	1.01	.35	.34	11 S4 stood
173	190	.91	.96	1.46	.28	1.25	1.2	1.35	.8	.83	.18	.32	20 S13 looked
174	190	.92	.97	1.53	.28	1.10	.5	1.31	.7	.91	.24	.32	3 S2 mother
184	190	.97	.99	2.70	.43	1.01	.1	.98	.2	.97	.18	.21	8 S4 travelled
186	190	.98	.99	3.15	.52	.98	.1	.65	.0	1.02	.20	.18	22 S13 farmer
187	190	.98	.99	3.45	.59	1.01	.2	2.69	1.3	.97	.11	.16	19 S13 father
188	190	.99	1.00	3.88	.72	.95	.1	.64	.2	1.03	.16	.13	9 S4 quietly
130.6	190.0	.69	.75	.00	.27	1.00	.0	1.09	.0		.36		Mean (Count: 24)
50.8	.0	.27	.28	2.14	.14	.12	1.3	.54	1.5		.14		S.D. (Population)
51.9	.0	.27	.29	2.18	.15	.13	1.3	.55	1.5		.14		S.D. (Sample)

Model, Populn: RMSE .31 Adj (True) S.D. 2.11 Separation 6.91 Strata 9.55 Reliability .98
Model, Sample: RMSE .31 Adj (True) S.D. 2.16 Separation 7.07 Strata 9.75 Reliability .98
Model, Fixed (all same) chi-square: 1196.6 d.f.: 23 significance (probability): .00
Model, Random (normal) chi-square: 22.2 d.f.: 22 significance (probability): .45

Rasch Analysis of Expert Ratings 24/02/2016 15:43:03
Table 8.1 Category Statistics.

Model = ?,?,?,D

Score	DATA Category Counts Total	Used	%	Cum. %	QUALITY CONTROL Avge Meas	Exp. Meas	OUTFIT MnSq
0	1426	1426	32%	32%	-.89	-.89	1.0
1	3134	3086	68%	100%	2.48	2.48	1.2

Rasch Analysis of Expert Ratings 24/02/2016 15:43:03
Table 4.1 Unexpected Responses (77 residuals sorted by u).

Cat	Score	Exp.	Resd	StRes	Num	rater	N	Dhivehi English Fam	Nu	Item
0	0	1.0	-1.0	-9.0	72	No Familiarity	1	No Familiarity	11	S4 stood
0	0	1.0	-1.0	-9.0	86	No Familiarity	1	No Familiarity	19	S13 father
1	1	.0	1.0	9.0	138	No Familiarity	1	No Familiarity	10	S4 passenger
0	0	1.0	-1.0	-9.0	158	No Familiarity	1	No Familiarity	9	S4 quietly
1	1	.0	1.0	8.7	14	No Familiarity	1	No Familiarity	4	S2 cat
1	1	.0	1.0	8.7	85	No Familiarity	1	No Familiarity	10	S4 passenger
0	0	1.0	-1.0	-8.0	166	No Familiarity	1	No Familiarity	8	S4 travelled
1	1	.0	1.0	7.5	162	No Familiarity	1	No Familiarity	10	S4 passenger
0	0	1.0	-1.0	-7.3	15	No Familiarity	1	No Familiarity	22	S13 farmer
0	0	1.0	-1.0	-7.3	90	No Familiarity	1	No Familiarity	22	S13 farmer
0	0	1.0	-1.0	-6.8	77	No Familiarity	1	No Familiarity	8	S4 travelled
0	0	1.0	-1.0	-6.6	22	No Familiarity	1	No Familiarity	3	S2 mother
0	0	1.0	-1.0	-6.6	31	No Familiarity	1	No Familiarity	3	S2 mother
0	0	1.0	-1.0	-6.6	109	No Familiarity	1	No Familiarity	3	S2 mother
1	1	.0	1.0	6.5	184	No Familiarity	1	No Familiarity	10	S4 passenger
0	0	1.0	-1.0	-6.3	26	No Familiarity	1	No Familiarity	20	S13 looked
0	0	1.0	-1.0	-6.3	31	No Familiarity	1	No Familiarity	20	S13 looked
0	0	1.0	-1.0	-5.8	186	Very Familiar	4	Very Familiar	8	S4 travelled
1	1	.0	1.0	5.7	158	No Familiarity	1	No Familiarity	4	S2 cat
1	1	.0	1.0	5.7	186	Very Familiar	4	Very Familiar	4	S2 cat
0	0	1.0	-1.0	-5.2	107	No Familiarity	1	No Familiarity	20	S13 looked
0	0	1.0	-1.0	-5.2	119	No Familiarity	1	No Familiarity	20	S13 looked
0	0	1.0	-1.0	-5.0	48	No Familiarity	1	No Familiarity	8	S4 travelled
1	1	.0	1.0	4.9	9	No Familiarity	1	No Familiarity	4	S2 cat
0	0	1.0	-1.0	-4.9	188	Very Familiar	4	Very Familiar	15	S8 theatre
0	0	1.0	-1.0	-4.8	32	No Familiarity	1	No Familiarity	11	S4 stood
0	0	1.0	-1.0	-4.8	59	No Familiarity	1	No Familiarity	11	S4 stood
0	0	1.0	-1.0	-4.6	100	Limited Familiarity	2	Limited Familiarity	5	S2 parted
0	0	1.0	-1.0	-4.5	16	No Familiarity	1	No Familiarity	3	S2 mother
1	1	.0	1.0	4.5	80	No Familiarity	1	No Familiarity	12	S4 bath
0	0	1.0	-1.0	-4.5	96	No Familiarity	1	No Familiarity	3	S2 mother
0	0	1.0	-1.0	-4.4	78	Limited Familiarity	2	Limited Familiarity	2	S2 laughed
0	0	1.0	-1.0	-4.4	100	Limited Familiarity	2	Limited Familiarity	2	S2 laughed
0	0	.9	-.9	-4.3	18	No Familiarity	1	No Familiarity	20	S13 looked
0	0	.9	-.9	-4.3	176	No Familiarity	1	No Familiarity	20	S13 looked
0	0	.9	-.9	-4.2	13	No Familiarity	1	No Familiarity	19	S13 father
1	1	.1	.9	4.2	104	No Familiarity	1	No Familiarity	10	S4 passenger
1	1	.1	.9	4.2	182	No Familiarity	1	No Familiarity	10	S4 passenger
1	1	.1	.9	4.1	127	No Familiarity	1	No Familiarity	4	S2 cat
1	1	.1	.9	4.0	75	No Familiarity	1	No Familiarity	6	S2 stick
0	0	.9	-.9	-3.9	23	No Familiarity	1	No Familiarity	16	S8 huge
1	1	.1	.9	3.9	77	No Familiarity	1	No Familiarity	12	S4 bath
0	0	.9	-.9	-3.9	125	Some Familiarity	3	Some Familiarity	24	S13 bulls
0	0	.9	-.9	-3.8	5	No Familiarity	1	No Familiarity	18	S8 shouting
0	0	.9	-.9	-3.8	99	Limited Familiarity	2	Limited Familiarity	3	S2 mother
0	0	.9	-.9	-3.7	17	No Familiarity	1	No Familiarity	5	S2 parted
1	1	.1	.9	3.7	53	No Familiarity	1	No Familiarity	10	S4 passenger
0	0	.9	-.9	-3.7	57	No Familiarity	1	No Familiarity	5	S2 parted
1	1	.1	.9	3.7	90	No Familiarity	1	No Familiarity	10	S4 passenger
0	0	.9	-.9	-3.7	110	No Familiarity	1	No Familiarity	20	S13 looked
1	1	.1	.9	3.7	158	No Familiarity	1	No Familiarity	10	S4 passenger
0	0	.9	-.9	-3.7	179	No Familiarity	1	No Familiarity	5	S2 parted
1	1	.1	.9	3.7	186	Very Familiar	4	Very Familiar	10	S4 passenger
0	0	.9	-.9	-3.5	4	No Familiarity	1	No Familiarity	15	S8 theatre
0	0	.9	-.9	-3.5	102	No Familiarity	1	No Familiarity	15	S8 theatre
0	0	.9	-.9	-3.5	189	Very Familiar	4	Very Familiar	7	S4 creature
0	0	.9	-.9	-3.4	65	No Familiarity	1	No Familiarity	11	S4 stood
1	1	.1	.9	3.4	92	No Familiarity	1	No Familiarity	4	S2 cat
1	1	.1	.9	3.4	159	Limited Familiarity	2	Limited Familiarity	4	S2 cat
1	1	.1	.9	3.3	39	No Familiarity	1	No Familiarity	12	S4 bath
1	1	.1	.9	3.3	52	No Familiarity	1	No Familiarity	12	S4 bath
1	1	.1	.9	3.3	150	Limited Familiarity	2	Limited Familiarity	12	S4 bath
0	0	.9	-.9	-3.3	158	No Familiarity	1	No Familiarity	3	S2 mother
0	0	.9	-.9	-3.2	5	No Familiarity	1	No Familiarity	16	S8 huge
0	0	.9	-.9	-3.2	35	No Familiarity	1	No Familiarity	16	S8 huge
0	0	.9	-.9	-3.2	150	Limited Familiarity	2	Limited Familiarity	16	S8 huge
0	0	.9	-.9	-3.1	6	No Familiarity	1	No Familiarity	13	S8 hag
1	1	.1	.9	3.1	12	No Familiarity	1	No Familiarity	10	S4 passenger
0	0	.9	-.9	-3.1	27	No Familiarity	1	No Familiarity	24	S13 bulls

0	0	.9	-.9	-3.1	36	No Familiarity	1 No Familiarity	19	S13 father
1	1	.1	.9	3.1	99	Limited Familiarity	2 Limited Familiarity	10	S4 passenger
0	0	.9	-.9	-3.1	109	No Familiarity	1 No Familiarity	24	S13 bulls
0	0	.9	-.9	-3.1	113	No Familiarity	1 No Familiarity	24	S13 bulls
1	1	.1	.9	3.1	136	No Familiarity	1 No Familiarity	10	S4 passenger
1	1	.1	.9	3.1	138	No Familiarity	1 No Familiarity	7	S4 creature
0	0	.9	-.9	-3.1	157	No Familiarity	1 No Familiarity	24	S13 bulls
0	0	.9	-.9	-3.1	163	Limited Familiarity	2 Limited Familiarity	24	S13 bulls
Cat	Score	Exp.	Resd	StRes	Num	rater	N Dhivehi English Fam	Nu	Item

Bibliography

Abercrombie, D., 1951. 'R.P. and local accent.' *The Listener*.
Reprinted in K. Bolton and Kachru, B.B. (eds.), 2006. *World Englishes: Critical Concepts in Linguistics. Vol. 1*. London: Routledge, p.219-223.

Aliakbari, M., 2002. Linguistic imperialism, linguistic democracy and English language teaching. *ITUA (Information Technology and Universities in Asia) 2002 Proceedings*.

Alderson, J. C., 1981. Report of the discussion on general language proficiency. In Alderson, J.C. and Hughes, A. (eds.), *Issues in language testing: ELT Documents*. London: The British Council, p.111.

Alderson, J. C. and Bachman L. F., 2004. Series editors' preface. In S. Luoma, S., 2004. *Assessing Speaking*. Cambridge: Cambridge University Press, p.ix-xi.

Alderson, J. C., Clapham, C. and Wall, D., 1995. *Language test construction and evaluation*. Cambridge: Cambridge University Press.

Anderson-Hsieh, J., Johnson, R. and Koehler, K., 1992. The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language learning*, 42(4), p.529-555.

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R. and Yamada, T., 2004. Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese/r/and English/l/and/r/. *Journal of Phonetics*, 32(2), p.233-250.

Avery, P. and Ehrlich, S., 1992. *Teaching American English Pronunciation*. Oxford: Oxford University Press.

Aziz, Y. Y., 1974. Some problems of English consonant sounds for the Iraqi learner. *ELT Journal*, 28(2), p.166-168.

Bachman, L. F., 2004. *Statistical analyses for language assessment*. Cambridge: Cambridge University Press.

Bachman, L.F. and Palmer, A.S., 1984. Some comments on the terminology of language testing. *Communicative competence approaches to language proficiency assessment: Research and application*, p.34-43.

Bachman, L.F. and Palmer, A.S., 1996. *Language testing in practice*. Oxford: Oxford University Press.

Bachman, L. and Palmer, A., 2010. *Language assessment in practice*. Oxford: Oxford University Press.

Bachman, L.F. and Savignon, S.J., 1986. The evaluation of communicative language proficiency: a critique of the ACTFL Oral Interview. *Modern Language Journal*, 70(4), p.380-390.

Ballmer, T.T., 1981. A typology of native speakers. In Coulmas, F. (ed.), *A festschrift for native speaker*. New York: Mouton Publishers, p.51-68.

Bamgbose, A., 1998. Torn between the norms: innovations in world Englishes. *World Englishes*, 17(1), p.1-14.

Barnwell, D., 1987. Oral proficiency testing in the United States. *British Journal of Language Teaching*, 25(1), p.35-42.

Bent, T. and Bradlow, A.R., 2003. The interlanguage speech intelligibility benefit. *The Journal of the Acoustical Society of America*, 114(3), p.1600-1610.

Bhatia, V.K., 2014. *Analysing genre: Language use in professional settings*. New York: Routledge.

Bond, T.G. and Fox, C.M., 2007. *Applying the Rasch model: Fundamental measurement in the human sciences*. New York: Routledge.

Boomershine, A., 2006. Perceiving and processing dialectal variation in Spanish: An exemplar theory approach. *8th Hispanic Linguistics Symposium*.

Boothroyd, A., 1984. Auditory perception of speech contrasts by subjects with sensorineural hearing loss. *Journal of Speech, Language, and Hearing Research*, 27(1), p.134-144.

Bradlow, A.R. and Bent, T., 2008. Perceptual adaptation to non-native speech. *Cognition*, 106(2), p.707-729.

Brown, A., 1995. The effect of rater variables in the development of an occupation specific language performance test. *Language Testing*, 12(1), p.1-15.

Brown, A., 1991. *Pronunciation Models*. Singapore: NUS Press.

Browne, K. and Fulcher, G., 2016. Pronunciation and intelligibility in assessing spoken fluency. In Isaacs, T. and Trofimovich, P. (eds.), *Second Language Pronunciation Assessment: Interdisciplinary Perspectives.* Bristol: Multilingual Matters, p.37-53.

Bybee, J., 2006. From usage to grammar: The mind's response to repetition. *Language*, 82(4), p.711-733.

Bygate, M., 1988. Units of oral expression and language learning in small group interaction. *Applied linguistics*, 9(1), p.59-82.

Cain, B.D., 2000. *Dhivehi (Maldivian): a synchronic and diachronic study*. (Doctoral dissertation, Cornell University, Ithaca, New York). Retrieved from http://pubman.mpdl.mpg.de/pubman/item/escidoc:407629/component/escidoc:407628/dhivehi_cain2000_o.pdf

Cain, B.D., and Gair, J.W., 2000. *Dhivehi (Maldivian)*. Munich: Lincom Europa.

Canagarajah, S., 2006. Changing communicative needs, revised assessment objectives: Testing English as an international language. *Language Assessment Quarterly: An International Journal*, 3(3), p.229-242.

Carey, M.D., Mannell, R.H. and Dunn, P.K., 2011. Does a rater's familiarity with a candidate's pronunciation affect the rating in oral proficiency interviews?. *Language Testing*, 28(2), p.201-219.

Carruthers, S.W., 2006. Pronunciation difficulties of Japanese speakers of English: Predictions based on a contrastive analysis. *Hawaii Pacific University TESOL Working Paper Series*, 4(2), p.17-24.

Catford, J.C., 1950. Intelligibility. *ELT Journal*, (1), p.7-15.

Clapham, C., 1996. *The development of IELTS: A study of the effect of on reading comprehension* (Vol. 4). Cambridge: Cambridge University Press.

Clark, J.L., 1979. Direct vs. semi-direct tests of speaking proficiency. In Brière, E.J. and Hinofotis, F.B. (eds.), 1979. *Concepts in Language Testing: Some Recent Studies.* Washington, DC: Teachers of English to Speakers of Other Languages, p.35-49.

Cohen, J., 1988. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. New Jersey: Routlage.

Cooper, C.R., 1977. Holistic evaluation of writing. In Cooper, C.R. and Odell, L. (eds.), *Evaluating writing: Describing, measuring, judging*, p.3-31.

Criper, C. and Davies, A., 1988. *ELTS validation project report*. University of Cambridge, Local Examinations Syndicate.

Cruz, N.C., 2007. Terminologies and definitions in the use of intelligibility: state-of-the-art. *Rev. Brasileira de Linguistica Aplicada*, 7(1), p.149-159.

Crystal, D., 1998. *English as a global language*. Cambridge: Cambridge University Press.

Dauer, R.M., 2005. The Lingua Franca Core: A new model for pronunciation instruction?. *Tesol Quarterly*, 39(3), p.543-550.

Davies, A., 1990. *Principles of Language Testing*. New Jersey: Blackwell.

Davies, A., 2003. *The Native Speaker: Myth and Reality.* UK: Multilingual Matters Ltd.

Derwing, T.M. and Munro, M.J., 1997. Accent, intelligibility, and comprehensibility. *Studies in second language acquisition*, 19(01), p.1-16.

Derwing, T.M. and Munro, M.J., 2009. Putting accent in its place: Rethinking obstacles to communication. *Language Teaching.*, 42(4), p.476–490.

Derwing, T.M., Rossiter, M.J. and Munro, M.J., 2002. 'Teaching native speakers to listen to foreign-accented speech', *Journal of Multilingual and Multicultural Development*, 23(4), p.245-259.

Deterding, D. and Kirkpatrick, A., 2006. Emerging South‐East Asian Englishes and intelligibility. *World Englishes*, 25(3‐4), p.391-409.

Dickerson, L. and Dickerson, W., 1977. Interlanguage phonology: Current research and future directions. In Corder, S.P. and Roulet, E. (eds.), *The notions of simplification, interlanguages and pidgins: Actes du 5ème Colloque de Linguistique Appliqué de Neufchatel*. Switzerland: Université de Neuchâtel, p. 18-30.

Everson, P. and Hines, S., 2010. How ETS Scores the TOEIC® Speaking and Writing Test Responses. *The research foundation for TOEIC: A compendium of studies*, p.8-1.

ETS., 2008. *TOEFL iBT Tips; How to prepare for the TOEFL iBT.* Princeton: Educational Testing Service.

ETS., 2012. TOEIC Examinee handbook speaking and writing. [PDF]. Retrieved from: http://www.ets.org/Media/Tests/TOEIC/pdf/TOEIC_Speaking_and_Writing_Examinee_Handbook.pdf

Fayer, J.M. and Krasinski, E., 1987. Native and nonnative judgments of intelligibility and irritation. *Language Learning*, 37(3), p.313-326.

Field, J., 2005. Intelligibility and the listener: The role of lexical stress. *TESOL quarterly*, 39(3), p.399-423.

Flege, J.E., 1981. The phonological basis of foreign accent: A hypothesis. *Tesol Quarterly*, 15(4), p.443-455.

Fox, J., 2004. Biasing for the best in language testing and learning: An interview with Merrill Swain. *Language Assessment Quarterly: An International Journal*, 1(4), p.235-251.

Francis, W.N., 1983. *Dialectology: an introduction*. Boston: Addison-Wesley Longman Ltd.

Frieda, E.M., Walley, A.C., Flege, J.E. and Sloane, M.E., 1999. Adults' perception of native and nonnative vowels: Implications for the perceptual magnet effect. *Perception & psychophysics*, 61(3), p.561-577.

Fritz, S., 2002. *The Dhivehi Language*. Germany: Ergon-Verlag.

Fulcher, G., 1993. The construction and validation of rating scales for oral tests in English as a foreign language (Doctoral dissertation, University of Lancaster, England). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.732.6006&rep=rep1&type=pdf

Fulcher, G., 2003. *Testing Second Language Speaking*. London: Pearson Education Limited.

Fulcher, G., 2010. *Practical language testing*. London: Hodder Education.

Gass, S. and Varonis, E.M., 1984. The effect of familiarity on the comprehensibility of nonnative speech. *Language learning*, 34(1), p.65-87.

Gescheider, G.A., 1976. *Psychophysics: Method and theory*. New Jersey: Lawrence Erlbaum.

Ghasemi, A. and Zahediasl, S., 2012. Normality tests for statistical analysis: a guide for non-statisticians. *International journal of endocrinology and metabolism*, 10(2), p.486-489.

Gimson, A.C., 1980. *An Introduction to the Pronunciation of English*. (3rd ed.). London: Edward Arnold.

Gnutzmann, C., 1999. *Teaching and learning English as a global language. Native and non-native perspectives*. Tübingen: Stauffenburg.

Goggin, J.P., Thompson, C.P., Strube, G. and Simental, L.R., 1991. The role of language familiarity in voice identification. *Memory & cognition*, 19(5), p.448-458.

Goldinger, S.D., 1992. Research on Speech Perception Technical Report No. 7. Indiana: Indiana University.

Goldinger, S.D., Pisoni, D.B. and Logan, J. S., 1991. On the nature of talker variability effects on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17(1), p.152.

Goldstein, B. and Washington, P.S., 2001. An initial investigation of phonological patterns in typically developing 4-year-old Spanish-English bilingual children. *Language, Speech, and Hearing Services in Schools*, 32(3), p.153-164.

Golombek, P. and Jordan, S.R., 2005. Becoming “black lambs” not “parrots”: A poststructuralist orientation to intelligibility and identity. *Tesol Quarterly*, 39(3), p.513-533.

Graddol, D., 1997. *The future of English?: A guide to forecasting the popularity of the English language in the 21st century*. London: British Council.

Graddol, D., 2006. *English next* (Vol. 62). London: British Council.

Green, R., 2013. *Statistical analyses for language testers*. London: Palgrave Macmillan.

Groom, C., 2012. Non-native attitudes towards teaching English as a lingua franca in Europe. *English Today*, 28(01), p.50-57.

Harding, L., 2011. *Accent and listening assessment: A validation study of the use of speakers with L2 accents on an academic English listening test*. Frankfurt: Peter Lang.

Hamp-Lyons, L., 1991. Scoring procedures for ESL contexts. *Assessing second language writing in academic contexts*, p.241-276.

Hill, C. and Parry, K., 1994. *From testing to assessment: English as an international language*. New York: Longman.

Hooper, J.B., 1976. Word frequency in lexical diffusion and the source of morphophonological change. *Current progress in historical linguistics*, p.96-105.

Huang, B.H., 2013. The effects of accent familiarity and language teaching experience on raters' judgments of non-native speech. *System*, 41(3), p.770-785.

Hughes, A., 2003. *Testing for language teachers*. Cambridge: Cambridge university press.

Ikeno, A., Pellom, B., Cer, D., Thornton, A., Brenier, J. M., Jurafsky, D. and Byrne, W., 2003. Issues in recognition of Spanish-accented spontaneous English. *ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*. Retrieved from http://www.dmcer.net/papers/isca2003.pdf

Isaacs, T., 2008. *Assessing second language pronunciation: A mixed methods study*. Germany: VDM Publishing.

Iverson, P. and Kuhl, P.K., 1995. Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America,* 97(1), p.553–562.

James, C., 1998. *Errors in language learning and use: Exploring error analysis*. London: Longman.

Jarvis, G.A., 1986. Proficiency testing: A matter of false hopes? *ADFL Bulletin,* 18(1), p.20-21.

Jenkins, J., 2000. *The phonology of English as an international language: New models, new norms, new goals*. Oxford: Oxford University Press.

Jenkins, J., 2002. A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics*, 23(1), p.83-103.

Jenkins, J., 2005. Implementing an international approach to English pronunciation: The role of teacher attitudes and identity. *Tesol Quarterly*, 39(3), p.535-543.

Jenkins, J., 2006. The spread of EIL: A testing time for testers. *ELT journal*, 60(1), p.42-50.

Jenkins, J., 2008. English as a Lingua Franca, Lecture 1. Waseda University, Tokyo, Japan. Retrieved from: http://www.jacet.org/2008convention/JACET2008_keynote_jenkins.pdf

Jenkins, J., 2009. English as a lingua franca: Interpretations and attitudes. *World Englishes*, 28(2), p.200-207.

Jenner, B., 1989. Teaching pronunciation: The common core. *Speak Out*, 4, p.2-4.

Johnson, K., 1990. The role of perceived speaker identity in F0 normalization of vowels. *The Journal of the Acoustical Society of America*, 88(2), p.642-654.

Johnson, K., 1997. Speech perception without speaker normalization: An exemplar model. *Talker variability in speech processing*, p.145-165.

Johnson, K., 2005. Speaker normalization in speech perception. In Pisoni, D.B. and Remez, R. (eds.), *The Handbook of Speech Perception*. Oxford: Blackwell Publishers.

Johnson, K., 2006. Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of phonetics*, 34(4), p.485-499.

Johnson, K., 2008. 15 Speaker Normalization in Speech Perception. In Pisoni, D.B. and Remez, R.E. (eds.), *The handbook of speech perception*, p.363.

Joo, M.J., 2007. The attitudes of students' and teachers' toward a Computerized Oral Test (COT) and a Face-To-Face Interview (FTFI) in a Korean university setting. *Journal of Language Sciences*, 14(2), p.171-193.

Kachru, B.B., 1985. Standards, codification and sociolinguistic realism: the English language in the Outer Circle. In Quirk, R. and Widdowson, H.G. (eds.), *English in the World: Teaching and Learning the Language and* Literatures. Cambridge: Cambridge University Press, p. 11–30.

Kachru, B.B., 1986. *The alchemy of English*: *The spread, functions, and models of non-native Englishes*. Champaign: University of Illinois Press.

Kachru, B.B., 1992. Models for non-native Englishes. *The other tongue: English across cultures*, 2, p.48-74.

Kachru, B.B., 1997. World Englishes and English-using communities. *Annual Review of Applied Linguistics*, 17, p.66-87.

Kachru, Y. and Smith, L.E., 2009. The Karmic cycle of world Englishes: Some futuristic constructs. *World Englishes*, 28(1), p.1-14.

Kennedy, S. and Trofimovich, P., 2008. Intelligibility, comprehensibility, and accentedness of L2 speech: The role of listener experience and semantic context. *Canadian Modern Language Review* 64, p.459–490.

Kenworthy, J. and Kenworthy, J., 1987. *Teaching English pronunciation* (Vol. 11). New York: Longman.

Kim, Y. H., 2009. An investigation into native and non-native teachers' judgments of oral English performance: A mixed methods approach. *Language Testing*, 26(2), p.187-217.

Kramsch, C., 1993. *Context and culture in language teaching*. Oxford: Oxford University Press.

Krisnawati, E., 2011. Pragmatic Competence in the Spoken English Classroom. *Indonesian Journal of Applied Linguistics (IJAL)*, 1(1), p.105-115.

Kubota, R., 2001. Teaching world Englishes to native speakers of English in the USA. *World Englishes*, 20(1), p.47-64.

Kuhl, P.K., 1991. Human adults and human infants show a ‘perceptual magnet effect’ for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50, p.93–107.

Kuhl, P.K., & Iverson, P., 1995. Chapter 4: Linguistic experience and the “Perceptual Magnet Effect,”. In Strange, W. (ed.), *Speech perception and linguistic experience: Issues in cross-language research*. York: York Press, p.121-154.

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N. and Lindblom, B., 1992. Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255(5044), p.606-608.

Lantolf, J.P. and Frawley, W., 1985. Oral proficiency testing: a critical analysis. *The Modern Language Journal,* 69, p.337-345.

Lazaraton, A., 1996. Interlocutor support in oral proficiency interviews: The case of CASE. *Language testing*, 13(2), p.151-172.

Lazaraton, A. and Riggenbach, H., 1990. Oral skills testing: A rhetorical task approach. *Issues in Applied Linguistics*, 1(2). p.196-217.

Levis, J., 1999. The intonation and meaning of normal yes-no questions. *World Englishes,* 18(3), p.373-380.

Linacre, M., 2012a. Many-facet Rasch measurement: Facets tutorial. Retrieved from www. winsteps. com/a/ftutorial1. pdf

Linacre, M., 2012b. Many-facet Rasch measurement: Facets tutorial. Retrieved from www. winsteps. com/a/ftutorial2. pdf

Lippi-Green, R., 1997. *English with an accent: Language, ideology, and discrimination in the United States*. Oxfordshire: Psychology Press.

Lively, S.E. and Pisoni, D.B., 1997. On prototypes and phonetic categories: a critical assessment of the perceptual magnet effect in speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, 23(6), p.1665.

Llurda, E., 2007. The representation of EFL teachers’ views on the role of English as a Lingua Franca. *VIAL: Vigo International Journal of Applied Linguistics* 4*,* p.11-23.

Lowenberg, P.H., 1993. Issues of validity in tests of English as a world language: whose standards?. *World Englishes*, 12(1), p.95-106.

Lowenberg, P., 2002. Assessing English proficiency in the Expanding Circle. World Englishes, 21(3), p.431-435.

Luoma, S., 2004. *Assessing Speaking*. Cambridge: Cambridge University Press.

Major, R.C., 2007. Identifying a foreign accent in an unfamiliar language. *Studies in Second Language Acquisition*, 29(4), p.539-556.

Martin, C.S., Mullennix, J.W., Pisoni, D.B. and Summers, W.V., 1989. Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(4), p.676.

Matsuda, A., 2003. Incorporating World Englishes in teaching English as an international language. *Tesol Quarterly*, 37(4), p.719-729.

McGarr, N.S., 1983. The intelligibility of deaf speech to experienced and inexperienced listeners. *Journal of Speech, Language, and Hearing Research*, 26(3), p.451-458.

McKay, S.L., 2002. *Teaching English as an international language: Rethinking goals and perspectives*. Oxford: Oxford University Press.

McNamara, T.F., 1996. *Measuring Second Language Performance*. London: Longman.

McNamara, T. F., 1997. 'Interaction' in second language performance assessment: Whose performance? *Applied linguistics*, 18(4), p.446-466.

McNamara, T.F., 2000. *Language Testing*. Oxford: Oxford University Press.

McNamara, T. and Roever, C., 2006. *Language testing: The social dimension*. Oxford: Blackwell.

Morley, J., 1991. The pronunciation component in teaching English to speakers of other languages. *Tesol Quarterly*, 25(3), p.481-520.

Mullennix, J.W., Pisoni, D.B. and Martin, C.S., 1989. Some effects of talker variability on spoken word recognition. *The Journal of the Acoustical Society of America*, 85(1), p.365-378.

Munro, M.J., 2008. Foreign accent and speech intelligibility. In Hansen Edwards, J.G. and Zampini, M.L. (eds), *Phonology and second language acquisition*. Amsterdam: John Benjamins Publishing Company, p.193-218.

Munro, M.J. and Derwing, T.M., 1994. Evaluations of foreign accent in extemporaneous and read material. *Language Testing*, 11(3), 253-266.

Munro, M.J. and Derwing, T.M., 1995. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language learning*, 45(1), p.73-97.

Munro, M.J., & Derwing, T.M., 1999. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 49(1), p.285-310.

Munro, M.J., Derwing, T.M. and Morton, S.L., 2006. The mutual intelligibility of L2 speech. *Studies in second language acquisition*, *28*(01), p.111-131.

Nelson, C.L., 1995. Intelligibility and world Englishes in the classroom. *World Englishes*, 14(2), p.273-279.

Norman, G., 2010. Likert scales, levels of measurement and the "laws" of statistics. *Advances in health sciences education*, 15(5), p.625-632.

Nosofsky, R.M., 1988. Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14(4), p.700.

O'Loughlin, K.J., 2001. *The equivalence of direct and semi-direct speaking tests*. Cambridge: Cambridge University Press.

Oyama, S., 1982. A sensitive period for the acquisition of a nonnative phonological system. In Krashen, S.D., Scarcella, R.C. and Long, M.H. (eds), *Child adult differences in second language acquisition*. New York: Newbury House Publishers, p.20-38.

Paikeday, T.M., 1985. *The native speaker is dead!: An informal discussion of a linguistic myth with Noam Chomsky and other linguists, philosophers, psychologists, and lexicographers*. Paikeday Pub.

Perneger, T. V., 1998. What's wrong with Bonferroni adjustments? *British Medical Journal, 316* (7139), p.1236-1238. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1112991

Pica, T., Holliday, L., Lewis, N. and Morgenthaler, L., 1989. Comprehensible output as an outcome of linguistic demands on the learner. *Studies in second language acquisition*, 11(1), p.63-90.

Pickering, L., 2006. Current research on intelligibility in English as a lingua franca. *Annual Review of Applied Linguistics*, 26, p.219-233.

Pierrehumbert, J., 2001. Exemplar dynamics: Word frequency, lenition and contrast. In Bybee, J. and Hopper, P. (eds), *Frequency and the emergence of linguistic structure*. Amsterdam: John Benjamins Publishing Company, p.137-57.

Pisoni, D.B., 1992. Some comments on talker normalization in speech perception. In Tohkura, Y.I., Vatikiotis-Bateson, E., & Sagisaka, Y. (eds.), *Speech perception, production and linguistic structure*. Amsterdam: IOS Press, p.143–151.

Pisoni, D.B., 1993. Long-term memory in speech perception: Some new findings on talker variability, speaking rate and perceptual learning. *Speech communication*, 13(1), p.109-125.

Prodromou, L., 2006. A reader responds to J. Jenkins's 'Current perspectives on Teaching World Englishes and English as a Lingua Franca'. *TESOL Quarterly* 41, p.409–13.

Quirk, R., 1981. International communication and the concept of nuclear English. In L. E. Smith (ed.), *English for cross-cultural* communication. London: Macmillan, p.151-165.

Rao, A. and Hashimoto, K., 1996. Intercultural influence: A study of Japanese expatriate managers in Canada. *Journal of International Business Studies*, 27(3), p.443-466.

Rajadurai, J., 2007. Intelligibility studies: A consideration of empirical and ideological issues. *World Englishes*, 26(1), p.87-98.

Rasch, G., 1980. *Some probabilistic models for intelligence and attainment tests*. Chicago: University of Chicago.

Reynolds, C., 2004. SONJA FRITZ: The Dhivehi language: descriptive and historical grammar of Maldivian and its dialects. *Bulletin of the School of Oriental and African Studies*, 67(02), p.247-249.

Riney, T.J. and Flege, J.E., 1998. Changes over time in global foreign accent and liquid identifiability and accuracy. *Studies in Second Language Acquisition*, 20(2), p.213-243.

Riney, T., Takagi, N. and Inutsuka, K., 2005. Phonetic Parameters and Perceptual Judgments of Accent in English by American and Japanese Listeners. *TESOL Quarterly*. 39(3), p.441-466.

Rooy, S.C.V., 2009. Intelligibility and perceptions of English proficiency. *World Englishes*, 28(1), p.15-34.

Royal‐Dawson, L. and Baird, J.A., 2009. Is teaching experience necessary for reliable scoring of extended English questions?. *Educational Measurement: Issues and Practice*, 28(2), p.2-8.

Ruscio, J., 2008. A probability-based measure of effect size: robustness to base rates and other factors. *Psychological methods,* 13(1), p.19–30.

Savignon, S.J., 1997. *Communicative Competence: Theory and Classroom Practice*. New York: McGraw-Hill.

Schnitzer, E., 1995. English as an international language: Implications for interculturalists and language educators. *International Journal of Intercultural Relations*, *19*(2), p.227-236.

Scovel, T., 1988. A time to speak: A psycholinguistic investigation into the critical period for human speech. New York: Harper and Row.

Seidlhofer, B., 2000. Mind the gap: English as a mother tongue vs. English as a lingua franca. *Views (Vienna English Working Papers)*, *9*(1), p.51-68.

Seidlhofer, B., 2001. Closing a conceptual gap: The case for a description of English as a lingua franca. *International Journal of Applied Linguistics*, *11*(2), p.133-158.

Seidlhofer, B., 2004. Research perspectives on teaching English as a lingua franca. *Annual Review of Applied Linguistics*, *24*, p.209-239.

Seidlhofer, B., 2009. Common ground and different realities: World Englishes and English as a lingua franca. *World Englishes*, *28*(2), p.236-245.

Seidlhofer, B., Breiteneder, A. and Pitzl, M.L., 2006. English as a lingua franca in Europe: Challenges for applied linguistics. *Annual Review of Applied Linguistics*, *26*, p.3-34.

Shafiro, V., Levy, E.S., Khamis-Dakwar, R. and Kharkhurin, A., 2013. Perceptual confusions of American-English vowels and consonants by native Arabic bilinguals. *Language and speech*, *56*(2), p.145-161.

Shohamy, E., 1994. The validity of direct versus semi-direct oral tests. *Language testing*, *11*(2), p.99-123.

Siegel, A. F., 1990. Multiple t tests: Some practical considerations. *TESOL Quarterly, 24* (4), p.773-775.

Skinner, M.W. and Miller, J.D., 1983. Amplification bandwidth and intelligibility of speech in quiet and noise for listeners with sensorineural hearing loss. *International Journal of Audiology*, *22*(3), p.253-279.

Smith, L.E., 1983. *Readings in English as an international language.* Oxford: Pergmaon Press.

Smith, L.E., 1987. *Discourse across cultures: Strategies in world Englishes*. New York: Prentice Hall.

Smith, L.E. and Nelson, C.L., 1985. International intelligibility of English: Directions and resources. *World Englishes*, *4*(3), p.333-342.

Smith, L.E., & Rafiqzad, K., 1979. English for cross-cultural communication: The question of intelligibility. *Tesol Quarterly*, 13(3), p.371-380.

Sobkowiak, W., 2005. Why not LFC? English pronunciation models: a changing scene. In Dziubalska-Kołaczyk, K. and Przedlacka, J. (eds), *English pronunciation models: a changing scene*. Bern: Peter Lang, p.131-149.

Spolsky, B., 1993. Testing across cultures: An historical perspective. *World Englishes*, *12*(1), p.87-93.

Stansfield, C.W., 1990. A Comparative Analysis of Simulated and Direct Oral Proficiency Interviews. Paper presented at the Annual Meeting of the Regional Language Centre Conference, Singapore. Retrieved from http://files.eric.ed.gov/fulltext/ED317073.pdf

Sung, C.C.M., 2013. English as a lingua franca and its implications for English language teaching. *JALT Journal*, *35*(2), p.177-194.

Sussman, J.E. and Lauckner-Morano, V.J., 1995. Further tests of the “perceptual magnet effect” in the perception of [i]: Identification and change/no-change discrimination. *J. Acoust. Soc. Am*, *97*(1), p.539-552.

Taylor, L., 2006. The changing landscape of English: Implications for language assessment. *ELT Journal*, *60*(1), p.51-60.

Taylor, L. and Galaczi, E., 2011. Scoring validity. Examining speaking: Research and practice in assessing second language speaking, 30, p.171.

Timmis, I., 2002. Native‐speaker norms and International English: a classroom view. *ELT journal*, *56*(3), p.240-249.

Tripathi, P.D., 1998. Redefining Kachru's ‘Outer Circle’of English. *English Today*, *14*(04), p.55-58.

TOEFL (2008). *TOEFL iBT Tips: How to prepare for the TOEFL iBT*. Princeton: ETS.

Van Wijngaarden, S.J., 2001. Intelligibility of native and non-native Dutch speech. *Speech communication*, *35*(1), p.103-113.

Van Wijngaarden, S.J., Steeneken, H.J. and Houtgast, T., 2002. Quantifying the intelligibility of speech in noise for non-native talkers. *The Journal of the Acoustical Society of America*, *112*(6), p.3004-3013.

Wells, J.C., 1982. Accents of English. Cambridge; New York: Cambridge University

Press.

Wenk, B.J., 1979. Articulatory setting and de-fossilization. *Interlanguage Studies Bulletin*, 4(2), p.202-220.

White, R., 1997. Going round in circles: English as an international language and crosscultural capability. In *Cross-cultural capability conference* (Vol. 97). Retrieved from http://host.uniroma3.it/docenti/boylan/text/white01.htm

Widdowson, H.G., 1994. The ownership of English. *TESOL Quarterly*, 28(2), p.77-389.

Williams, L., 1977. The perception of stop consonant voicing by Spanish-English bilinguals. *Perception & Psychophysics*, *21*(4), p.289-297.

Winke, P. and Gass, S., 2013. The influence of second language experience and accent familiarity on oral proficiency rating: A qualitative investigation. *TESOL Quarterly*, *47*(4), p.762-789.

Winke, P., Gass, S. and Myford, C., 2011. The relationship between raters' prior language study and the evaluation of foreign language speech samples. *ETS Research Report Series*, 2011(2), p.i-67.

Winke, P., Gass, S., & Myford, C., 2013. Raters' L2 background as a potential source of bias in rating oral performance. *Language Testing*, *30*(2), p.231-252.

Winsteps, 2013. Logit and Probit: what are they? Retrieved from http://www.winsteps.com/winman/whatisalogit.htm

Wright, B.D., Linacre, J M., Gustafson, J. E. and Martin-Lof, P., 1994. Reasonable mean-square fit values. *Rasch measurement transactions*, 8(3), p.370.

Xi, X., 2007. Evaluating analytic scoring for the TOEFL® Academic Speaking Test (TAST) for operational use. *Language Testing*, *24*(2), p.251-286.

Xi, X. and Mollaun, P., 2009. How Do Raters from India Perform in Scoring the TOEFL iBT [TM] Speaking Section and What Kind of Training Helps? *TOEFL iBT [TM] Research Report*. RR-09-31. Princeton: ETS.

Young, T. J., & Walsh, S. (2010). Which English? Whose English? An investigation of 'non-native' teachers' beliefs about target varieties. *Language, Culture and Curriculum*, *23*(2), p.123-137.

Zhang, Y. and Elder, C., 2010. Judgments of oral proficiency by non-native and native English speaking teacher raters: Competing or complementary constructs?. *Language Testing*, 28(1), p.31-50.

Zuengler, J., 1988. Identity Markers and L2 Pronunciation. *Studies in Second Language Acquisition,* 10(1), p.33-50.