The Development of a Valid Informant-Report Measure of Children's Cattell-Horn-Carroll Cognitive Abilities

Lydia Soh

Bachelor of Commerce, The University of Melbourne

Postgraduate Diploma of Psychology, Monash University

A thesis submitted in partial fulfilment of the requirements for the degree of

Master of Educational & Developmental Psychology/Doctor of Philosophy

Faculty of Education

Monash University

July 2018

Copyright Notice

Abstract

Performance-based maximal assessments have long since been the dominant method used to obtain an index of an individual's cognitive functioning. However, the iterative nature of cognitive ability assessments and high hourly costs associated with the assessment process indicate that it could quickly become a lengthy and expensive exercise. A possible economical solution to these limitations is to obtain valid estimations of children's cognitive functioning from their parents and teachers. In using this information as an adjunct, the practitioner could streamline the process by focusing on the cognitive abilities relevant to the referral concern. The use of informant-report methods in measuring cognitive functioning has been limited and viewed as problematic, largely due to a lack of measures based on theoretically and psychometrically robust ability models, an issue that the advent of Cattell-Horn-Carroll (CHC) theory can potentially circumvent. This dissertation capitalises on recent theoretical advances in cognitive abilities by developing and piloting an informant-report measure underpinned by CHC theory that includes measurement of the eight broad abilities known to be important for academic achievement: Fluid Reasoning (*Gf*), Comprehension-Knowledge (*Gc*), Short-Term Working Memory (*Gwm*), Learning Efficiency (*Gl*), Retrieval Fluency (*Gr*), Visual-Spatial Processing (*Gv*), Auditory Processing (*Ga*), and Processing Speed (*Gs*).

This research endeavour commenced with meta-analytic reviews of the current literature pertaining to parent-report and teacher-report of children's cognitive and academic abilities to explore the human-related characteristics, and more importantly, measurement conditions that were most conducive to eliciting valid parent- and teacher-reports of ability. Using established scale development procedures, these insights were then applied in conjunction with CHC theory to develop the Estimates of Children's Cognitive Abilities (ECCA) instrument, a multi-item, inter-individual measure comprising of 58 scale items that

reflected behavioural operationalisations of cognitive abilities, such that items referred to cognitive tasks, observable behaviours, and skills. The instrument's content validity was established through the use of expert panel ratings and cognitive interviewing with members (two psychologists and two educational specialists) of the target audience.

A sample of 309 primary and secondary school teachers were recruited for scale validation purposes, conducted in two phases. Using exploratory factor extraction, a three-factor solution was found to be most optimal for the obtained dataset. Thus, the expectation of an eight-factor model was not met in the first phase of the scale validation study. However, these results provided some evidence for a factor structure similar to conceptualisations of intelligence in Carroll's (1993) factor-analytic survey of human cognitive abilities, and are consistent with the statistical methodologies used at that time. The second phase using confirmatory factor analytic methods endorsed the proposed factor structure in accordance with the eight broad abilities as defined in CHC theory, with results revealing excellent convergent validity of items with their respective ability constructs, and adequate discriminant validity across factors.

The current study thus provides evidence for the capacity of important adults in a child's life, such as teachers, to accurately estimate a child's level of cognitive functioning. The teachers' data adequately demonstrated the reliability and validity of the newly developed ECCA scale to report children's CHC abilities. This dissertation also supports the notion of using a strong theoretical model containing robust and well-validated factors when developing informant-report measures of cognitive ability, and brings us a step closer to developing a more effective multi-informant, multi-method measurement of abilities. It is recommended that future research prioritise the application of robust theoretical models in the development of informant-reports of cognitive abilities.

Declarations

This thesis contains no material which has been accepted for the award of another degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Lydia Soh

21 July 2018

Ethics Approval

The research for this project received the approval of the Monash University Standing Committee for Ethical Research in Humans.

Project 2016-0228

Publications during Enrolment

Soh, L. & Jacobs, K. (2016). Accuracy of parent reports of children's cognitive abilities: A meta-analysis. Paper presented at the 15th annual Australian Conference on Personality and Individual Differences. Melbourne, Victoria, Australia.

Acknowledgements

This thesis would not exist without the numerous people in my life who were convinced, even when my own convictions were tested, that I would eventually succeed. I would first like to express my gratitude to Dr Kate Jacobs for inspiring me to pursue this degree, for her continuous faith in me throughout my postgraduate career, and for sharing her immense knowledge of the research domain. I would also like to thank Dr Louise Mclean for her careful attention to detail, and Dr Shane Costello for imparting his wisdom on statistical methods. I am grateful to you both for adopting me as supervisee so late into my candidature, and for your insightful comments and encouraging pep talks to help me get this thesis over the finish line.

To the many PhD friends I have made along the way, in particular Simone, Rachael, Simon, and Capella. Thank you for bringing light and laughter into my academic life, and providing intellectual and emotional advice whenever I needed it. No day was too dark when long lunch breaks over a large bowl of chips were had. Your companionship has transformed a would-be isolating process into a joyous one, and I am so glad to have met you all. I hold each and every one of you in high esteem, and will be cheering you on enthusiastically as you progress toward the completion of your own PhD journey.

To all my loved ones in various parts of the world, thank you for your moral support throughout the years and for always encouraging me to pursue my dreams. I am especially grateful to my parents for their emotional and financial support in getting me to where I am today. Finally, to Brad, my selfless partner in crime, dance, and life. Thank you for being my eternal cheerleader, for celebrating my triumphs and commiserating my woes, and for taking over the household duties without complaint while I was otherwise preoccupied. Your steadfast love and unwavering belief in me has carried me through this experience, and I truly could not have done this without you.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1

Introduction

Parents and teachers possess invaluable knowledge about the daily functioning and behaviours of children referred for psychoeducational assessment. This information is typically used by practitioners to form a cohesive picture of the child's background, make diagnostic decisions, and develop interventions. This is evident in psychologists' typical practice of collecting information about the child via clinical interviews with important adults, and with informant-report tools that assess children's functioning in various psychosocial and psychological constructs (e.g., Behaviour Assessment System for Children that provides an index of children's emotional functioning through parent- and teacher-reports; Reynolds & Kamphaus, 2015). Yet, information gleaned from parents and teachers is currently infrequently used to report on children's cognitive abilities, as the extent to which they are accurate informants of this domain is not well understood. Studies that have investigated the validity of parent- and teacher-reports of children's cognitive abilities have shown that the measurement of cognitive abilities via informant-report methods is a flawed practice.

Another point of contention regarding this research has been the lack of well-defined abilities on which to base these measures. Despite great advancements in the field of cognitive theory and testing practices as seen in the seminal work of Raymond Cattell, John Horn, and John Carroll (Jacobs, 2012; McGrew, 2009), there are currently very limited applications of well-established and validated cognitive ability theories in the development of parent- and/or teacher-report measures. The aim of the current dissertation was thus to take preliminary steps in exploring the applicability of more theoretically robust ability constructs for the development of valid and reliable informant-report measures of children's cognitive ability.

A secondary objective of the dissertation was to identify the measurement conditions that are most favourable to eliciting a valid index of children's cognitive ability from parents and teachers. It is anticipated this research will be foundational for future validation studies of the newly developed informant-report scale, or in the determination of appropriate methodological decisions that scale developers should take into consideration in the development of valid informant-report measure of cognitive abilities. This thesis also presents a method for cognitive interviewing as an additional qualitative step in maximising the content validity of scale items.

The remainder of this chapter delineates the research context and elaborates on the rationale of a dissertation based on CHC theory. Subsequently, the scope and focus of the dissertation is discussed, followed by the research problem and accompanying research questions that the current thesis aims to answer. The design, method and data analysis plan of the separate phases of studies is then outlined. The chapter then concludes with an outline of each of the chapters of this thesis

1.1 The Context

In attempting to explain and clarify the complex phenomena that differentiates individuals' differing patterns of cognitive capabilities, the construct of human intelligence has induced a considerable amount of interest within the field of psychology, and in modern society, for over a century. Cognitive ability, as defined by Carroll (1993), is an ability that requires the processing of mental information appropriately to perform tasks related to that information. The measurement of cognitive abilities has traditionally been completed using performance-based assessments and tests, such as the Wechsler series of intelligence scales. These assessment processes constitute a significant part of psychologists' professional responsibilities (Ortiz & Lella, 2004). The prominence of intelligence testing in psychology stems, in large part, to research that has demonstrated that psychometric tests of intelligence can facilitate the prediction of many important life outcomes such as academic achievement, vocational success, socio-economic status, and positive psychological development (Dreary, Leaper, Murray, Staff, & Whalley, 2003; Gottfredson, 1997; Hofer & Clouston, 2014; Lubinski, 2004). Within the field of educational psychology, the assessment of cognitive abilities provides a framework which important stakeholders in a child's life, such as their parents and teachers, can use to develop an understanding of the different ways in which children learn (Logsdon, 2014). Cognitive ability assessments also provide practitioners with a means to identify academic underachievement or specific learning disabilities (Baudson & Preckel, 2013), thereby facilitating the planning and implementation of educational interventions that are specifically tailored to the child's cognitive learning profile.

The performance-based method of obtaining an index (or multiple ability indexes) of an individual's cognitive abilities is widely accepted practice. However, a comprehensive psychoeducational assessment can be a lengthy endeavour due to its inherent iterative process. Practitioners may have already engaged in what they believed to be a comprehensive evaluation, only to find that additional data needs to be collected to inform and supplement the results of the initial assessment (Dombrowski, 2015). The overall process can quickly become a time-consuming one, for both practitioner and client. Coupled with the fact that the recommended hourly fee for Australian psychologists is $251 (Australian Psychological Society [APS], 2018), the accessibility of psychological assessment could be limited for some segments of society, particularly those of lower socio-economic statuses who are likely most in need of such services. Although there are ways to obtain psychological assessment services through the public sector (e.g., school psychologists working in government schools), long waiting lists and having to meet specific requirements (only referral concerns of Intellectual Disability and Severe Language Disorder are accepted) means that the likelihood of a student obtaining assessment in a timely manner is low.

There is thus a need for economical solutions to ensure the financial efficacy of psychoeducational assessments. One option is to ensure the assessment process is focused on the ability domains that are most relevant to answering the referral concern and planning of interventions (Groth-Marnat, 1999). The advent of well-validated cognitive ability theories such as the Cattell-Horn-Carroll (CHC) theory has afforded practitioners with a psychometrically defensible technique (i.e., cross-battery assessment approach, XBA) for designing such time and cost effective assessments. The CHC framework also assists test assessment developers to better operationalise and design time efficient assessment tools.

1.2 Rationale for a CHC-Based Dissertation

This dissertation sought to investigate the use of a psychometrically defensible theory as a foundation on which to develop a valid informant-report measure of children's cognitive abilities. As a taxonomy of cognitive abilities and as a set of theoretical explanations regarding the existence of variations in cognitive ability levels across and within individuals, CHC theory is currently known as the "consensus framework from which cognitive abilities are now most often conceptualised and measured" (McGrew & Wendling, 2010, p. 651). Indeed, CHC theory is highly regarded by many researchers in the extant intelligence theory and assessment literature as the most influential and well-validated theory of cognitive abilities currently available (e.g., Ackerman & Lohman, 2006; Detterman, 2011; Flanagan & Dixon, 2014; Keith & Reynolds, 2010; McGrew, 2009; Newton & McGrew, 2010). As such, there is potential for this model to be used as a much needed robust theoretical basis for the development of a valid informant-report measure of children's cognitive abilities.

CHC theory is the amalgamation of two prominent psychometric models of cognitive abilities; namely, Cattell-Horn's *Gf-Gc* theory and Carroll's Three-Stratum Theory (McGrew, 2009), making CHC theory an integration of more than 60 years of factor analytic research in the domain of cognitive abilities (Pässler, Beinicke, & Hell, 2015). CHC theory assumes a hierarchical model of cognitive abilities with three strata, starting with general intelligence (*g*) at Stratum III, 16 broad cognitive abilities (Flanagan, Ortiz, & Alfonso, 2013; Schneider & McGrew, 2012) at Stratum II, followed by over 70 narrow abilities located at Stratum I (Kaufman, Reynolds, Liu, Kaufman, & McGrew, 2012; Vanderwood, McGrew, Flanagan, & Keith, 2002). Definitions of the 10 broad ability areas related to cognitive ability and academic achievement are included in Table 1.1, accompanied by a visual representation of the model in Figure 1.1.

The conception of CHC theory has since brought about an evolution in cognitive testing practices, such that the use of global IQ scores like the Full Scale Intelligence Quotient (FSIQ), as a single measure in ascertaining an individual's level of cognitive functioning can no longer be considered as valid practice (Jacobs, 2012). This is further amplified in the finding that the FSIQ is not an adequate reflection of global intellectual functioning for either children with disabilities or children who are typically developing but display significant variability in their learning profiles (Fiorello, Hale, McGrath, Ryan, & Quinn, 2002). It has been noted that more than 50% of the variance in academic achievement cannot be explained by measures of general intelligence alone (Jensen, 1998), and that specific abilities such as processing speed, short-term working memory, and visual-spatial processing may be better candidates for explaining variance in academic skills beyond what is accounted for by general intelligence (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Flanagan, Alfonso, & Reynolds, 2013; Luo & Petrill, 1999; Rohde & Thompson, 2007). Additionally, establishing an individual's pattern of cognitive and academic strengths and weaknesses in the various broad and narrow ability areas can also help to inform the development of individually tailored interventions that would be more effective in addressing specific referral concerns (Mascolo, Flanagan, & Alfonso, 2014). Hence, the practical implication of CHC theory is that individuals should be assessed for the total range of abilities as specified in the framework – however, as there is typically insufficient time to conduct such comprehensive assessments, research is needed to guide the process of referral-focused selective testing, such that practitioners can easily select which abilities should form the focus of cognitive ability testing in particular cases (Carroll, 1997).

In response to this, the cross-battery assessment approach (XBA) was developed by Flanagan and colleagues (Flanagan & McGrew, 1997; Flanagan, McGrew, & Ortiz, 2000; Flanagan & Ortiz, 2001; McGrew & Flanagan, 1998), and allows for the construction of cognitive ability assessments that more closely approximates the complete range of ability and processes (Flanagan, Ortiz, et al., 2013). The approach was born out of the realisation that single test batteries are often insufficient in providing a measure of all abilities that may be relevant to referral concerns (Jacobs, 2012). For example, even in its most recent revision, the Wechsler Intelligence Scale for Children – Fifth Edition (WISC-V) does not include subtests measuring the broad ability of *Ga*, while subtests measuring narrow abilities subsumed under *Gl* (associative memory*)* and *Gr* (naming facility) are not part of the primary index scales, despite research showing that these ability areas are instrumental in predicting academic achievement (Flanagan, Ortiz, Alfonso, & Mascolo, 2006). The XBA technique thus provides a solution for practitioners with psychometrically defensible means to "cross" (i.e., use more than one) intelligence batteries to supplement one battery with subtests from other batteries, to ensure that the assessment process effectively addresses the referral concern by capturing a holistic picture of all important ability areas.

Table 1.1

Definitions of Cattell-Horn-Carroll Broad Cognitive Abilities Represented in Cognitive and Achievement Batteries

Broad ability area	Definition
Comprehension-knowledge (Gc)	The ability to comprehend and communicate culturally-valued knowledge. Gc includes the depth and breadth of both declarative and procedural knowledge and skills such as language, words, and general knowledge developed through experience, learning and acculturation.
Reading and writing (Grw)	The depth and breadth of declarative and procedural knowledge and skills related to written language
Quantitative knowledge (Gq)	The depth and breadth of declarative and procedural knowledge related to mathematics. The *Gq* domain is likely to contain more narrow abilities than currently listed in the CHC model
Fluid reasoning (*Gf*)	The use of deliberate and controlled procedures (often requiring focused attention) to solve novel "on the spot" problems that cannot be solved by using previously learned habits, schemas, and scripts
Short-term working memory (Gwm)	The ability to maintain and manipulate information in active attention. The mind's mental "scratchpad" or "workbench".
Visual-spatial processing (Gv)	The ability to make use of simulated mental imagery to solve problems. Perceiving, discriminating and manipulating images in the "mind's eye.
Auditory processing (Ga)	The ability to discriminate, remember, reason, and work creatively (on) auditory stimuli, which may consist of tones, environmental sounds, and speech units.
Learning Efficiency (Gl)	The ability and efficiency to learn, store, and consolidate new information over periods of time measured in minutes, hours, days, and years.
Retrieval Fluency (Gr)	The rate and fluency at which individuals can access information stored in long-term memory
Processing Speed (Gs)	The ability to control attention to automatically, quickly and fluently perform relatively simple repetitive cognitive tasks. Attentional fluency or attentional speediness.

Adapted from Schneider, W. J., & McGrew, K. S. (in press). The Cattell-Horn-Carroll Theory of Cognitive Abilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (4th ed.,) New York: Guilford Press. Adapted with permission.

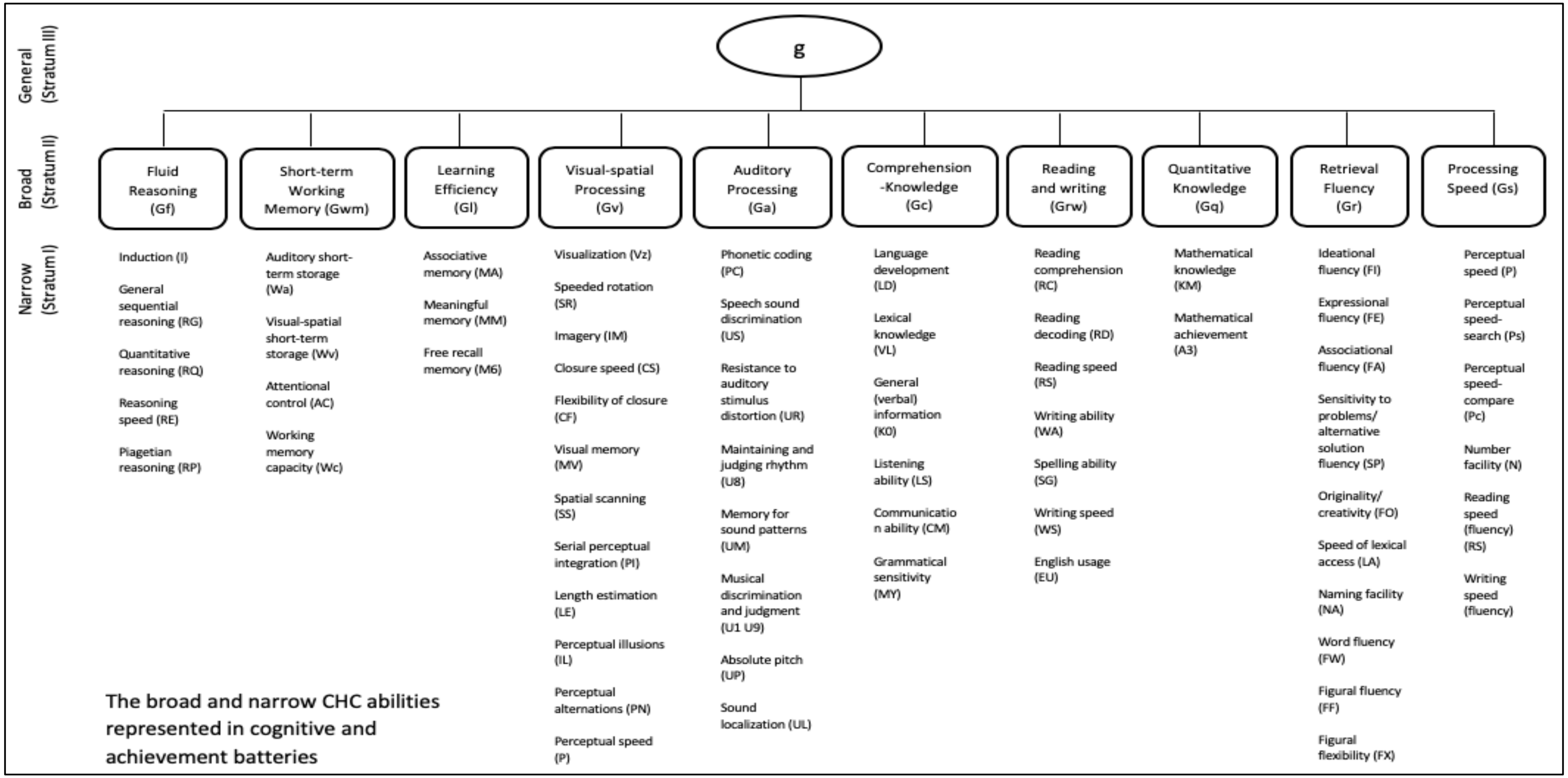


Figure 1.1 The Cattell-Horn-Carroll model of cognitive abilities

Adapted from Schneider, W. J., & McGrew, K. S. (in press). The Cattell-Horn-Carroll Theory of Cognitive Abilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests and issues.* (4th ed.,) New York: Guilford Press. Adapted with permission

The XBA method also assists practitioners in formulating *a priori* hypotheses about the possible locations of cognitive deficits of individuals before conducting the assessment process. In an example of a young client presenting with difficulties in math achievement, a clinician can review empirical research regarding relations between CHC abilities and this achievement domain (e.g., McGrew & Wendling, 2010; Taub, Keith, Floyd, & McGrew, 2008) to determine which abilities should be emphasised for adequate measurement during assessment (the narrow abilities and processes of *Gf*, *Gc*, *Gwm*, and *Gs* appear to be most important; Flanagan, Ortiz, et al., 2013). Further, if time and/or financial concerns were present, the clinician can reasonably exclude ability areas that are not crucial in answering the referral concern. Using the same example, since *Gv* may be important primarily for higher-level mathematics and is less pertinent to the successful acquisition of math-related skills for younger clients (Flanagan, Ortiz, et al., 2013), it could be excluded to reduce the amount of time spent on the process.

Another potential source of information that practitioners can use to develop hypotheses about an individual's cognitive profile is from the parent and teacher of the child being referred for assessment. In asking these important adults about the performance of the child in various cognitive tasks, hypotheses can be formed about the child's specific cognitive strengths and weaknesses, thereby streamlining the process further. Given the high-standing status and significance that parents and teachers have in a child's life, and that they interact with the child in different contexts, these adults likely possess instrumental first-hand knowledge of the child's cognitive strengths and weaknesses. Hence, if the clinician is able to obtain valid information from these "experts" of the child prior to conducting a formal assessment, the measurement of abilities that are not relevant to the referral concern could be justifiably excluded from the process.

However, the utility of parents and teachers as informants of children's cognitive functioning is not clearly established in the current literature. Although this could be due, in part, to the prevailing method of performance-based tests as official measures of cognitive ability, the validity of an informant-report measure could also be argued as questionable. This is because such scales are essentially quantifications of the subjective opinions of a child's intelligence from others (Reid & Maag, 1994), and hence could be susceptible to the influence of an array of external variables unrelated to the child's actual ability levels. The extant literature has highlighted the existence of several variables related to individual differences in people, such as child gender (e.g., Furnham, Reeves, & Budhani, 2002; Hinnant, O'Brien, & Ghazarian, 2009), parent gender (e.g., Chamorro-Premuzic, Arteche, Furnham, & Trickot, 2009; Furnham & Petrides, 2004), years of teaching experience (e.g., Babad, 1985; Hofer, 2015), that influence how accurate parents and teachers are in their estimations of children's cognitive abilities. Nonetheless, because these individual differences amongst people are naturally occurring, they are invariably unchangeable. Therefore, the onus is on researchers and scale developers to determine the measurement conditions (i.e., item-level factors) that facilitate valid estimations from parents and teachers when reporting on children's cognitive functioning, since these variables are controllable.

Much less attention has been paid to the role that measurement conditions play in influencing informant accuracy, which could be attributed to the fact that current research has largely failed to take into consideration basic psychometric principles when developing methods to measure children's cognitive abilities by informant-report. The scarcity of informant-report measures that have been developed based on well-established cognitive ability theories is apparent in the present literature. For example, in some studies investigating informant accuracy, raters are asked to make predictions on the child's performance on the various tasks in objective measures (e.g., Miller, Manhal, & Mee, 1991), or to estimate the child's "overall intelligence" or "school achievement" alone (e.g., Li, Pfeiffer, Petscher, Kumtepe, & Mo, 2008). Laypersons' unfamiliarity with standardised cognitive test tasks and the existing differences in laypersons' implicit conceptions of intelligence implies that these methods are likely to lead to ambiguity in respondent interpretations, and a subsequent increased risk of inaccurate judgments. Additionally, conclusions about informant accuracy have been drawn from studies using measures underpinned by Gardner's (1983) multiple intelligences (MI) theory (e.g., Hernández-Torrano, Ferrándiz, Ferrando, & Prieto, 2014). However, since there are no standardised assessment tools that are performance-based measures of the MI constructs, what the informants are asked to report on could be inherently different to what is being objectively measured; hence, the validity of these findings is questionable. Therefore, in determining how accurate parents and teachers can be in their estimations of children's cognitive ability levels, researchers should endeavour to obtain informant-reports using measures based on psychometrically and theoretically robust theoretical frameworks (i.e., CHC theory), so that the comparison of these informant-reports to theoretically analogous, standardised performance-based measures is possible.

1.3 Scope and Focus of the Dissertation

Since the aim of the present dissertation is to develop a measure of informant-report of children's cognitive abilities that could be used as a tool to inform and support the results of traditional psychoeducational assessment, a selection of eight CHC broad abilities were chosen for this initial investigation. These were: Fluid Reasoning (*Gf*), Comprehension-Knowledge (*Gc*), Short-Term Working Memory (*Gwm*), Learning Efficiency (*Gl*), Retrieval Fluency (*Gr*), Visual-Spatial Processing (*Gv*), Auditory Processing (*Ga*), and Processing Speed (*Gs*). These eight abilities were selected due to extensive research showing evidence for the importance of these various individual broad abilities in the prediction of achievement in specific academic domains (Cormier, Bulut, McGrew, & Singh, 2017; Flanagan et al., 2006; McGrew & Wendling, 2010). Although the abilities of *Grw* and *Gq* which pertain to reading, writing and math ability are considered to be cognitive abilities within the CHC model, these were not included for investigation within the current dissertation. This was because the intention was to create a tool for parents and teachers to report only on the cognitive abilities that were important for academic achievement, and not to report on both cognitive and academic abilities. The focus on these specific broad ability areas is also reflective of the paradigm shift in focus from general intelligence to specific cognitive abilities that has resulted from the advent of CHC theory (Jacobs, 2012).

1.4 Research Problem and Questions

There is a need for a practical tool that would assist practitioners in deciding which of the multitude of broad and narrow abilities should be at the forefront of a cognitive assessment, so that the financial efficacy of this traditionally lengthy and costly process can be ensured. While there have been forays into developing parent- and teacher-report measures of cognitive ability (e.g., Lamb, 2008; Williams, Ochs, Williams, & Mulhern, 1991), the development of informant-report measures that are based on well-validated theories of cognitive ability is lacking in the extant literature, with only one parent-report measure (Waschbusch, Daleiden, & Drabman, 2000) based on *Gf-Gc* theory (the predecessor of CHC theory), and one teacher-report measure, the Children's Psychological Processes Scale (CPPS; Dehn, 2012) based on CHC theory and Luria's (1970) neuropsychological theory of human cognitive processes.

Hence, the intention of this dissertation is to replicate and extend the findings of a previous thesis by Jacobs (2012), in which an informant-report measure of three CHC broad cognitive ability areas (*Gf, Gc*, and *Gv*), named Self-Report Measure of Cognitive Abilities (SRMCA) was developed and validated. The findings of this previous study highlighted that the use of psychometrical principles and theoretical constructs allowed for a valid measure of cognitive abilities via self-report methods to be created. Similarly, the aim of the current study was to determine if these conditions would also allow for the development of a parent-and teacher-report measure of children's cognitive abilities that would be reliable and valid. To test this, a multi-item, informant-report measure of CHC cognitive abilities (named the Estimates of Children's Cognitive Abilities, ECCA) was developed, and its validity investigated.

Prior to developing the informant-report measure in this dissertation, an investigation of the human and method factors that are currently known to have influences on parent and teacher accuracy in rating children's abilities was undertaken via meta-analytic reviews of the extant literature. The aim at this stage of the research was to answer the following:

1. What are the factors that influence the validity of parent-reports of children's cognitive abilities?
2. What are the factors that influence the validity of teacher-reports of children's cognitive abilities?

Insights gained from the meta-analytic endeavour were applied in the scale development process. The overarching aim at this stage was to devise an appropriate methodology for devising scale items that facilitate valid estimations from parents and teachers regarding children's abilities. To establish the content validity of the scale, expert panel ratings of item clarity, conciseness, and relevance were used, and cognitive interviewing as a pretesting technique were used. Another aspect of validity that was addressed was construct validity, specifically, convergent and discriminant validity. This stage of the dissertation aimed to answer the following research questions:

3. Does the ECCA display adequate content validity?
4. Does the ECCA display internal validity? That is, does it conform to *a priori* expectations by forming eight factors corresponding to G*f, Gc, Gwm, Gl, Gr, Gv, Ga,* and *Gs?*

1.5 Design, Method, and Data Analysis

As delineated in Figure 1.2, the present dissertation was segregated into several phases to address the various aspects of validity that were of interest. First, meta-analytic reviews of the literature pertaining to the validity of parent- and teacher-reports of children's cognitive abilities were conducted, and their findings guided the development of the ECCA. Validity was assessed by the relationship between parent-/teacher-reports of children's cognitive abilities, and children's actual performance on standardised tests. Second, the ECCA scale was created following specific procedures to optimise the content validity of the scale, via the use of expert panel ratings and cognitive interviewing with members of the target audience. Third, the ECCA was administered to an initial development sample, and the internal validity of the measure investigated in two stages of data analysis, the first phase involving exploratory factor analysis (EFA) techniques to determine the initial dimensionality of the ECCA. The second phase was conducted with confirmatory factor analysis (CFA) to establish the measurement properties of the instrument at the construct level by conducting a series of single congeneric measurement models, and to consider the convergent and discriminant validity of the ECCA items with an overall structural model. This occurred using a structural equation modelling (SEM) framework, using a two-step method: the Holmes-Smith and Rowe (1994) approach to fitting complex models.

Participants in Study 3 comprised of primary and secondary school teachers who were currently teaching in either government, Catholic, or independent schools in Victoria, conveniently recruited through requesting participation from directly emailing school principals. The ECCA measure was in the form of an online questionnaire on Qualtrics (web-based survey platform), and was completed by participants in their own time.

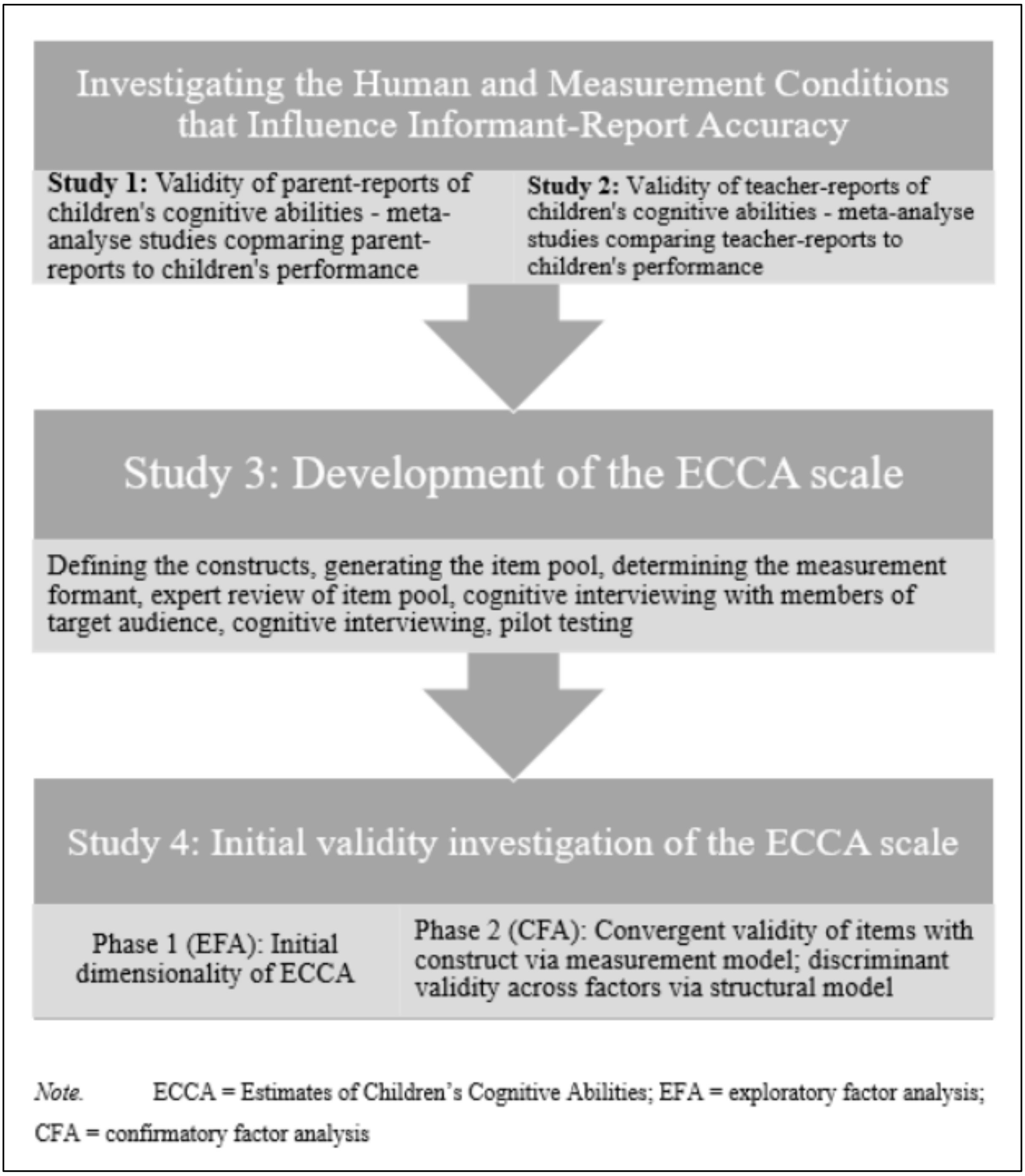


Figure 1.2 Overview of the studies and associated phases involved in the research project

1.6 Organisation of the Dissertation

In this chapter, the scope and rationale of the dissertation was outlined, and the planned design, methodology, and data analysis plan of the research project was described.

Chapter 2 presents an overview of the current literature that examined the traditional methods of measuring children's cognitive abilities, and the existing barriers that prevent certain segments of society from accessing such services. It then details how utilising parent and teachers as informants of children's abilities could not only provide a feasible financial solution to the often time-consuming and costly performance-based method, but has practical benefits within the educational context as well. Subsequently, a review of how parent and teacher ratings have been measured in the extant research is explored, followed by a discussion of the possible child-level, informant-level, and item-level factors that could influence the accuracy of parent and teacher ratings. Finally, the chapter concludes with a discussion of the theoretical and methodological limitations of existing parent- and teacher-report measures, and how future research can improve on such limitations to ensure that valid and reliable ratings from parents and teachers can be obtained.

Chapter 3 presents the methodology of Study 1, which was concerned with answering the first and second research questions of this dissertation – the validity of parent- and teacher-reports of children's cognitive abilities. The aim was to determine the validity of informant-reports by meta-analysing the existing literature on the relationship between parent (and teacher) estimates of children's abilities, and children's actual performance on standardised tests. The analyses further sought to identify the respective child-, informant-, and item-level factors that moderate the strength of the relationship between the two variables, either increasing or decreasing the accuracy of informant-reports. The human-related characteristics under investigation included the child's age and gender, the parent's gender, and teachers' years of teaching experience. The measurement conditions that were explored included the match between criterion measures and informant-report measures, the number of items on the informant-report measure, type of comparison used, and whether informant-report items referred to cognitive constructs or cognitive tasks. Subsequently, Chapter 4 provides an integration of the results found from the two meta-analytic reviews, and a discussion that ties the findings together. The measurement conditions shown to be conducive to the valid assessment of children's cognitive ability via parent- and teacher-report are highlighted.

Chapter 5 describes the development of the ECCA instrument that was the primary focus of the third study of this dissertation. The principles and recommendations made by DeVellis (2017) were closely adhered to throughout the development process. The constructs that were to be measured are in accordance with the CHC theory of cognitive abilities, in which the following eight broad abilities that are known to be important for academic achievement are delineated: *Gf, Gc, Gwm, Gl, Gr, Gv, Ga,* and *Gs*. The scale development process involved the following steps: defining the constructs to be measured, determining the response format, generating an item pool, consulting with an expert panel to review the item pool, cognitive pre-testing of the scale with members of the target audience, and pilot testing the finalised measure. The aim of this study was to establish the content validity of the ECCA, which was the third research question of this dissertation.

Chapter 6 presents the first phase of Study 4 that was concerned with the fourth research question of this dissertation; that is, to determine if the factorial structure of the ECCA sufficiently replicates *a priori* expectations of eight factors corresponding to the eight broad CHC cognitive ability areas under investigation. A brief discussion of exploratory factor analysis and confirmatory factor analysis precedes the rationale for using both methods within this dissertation. The methodology and data analysis approach for exploratory factor analysis is outlined before results are presented and discussed.

Chapter 7 presents the second phase of Study 2, in which the confirmatory factor analysis technique of structural equation modelling used was focused on determining the convergent validity of scale items with their respective ability constructs as delineated in CH theory, and the discriminant validity across factors contained in the overall scale. The method of data analysis used in this chapter is outlined before results are presented and discussed.

The final chapter of this dissertation, Chapter 8, is an integrative overview of the major findings of this dissertation and conclusions that can be drawn. Implications of these results are also detailed. Suggestions for what direction future research can take into the quest for developing psychometrically defensible parent- and teacher-report measures of children's cognitive ability that provide reliable and valid indications of children's performance in specific cognitive domains are provided.

Chapter 2

A Review of the Literature

This chapter will review the literature relevant to the accuracy of parent and teacher ratings of children's cognitive abilities. To set the context, the first part of this review will discuss the traditional methods of measuring intelligence and its associated limitations, and how such limitations could be addressed by using parents and teachers as informants of children's cognitive functioning. Next, the current literature pertaining to the accuracy of parents and teachers as reliable informants will be explored, followed by a discussion of the possible factors (child-level, informant-level, and item-level) that could influence the accuracy of parent and teacher ratings. Previous studies investigating the relationship between informant ratings and actual performance will be presented, with limitations and gaps in the extant research highlighted throughout.

The second part of the chapter will explore the current literature pertaining to methodological aspects of the scale development process. Issues surrounding the different theories of cognitive abilities that have been used in the development of informant-report scales will be explored, and the importance of using psychometrically-validated theories such as the Cattell-Horn-Carroll (CHC) theory of cognitive abilities as a foundation for the development of an adequate informant-report scale will be discussed. Further, the use of cognitive interviews to translate psychological constructs into layperson language in the scale development process will be investigated. Finally, general conclusions in light of previous studies will be provided, and suggestions for future research aimed at extending current knowledge in this field will be presented.

2.1 The Measurement of Intelligence

The question of how intelligence should be conceptualised and measured has been a hotly debated issue amongst psychologists for over a century. Amongst the differing approaches (e.g., biological and developmental approaches) developed to explain intelligent human behaviour, the methodology that has generated the most prominent research is based on psychometric testing (McGrew, 2005; Neisser et al., 1996). The psychometric approach is based on the presumption that intelligence is a measurable construct, and attempts to measure performance along dimensions (Hunt, 1995; Taylor, 1994). This method to understanding intelligence is not only the oldest, but is also viewed as the most established research-based technique that has produced efficient and practical instruments for measuring human cognitive ability (Neisser et al., 1996; Taylor, 1994).

All psychological constructs can be tested with measures of typical performance or maximal performance (Reynolds, Livingston, & Wilson, 2006). Tests of typical performance attempt to measure what a person does on a regular basis (Cronbach, 1949, 1960; Fiske & Butler, 1963), and are normally used to obtain an index of psychological constructs such as personality and behaviour. In contrast, tests of maximal performance attempt to measure an individual's fullest potential, or the best possible performance at a given task (Lamb, 2008), with intelligence and academic achievement being commonly assessed in this way. As proposed by Fiske and Butler (1963), the dominant method of obtaining an index of an individual's level of cognitive functioning has been through the administration of maximal performance measures due to two main reasons. Firstly, maximal performance measures are thought to obtain a "pure measure" (p.253) of ability that is determined mainly by capacity and with the removal of external influences. Secondly, maximal performance in a controlled setting is considered to be more stable "than performance under more life-like conditions" (p. 253).

Maximal performance-based assessments have thus been the dominant method of measuring an individual's level of cognitive functioning, since they function on the basis of behavioural observations considered to be purely reflective of one's true ability (Fiske & Butler, 1963; Jacobs, 2012). Psychological tests of intelligence constructs engage test-takers in "real-world" tasks, and test-takers are appraised in accordance to criteria that are central to actual performance in a day-to-day setting (Wiggins, 1998), and yield quantitative scores that are compared against standardised age or grade norms. Norm-referenced tests are just one of the many necessary components in any psychological assessment that help a clinician arrive at some hypotheses to make diagnostic decisions about an individual and their behaviour, personality, and capabilities (Framingham, 2016). Standardised tests of cognition are, therefore, useful tools that provide clinicians with tangible information that is critical to the development and implementation of interventions aimed at mitigating presenting referral concerns, and this information is also commonly used to determine educational placements and entitlements for students.

Some researchers claim that of all the social science variables that have been brought under investigation, intelligence tests are one of the best predictors of important life outcomes (Jensen, 1981; Neisser et al., 1996; Sattler, 1990), including academic achievement, psychosocial adjustment, and vocational success. Given the diversity in correlates, cognitive ability assessments have become a critical tool in the assessment of an individual's cognitive functioning, along with the prediction of associated behavioural, social, and economic consequences (Reschly, Myers, & Hartel, 2002). Cognitive ability assessment is also important within the educational context, as the process seeks to determine how well students are learning and is an integral part of the quest for improved education. The results of the assessment process can provide valuable feedback to students, educators, parents, policy makers, and the general public about the effectiveness of educational services. Hence, there is little doubt about the importance of cognitive ability assessment and the prominent role it plays in modern society.

2.1.1 Barriers to the maximal performance-based method

With all the advantages that intelligence testing brings to the educational scene, there are however, numerous barriers that prevent students that are likely to need it most from accessing cognitive ability assessments administered in the traditional, maximal performance-based method. Although cognitive assessments are offered at a minimal cost through government services in Australia, long waiting lists and having to meet numerous eligibility requirements imply that the prospect of a student undergoing an intelligence test in a timely manner through this service is not likely. For example, the Department of Education and Training in the state of Victoria provides an assessment service to support applications for the *Program for Students with Disabilities* (PSD) for the specific categories of Intellectual Disability and Severe Language Disorder only, and it is the responsibility of schools to identify such students and collect enough information to support a referral to the government's assessment service. This indicates that any student who does not fall into either diagnostic category who might still require a cognitive ability assessment (e.g., for giftedness, specific learning disability, etc.) would be unable to access this service.

Underdoing assessment through referrals to private psychology clinics is therefore the more common route that is taken in Australia. Unfortunately, the maximal performance-based method of assessment is not only typically lengthy, but also expensive, as it requires a trained psychologist to administer the measures (Jacobs, 2012). The Australian Psychological Society (APS) currently recommends an hourly fee of $251 for psychological assessment (APS, 2018). As in-depth testing could take place over several sessions, the total amount clients would have to pay can add up very quickly. The fact that psycho-educational assessment is often a time-consuming and expensive process indicates that certain segments of society who are likely most in need of these assessments, have a reduced capacity to obtain an adequate measure of cognitive ability (Jacobs, 2012). For example, research investigating links between socio-economic status (SES) and educational attainment have come to the conclusion that low SES adversely affects a range of educational outcomes, such as lower levels of literacy and numeracy, lower retention rates in school, and increased difficulty with academic performance (Considine & Zappalà, 2002). When educational difficulties become apparent, undergoing a cognitive ability assessment can help uncover the student's innate ability and identify cognitive strengths and weaknesses, and determine the necessity for, as well as type of, intervention that would be most beneficial to the child (Cockcroft, Bloch, & Moolla, 2016). However, the financial disadvantage faced by students from low SES backgrounds precludes them from routinely accessing psycho-educational assessments via the private route, and the restrictive criteria (as outlined above) that students must meet in order to obtain assessment through government-aided services means that those who need cognitive assessment most are likely not able to obtain one in a timely manner. As such, the failure to ensure the financial efficacy of psychological assessment is a real concern.

Additionally, as is characteristic of any typical maximal performance test, cognitive assessments do not occur in a natural-based setting. The nature of performance-based cognitive testing means that these tests are often administered within a highly controlled and structured test-taking environment. This is reflected in the administration instructions for commonly used performance-based measures such as the Woodcock Johnson measures and Wechsler scales, wherein examiners are advised to carry out testing in "a testing room that is quiet, comfortable, and has adequate ventilation and lighting" (Mather & Woodcock, 2001, p. 22). Test administrators are also required to build rapport, especially when working with children and adolescents, as it is "crucial to obtaining valid testing results" (Flanagan & Kaufman, 2004, p. 47) and prompts the test-taker's optimum test performance. However, the day-to-day classroom environment shares few similarities with the testing situation, indicating that scores obtained on cognitive assessments do not guarantee that the child will exhibit the same level of performance under everyday motivations and situations (Cronbach, 1949).

Despite criticisms that a single measure of intelligence, such as the widely used Intelligence Quotient (FSIQ), is inadequate to capture all of the differences in human cognitive ability, the FSIQ score is still widely depended upon for making decisions regarding educational placement and entitlements for gifted, as well as learning disabled and intellectually challenged students (Machek, 2003). For example, eligibility for disability funding from the Victorian government is dependent upon the student's FSIQ score being two standard deviations or more below the mean (i.e., 70 or below). Should the student obtain a score of 71, he or she is not eligible to access this funding. Additionally, learning disabilities have been diagnosed for decades using the IQ-achievement discrepancy model, which purports that children whose achievement scores are a standard deviation or more below their FSIQ scores are identified as learning disabled. However, a major limitation in terms of the model's clinical utility is that the discrepancy does not reveal where the child's weaknesses (or strengths) are, and subsequently, adherence to the IQ-achievement discrepancy model is unlikely to reveal substantial insight regarding the specific kind of educational intervention that might be most effective in helping the child learn.

The dependency on the FSIQ score as the sole measure in identifying children's cognitive functioning and diagnostic decisions of intellectual disability, giftedness, and learning disability is hence becoming a major concern amongst researchers and practitioners alike (Pfeiffer, 2003). Such criticisms of intelligence testing have thus brought about the notion of *intelligent intelligence testing* as first articulated by Alan Kaufman (1979), in which his philosophy of how intelligence testing should be applied within the educational context is reflected. As is reiterated nearly 30 years later, *The focus of the assessment is on the child, with … communication of the test results in the context of the child's particular background, behaviours, and approach to the test items as the main goals. Global scores are deemphasised, flexibility and insight on the part of the examiner are demanded, and the test is perceived as a dynamic helping agent rather than an instrument for placement, labelling, or other types of academic oppression. (Kaufman, Raiford, & Coalson, 2016, p. 5).*

This shift in emphasis on intelligent intelligence testing thus indicates a real need for not only a change in the way assessment scores are interpreted, but that the assessment process should also be child-centred.

2.1.2 The informant-based method as a solution

One way to circumvent these problems is to streamline the assessment process and focus on the cognitive domains most relevant to the referral concern, by using both parents and teachers as informants of the cognitive abilities of children referred for psycho-educational assessment. Parents have the unique opportunity to witness their child's growth and development across different life stages, and are able to observe their child's behaviour and accomplishments in various situations. In doing so, they are likely to be capable of providing reasonable judgments of the child's ability (Sommer, Fink, & Neubauer, 2008). Additionally, teachers are already the most frequently utilised source of information regarding school-related abilities for students, as they are able to observe first-hand the child's rate of learning and achievement in relation to that of the child's same-aged peers, thus making teachers experts on children's learning trajectories in the school setting (Spinath & Spinath, 2005). Hence, the information that could be gleaned from these adults that experience the child in different contexts is helpful in obtaining a more holistic view of the child's abilities (Geiser, Mandelman, Tan, & Grigorenko, 2016).

The relatively high-standing status and significance that parents and teachers hold in a child's life during the formative years also imply a likely possession of invaluable knowledge regarding the child's cognitive strengths and weaknesses. If reliable and valid reports from these adults could be obtained before formal cognitive assessment of the child commenced, a practitioner could justifiably exclude the measurement of some cognitive abilities that are not practically relevant to the referral concern. This would then ensure the economical efficacy of psychological assessments. Since the performance-based method is the gold standard for assessing cognitive ability, it is not recommended that information obtained from parents and teachers should be used as a substitute (Waschbusch et al., 2000), but instead should complement the results from a traditional intelligence test to make any diagnostic decisions. Using a multiple-rater tool to obtain information about a child's cognitive functioning also appropriately reflects the complexity of the multidimensionality of human cognitive abilities, in that the manifestation of specific abilities in children's behaviours can differ across contexts.

Additionally, in focusing on the abilities relevant to the referral concerns, the development of a valid parent- and teacher-report measure of children's cognitive abilities could not only help to reduce the financial costs of the performance-based method, but it could also provide a deeper understanding for these adults in learning how cognitive abilities are related to and manifested in their child's academic life. Arguably, this would provide more meaning than arbitrary numbers on a standardised test. Such a measure could also benefit schools and students in other practical ways. For example, it could be used with all students within a class to understand the cognitive profiles of students who would not necessarily require a traditional assessment. It could also be used as a screening tool to identify children at risk for learning difficulties, or could benefit from further traditional assessments to pinpoint their cognitive weaknesses. It can thus be seen that utilising parents and teachers as informants of children's cognitive abilities has the potential to provide a way to ensure the assessment process is indeed child-centred, and is advantageous in the educational context as well.

2.2 Parents and Teachers as Informants of Children's Cognitive Abilities

All psychological constructs are intangible abstractions that cannot be observed directly, and so are inferred and assessed through patterns of response behaviour in various settings (Teglasi, Simcox, & Kim, 2007). In the case of assessing children's functioning in these areas, parents and teachers are already considered and used as important sources of information; given their frequent interactions with the child, they could be thought of as experts on the child's behaviour. The reliance that both researchers and clinicians have on parent- and teacher-reports of children's functioning is evident in a wide variety of constructs, such as personality (e.g., Hierarchical Inventory Personality for Children; Mervielde & De Fruyt, 1999), attentional disorders (e.g., Behaviour Rating Inventory of Executive Function; Gioia, Isquith, Guy, & Kenworthy, 2000), and emotional functioning (e.g., Behaviour Assessment System for Children; Reynolds & Kamphaus, 2015). With research supporting the diagnostic utility of behavioural ratings provided by these adults (Bradstreet, Juechter, Kamphaus, Kerns, & Robins, 2017; Goldin, Matson, Konst, & Adams, 2014; McCandless & O'Laughlin, 2007), it is thus reasonable to suggest that a parent- and teacher-report measure of a child's cognitive functioning could inform practitioners' decisions in the same way.

However, the capacity of parents and teachers to provide reliable and valid estimations of a child's cognitive abilities is not clearly established in the extant literature. This could be due to the fact that practitioners typically do not obtain such information from these adults as an official measure of children's cognitive functioning, instead opting for the prevailing method of performance-based measures such as intelligence tests. Additionally, some could argue that the validity of an informant-report measure could be questionable, as such a scale is essentially seeking to obtain subjective opinions about a child's intelligence from others. As such, they are not truly objective measures of an individual's competencies and problems, and are "simply quantifications of adult opinions" (Reid & Maag, 1994, p. 348). Hence, the accuracy of such estimations could be vulnerable to the influence of non-cognitive and external variables (Jacobs, 2012). These factors are outlined subsequently.

The validity of instruments measuring most psychological constructs is usually evidenced by the presence of cross-informant concordance (Balsis, Cooper, & Oltmanns, 2015; Jansen, Bodden, Muris, van Doorn, & Granic, 2017; Sharp, Mosko, Chang, & Ha, 2010), or consistency in scores obtained on other well-validated instruments measuring similar constructs (Galvin, Roe, Xiong, & Morris, 2006; Tripp, Schaughency, & Clarke, 2006), as it is not possible to develop performance-based criterion measures of constructs such as personality and anxiety. However, in the case of cognitive ability, over a century's work in refining objective, reliable, and standardised intelligence tests has resulted in validated performance-based tests of cognitive ability which can be used as criterion measures when assessing how valid subjective reports of cognitive ability are (Jacobs, 2012). Hence, the accuracy and validity of informant-reports of cognitive ability in the current literature has most often been investigated by correlating informant-reports with traditional performance-based measures.

In general, empirical studies investigating parental accuracy yield high to medium correlations (i.e., above r =.50) between parental estimations and actual test results of children (Delgado-hachey & Miller, 1993; Hunt & Paraskevopoulos, 1980; Miller, 1986). However, there is little evidence to show that parents are consistent in their accuracy across different aspects of their children's development (Miller & Davis, 1992). It has therefore been concluded that while parents are typically better than chance in their estimations of children's abilities, they often show a marked tendency towards overestimation (Deimann & Kastner-Koller, 2011; Miller, 1988; Miller et al., 1991). Consequences of overestimation may include parents setting inappropriate and unrealistic expectations of the child, and because the child is unlikely to meet them, this subsequently could result in the development of feelings of powerlessness and failure.

To date, most studies on teacher judgments have focused exclusively on children's academic achievement (i.e., reading, writing, and mathematics ability), rather than their cognitive potential (Baudson, Fischbach, & Preckel, 2016). Overall, studies that investigate teacher accuracy have found moderate to high correlations (r = .50 to .90) between teacher ratings and measured student performance, implying that like parents, teachers are also relatively accurate in their estimations (Begeny, Krouse, Brown, & Mann, 2011; Demaray & Elliott, 1998; Feinberg & Shapiro, 2003; Ferguson, 2003). A recent meta-analysis of 75 studies conducted by Südkamp, Kaiser, and Möller (2012) reported a mean effect size of r = .63 between teacher judgment and student academic achievement, which is well-aligned with correlations reported by DeYoung's (2009) study of r = .45 and .80 between teacher judgments of student intelligence and measured FSIQ range. However, there are wide variations in the accuracy of individual teachers, as revealed by the results found by Helmke and Schrader (1987) of the correlations between estimates and criterion measures of academic achievement ranging from .03 to .90 for a group of teachers. Further, Hopkins, George, and Williams (1985) reported a range of correlations from .44 to .88 across individual teachers. These findings suggest that some teachers show very high judgment accuracy, while others, very low judgment accuracy.

Hence, the results found thus far are promising in that informant estimations of children's cognitive and academic functioning are largely accurate. However, the range of correlations found clearly suggests the existence of factors that impact the accuracy of these judgments, making them far from being objective, reliable and valid. Examining the extent to which parents and teachers are able to be accurate informants of children's cognitive abilities, and the extraneous factors that could contribute to greater inaccuracy, is important for a number of reasons. Firstly, being aware of the human characteristics and measurement conditions that influence these adults' estimations allows for the creation of a parent- and teacher-report measure that is of sufficient reliability and validity to be used as an adjunct (or in certain situations, as an alternative) to traditional methods of assessment. Beyond this, previous research has shown that the beliefs and opinions that parents and teachers hold of a child's cognitive functioning can shape the child's related self-beliefs and academic outcomes (Bleeker & Jacobs, 2004; Eccles & Wigfield, 2002), and influence their motivation and intellectual development (Dweck, 2000; Räty, Kasanen, Kiiskinen, Nykky, & Atjonen, 2004). These findings, and their significant implications for children's later life outcomes, thus indicate that the accuracy of parent and teacher reports warrant thorough evaluation.

Thus, differences in the validity of parent and teacher estimations should be further investigated to determine if these discrepancies are due to human characteristics such as child gender, parent gender, or teaching experience, or if they are more a function of the methodological differences in the way researchers have obtained these informant estimations. The following section details the human characteristics (at the child- and informant-level, respectively) and measurement conditions (at the scale item-level) that potentially have a significant impact on parents' and teachers' accuracy while reporting on children's cognitive abilities.

2.3 Human Characteristics that Influence Informant Accuracy

2.3.1 Gender of child

The general beliefs that parents and teachers hold about the relative competence of boys and girls could influence informant accuracy when reporting on children's cognitive abilities. As defined by McGarty, Yzerbyt, and Spears (2002), stereotypes are formed general beliefs about a particular social group that have a basis in reality. It has been highlighted throughout the previous research that stereotypes about social groups can function consciously and subconsciously to impact on people's expectations and perceptions of, and attitudes and behaviours toward others (Kunda & Spencer, 2003; Payne, 2006; Schmader, Johns, & Forbes, 2008). Although such views about social group differences may be formed on the basis of people's observations of, and interactions with, the world around them, they could give rise to unjustifiably negative perceptions of individual members of the stigmatised group (Wood, Kurtz-Costes, Okeke-Adeyanju, & Rowley, 2010).

Previous studies looking into parent beliefs of their children's intelligence has found evidence for gender stereotypes, such that boys are rated higher than girls in general ability (Furnham & Gasson, 1998; Furnham, Reeves, et al., 2002), and in specific ability areas such as mathematics and visual-spatial skills (Furnham, 2000). Although there is some evidence to show that on average, males outperform females on tests that measure visual-spatial ability (Halpern & Collaer, 2005; Kimura, 1999; Reilly, Numann, & Andrews, 2017), the same cannot be said for math ability, with previous research showing non-significant practical gender differences in math performance (Ganley & Lubienski, 2016; Hyde, Fennema, & Lamon, 1990; Isiksal & Cakiroghi, 2008). That these gender-stereotypical parental ratings are found even when there are no differences in performance or school grades between male and female children indicates that the erroneous beliefs that parents have of boys having higher capabilities than girls is unfounded.

As for daughters, parents are inclined to give girls better ratings than boys in the areas of literacy and verbal intelligence (Frome & Eccles, 1998; Furnham, Reeves, et al., 2002; Pajares & Valiante, 1997). Although this could still be considered as gender stereotypic thinking, their ratings are consistent with research that has shown female children and adolescents performing better than males in measures of spelling (Allred, 1990; Kimura, 1999; Reynolds, Scheiber, Hajovsky, Schwartz, & Kaufman, 2015), written expression (Fearrington et al., 2014), word fluency (Halpern, 2000; Kimura, 1999), and reading comprehension (Logan & Johnston, 2010; Lynn & Mikk, 2009). Taken together, the results indicate that parents may accurately perceive daughters to be better in literacy skills, but may underestimate their performance (or overestimate boys' performance) in math ability.

From a culturally differing perspective, Furnham and Fukumoto (2008) found that Japanese parents equally estimated their sons' and daughters' overall intelligence, unlike in Western societies where parents are likely to rate their sons as more intelligent than their daughters. It was proposed that this could be due to differences in cultural beliefs that effort and endurance, rather than intelligence, are more important than success (Furnham & Fukumoto, 2008). This finding suggests that culturally distinct beliefs relating to intelligence may result in decreased stereotypical parental perceptions of the capability between male and female children. However, definitive conclusions about this cannot be based on a single study, thereby highlighting the need for more research in cross-cultural contexts.

Earlier studies investigating the existence of gender differences in achievement-related beliefs in teachers have been found in the United States (Jussim, 1989; Jussim & Eccles, 1992) as well as in Europe (Tiedemann, 2000; Tiedemann & Steinmetz, 1998), where teachers believed boys were more competent in math skills than girls, even when there was no gender differences in standardised test scores. A more recent study by Hinnant et al. (2009) found that teachers tended to overestimate girls, and underestimate boys, in their ability to read. However, non-significant effects of gender on influencing the accuracy of teacher judgments of their students' math ability have also been found (Helwig, Anderson, & Tindal, 2001). Nonetheless, the majority of these studies were conducted in with middle class, mostly Western samples, calling into question the generalisability of these findings to other socioeconomic statuses and cultures. Hence, the effects of student gender on teacher ratings is still equivocal and as these findings are relatively dated, this emphasises the need for further empirical studies in this area.

Gender stereotypes in the context of education typically refer to the presumption that parents and teachers have of males and females differing in their traits, abilities and motivations (Schmenk, 2004). Research investigating this has mainly focused on stereotype threat, which describes the self-fulfilling nature of stereotypes in certain situations (Aronson & Steele, 2005). There is evidence of the negative impact of stereotype threat in the performance of females on math tasks (Nguyen & Ryan, 2008), and on boys in their reading, writing and math skills (Hartley & Sutton, 2013). Moreover, research has shown that these adults' biases towards boys have a differential impact by gender – positive effects on boys' achievement and enrolment into advanced level math courses in high school, and negative for girls (Lavy & Sand, 2015). These results indicate that gender biases from parents and teachers that is evident at the early stages of children's academic life in school can have long-term implications for vocational choices in adulthood, since enrolment into advanced math and science at high school is a prerequisite for tertiary education in specialised fields such as engineering and computer science.

Hence, even though there is evidence to suggest that performance in various ability areas varies by gender, it is important to highlight that these studies focus on group differences that are based on averages. The large variability around the mean, coupled with the small sex differences imply that it is indeed possible for either gender to contradict the general tendency. Additionally, the gender similarities hypothesis, as proposed by Hyde (2005) on the basis of large-scale reviews of studies, deems that boys and girls are more alike than dissimilar on many psychological variables, including academic skills such as reading and math. This suggests that the perceptions of these adults of children's academic capabilities that arise from stereotyping behaviours of males and females is erroneous and potentially damaging for children's self-related academic beliefs, motivations, and resulting academic performance.

2.3.2 Age of child

There have been inconclusive results in the existing literature regarding the age of the child being reported on having an effect on parent and teacher accuracy. It could be said that parent ratings of older children might be more accurate than younger children, due to a greater amount of feedback regarding their child's learning trajectory and academic progress that is naturally accumulated from more years of formal education. Miller et al. (1991) explored this variable by comparing parent estimations of second- and fifth-grade children, but did not find significant differences in accuracy as a function of the child's age. In this study, parent estimations were obtained by asking parents to make specific predictions of the point total that the child would earn on intelligence test tasks. This methodological aspect of the study may have been a limitation as it reflects a very stringent test of parental knowledge of the child's capability to complete tasks that would have been unfamiliar to the average layperson. Parental knowledge of a child's capabilities may thus be inadequately represented in discrepancy scores and correlations based on absolute differences between predicted and actual performance (Miller et al., 1991).

Conversely, parents of younger children may be more accurate when reporting on their child's abilities because in general, they are able to observe a higher proportion of the child's endeavours (Miller et al., 1991). In a study where parents were asked to estimate the gifted status of their four-year-old, it was found that 61% of children who were referred for a gifted assessment had an IQ of 132 or above (Louis & Lewis, 1992). This parental identification rate is consistent with past research concerning parents of pre-school children (Jacobs, 1971; Silverman, Chitwood, & Waters, 1986), and indicates that the majority of parents of intellectually gifted children become aware, in the early years, that their child is very bright (Louis & Lewis, 1992). However, the well above-average nature of the study samples suggests that this finding may not be generalisable to the typically developing child, and given that there are few studies that have investigated this variable with such a population, this indicates a gap in the research domain.

Looking toward the teacher-report domain, there has been no research examining whether there is an association between the child's age and how accurate a teacher is when reporting on the child's cognitive abilities. Previous attempts to synthesise the teacher-report literature to examine age effects has not been successful, as information on such student characteristics is inconsistently reported, and is therefore not readily comparable across studies (Südkamp et al., 2012). For example, only the range of student grade level was reported in several studies (e.g., Herbert & Stipek, 2005; Johansson, Myrberg, & Rosén, 2012; Maguin & Loeber, 1996), while others reported the mean age of the student sample without the standard deviation (e.g., Alvidrez & Weinstein, 1999; Commodari & Guarnera, 2005; Sommer et al., 2008). Additionally, much of the research to date has largely focused on preschool and primary school aged children. Future studies that includes students from higher grade levels would therefore be necessary to further explore the possible influence of children's age on teacher judgment accuracy.

2.3.3 Gender of informant (parent)

While some attention has been devoted to the influence of parent gender on the accuracy of their reports on their children's functioning, in studies looking at the accuracy of teacher-reports, very little information has been reported about the gender distribution within teacher samples. Therefore, it is not possible at this stage to consider teacher gender as a variable that may influence how accurately teachers can report on their students' cognitive abilities.

Given that gender stereotypes are apparent in children being rated (as was outlined in section 2.3.1 above), it is reasonable to propose that such biases may also exist within the parents as informants, in that there may be significant differences between paternal and maternal accuracy. Disparity between mothers and fathers in the types of information they provide about their child's emotional and behavioural functioning is known to exist (Hay et al., 1999), and this could also occur in the context of reporting on a child's functioning in cognitive ability areas. Further, mothers are typically treated as the primary source of information of children's cognitive functioning based on researcher and clinician perception that mothers are likely to be accurate reporters, but tend to neglect fathers as possible informants (Truetler & Epkins, 2003), as is apparent in the research domain of parent-report studies where mothers make up the lion's share of the parent participant samples (Bouffard & Hill, 2005; Brummelman, Thomaes, Nelemans, de Castro, & Bushman, 2014; Furnham & Bunclark, 2006; Maguin & Loeber, 1996). This is likely due to differing levels and types of parental involvement and interaction within a typical nuclear family structure, where for example, mothers may assist with homework tasks while fathers play games and sport. Thus, it could be said that fathers would have different data, rather than less data, on their children as compared to mothers. Additionally, fathers typically have less feedback on their child's performance in school, or have less contact with different children and so have fewer sources of comparison, which may lead to higher chances of paternal inaccuracy when estimating their child's ability (Chamorro-Premuzic et al., 2009).

In one of the first studies to investigate parents' beliefs about their child's cognitive development, Hunt and Paraskevopoulos (1980) found that mothers proved to be moderately accurate in predicting how well their child would perform on a battery of IQ items, as shown by a correlation of $r = .53$ between maternal prediction and child performance. This result was consistent with a later study conducted by Miller (1986), who found that mothers were very accurate ($r = .85$) in predicting their child's performance on an intelligence test. However, these studies did not include the children's fathers as participants, thereby precluding the ability to compare accuracy in judgments across parents of different genders.

As for research that included both parents as participants, Furnham and Gasson (1998) determined that the gender of the child was a better predictor than the gender of the parent in overall intelligence estimates. In contrast, a later study conducted by Furnham and Petrides (2004) revealed that fathers tend to give higher estimates than mothers for their child's general, analytic, and creative intelligences. However, these two studies did not compare parent-estimates with children's actual performance, so the effects of parent gender on parent accuracy were not investigated. In studies that did, it was found that over-estimation by fathers was more prevalent as compared to mothers and teachers as informants, resulting in fathers providing the least accurate estimations of their child's cognitive ability (Chamorro-Premuzic et al., 2009). Moreover, only a small to moderate mother-father agreement ($r = .22$ and $r = .35$) in the prediction of their child's performance on a standardised test was found in Miller et al.'s (1991) study, indicating that spouses are likely to have incongruent beliefs of their child's cognitive ability. Although discrepant results may indicate an offering of a unique perspective of the child's behaviour (Achenbach, McConaughy, & Howell, 1987), this calls into question the ability to confidently state that mothers and fathers are equally valid in their estimations of children's cognitive functioning.

2.3.4 Professional experience of teacher

It could be expected that teachers with more years of professional experience have more accurate judgments of students' cognitive abilities than teachers with less teaching experience. This is because such teachers usually have taught a variety of children with differing levels of cognitive and academic functioning, and thus would have a larger basis of comparison than less experienced teachers. Contrastingly, early career teachers are still in the midst of trying to make sense of their experience and have not yet developed extensive case knowledge of the degree to which children's learning profiles can vary (Mulholland & Berliner, 1992).

However, the extent to which the professional experience of teachers can affect their capacity to accurately report on students' ability levels is not clearly established in the extant literature. For example, in an earlier study by Babad (1985), it was shown that more experienced teachers tend to provide more accurate ratings of student's abilities. A study conducted twenty years later showed similar results, wherein consistent and clear gender biases against girls were apparent in the first part of physics teachers' careers, but declined with increasing teaching experience for Swiss, Austrian and German teachers (Hofer, 2015).

Mulholland and Berliner (1992) compared the accuracy of experienced teachers (practicing teachers with varying years of experience) and novice teachers (education students beginning field placements) in making predictions of student scores on standardised tests of reading and mathematics. Although experienced teachers were much more accurate than the novice teachers in their predictions, in contrary to expectations, there was no significant relationship found between years of experience and judgmental accuracy. That some of the most accurate judgments of student achievement ($r = .71$) came from practising teachers in the early years of their career (0 to 5 years of teaching experience) led the authors to postulate that the relationship between judgment accuracy and teaching experience could be nonlinear, and that further research was needed to better address the issue of teaching experience as a moderator by a including a larger sample of beginning teachers.

The inconsistencies found in the research domain suggest that although it is encouraging that teachers' judgment of student achievement tend to be accurate, there exists wide variation in the judgmental accuracy among individual teachers. What qualities separate the very accurate teacher from the very inaccurate teacher is a question worthy of further exploration, since teachers who cannot estimate their students' level of mastery over a content area with a reasonable degree of accuracy are far more likely to make erroneous instructional decisions for students.

2.4 Measurement Conditions that Influence Informant Accuracy

Although previous research indicates that the individual differences of informants, as well as those of the children being rated, can or does impact the validity of ratings of children's cognitive abilities, the way in which questionnaire items are presented has also been found to play a role. A variety of measures that differ in their technical design have been used to obtain ratings of children's abilities from parents and teachers. Such test characteristics, which are dependent on the methodological decisions made by scale developers or author(s) of the study, can be assumed to influence the strength of the relationship between informant reports and children's performance (Südkamp et al., 2012). Thus, the variability in item-level factors and how it could affect informant accuracy is worthy of investigation. To the author's knowledge, the potential influence of the following measurement conditions on the accuracy of parent and teacher reports of children's cognitive functioning have not been explicitly investigated.

2.4.1 Criterion test match (parallel vs. non-parallel)

As mentioned above, the predominant method that has been used by the extant literature to assess the accuracy of informants when rating a child's cognitive and academic ability has been to compare the relationship of scores obtained on an informant-report measure and scores obtained by the child on a performance measure. However, whether or not the informant-reported ability is theoretically parallel to the ability that is being objectively measured and compared with has not been given an equal amount of consideration as other variables have received within the extant literature. Within the current dissertation, whether a criterion test match is deemed to be parallel or non-parallel refers to the extent to which the abilities that informants are asked to report on, and the abilities that are assessed with maximal performance-based measures, are theoretically similar.

Theoretically, the congruence between what is being estimated and what is being measured could be said to be an important and somewhat obvious condition (Mabe & West, 1982) that would facilitate more valid reports from informants. For example, if the informant's task is to estimate a child's overall academic achievement, and the criterion measure is an assessment of the child's reading ability, it is reasonable to expect that the relationship between the two would be smaller than if the informant was asked to specifically rate the child's reading level. This measurement condition was first analysed within the self-report domain in Mabe & West's (1982) meta-analysis of the accuracy of self-reports of cognitive abilities, who hypothesised that self-reports would be more valid if criterion tests were parallel. Yet, the authors found no evidence to support the hypothesis, and only 10% of the effect sizes reported by studies included for the analysis were compared with non-parallel criterion tests (Jacobs, 2012). This would seem that concerns over the use of non-parallel criterion measures is unsubstantiated. However, a subsequent meta-analysis conducted by Jacobs (2012) investigated the use of parallel versus non-parallel criterion tests in self-reports of ability, and found significant moderating effects for this condition, such that self-reports of both general ability and specific abilities were considerably more valid when correlated with congruent performance-based measures. Jacobs (2012) concluded that the inclusion of effect sizes based on self-reports of ability that were compared with non-parallel tests likely diminished the summary effect size obtained in a concurrent meta-analysis of self-reports conducted by Freund and Kasten (2012).

Although comparisons could not be made for parent-reports since there have not been systematic reviews of the current literature, a similar finding has been found within the teacher-report domain. Südkamp et al.'s (2012) meta-analytic review included the congruence between the teacher's judgment task and the achievement test completed by students as a condition that could moderate teacher's accuracy in their judgments of students. The findings indicated that higher congruence between the two was associated with higher accuracy levels, and that a “mismatch” (i.e., use of non-parallel measures) lead to lower accuracy from teachers. These results provide support for the argument that ensuring the match between the informant-reported ability and objectively measured ability is an important measurement condition that facilitates valid assessments of children’s cognitive functioning by parents and teachers. Criterion tests should thus be appropriately matched such that the specificity of the informant-report measure and performance-based measure is parallel (Jacobs, 2012). Given that there has been no investigation about this measurement condition within the parent-report literature, and the surprising lack of discussion regarding the moderating effect of this measurement condition on teacher accuracy as acknowledged by Südkamp et al. (2012), this indicates a gap in the current literature that needs to be addressed.

2.4.2 Type of comparison (intra-individual vs. inter-individual)

The variability of human cognition is present both between people (inter-individual) and within a given person (intra-individual); as such, the range of methods used to develop informant-report measures of cognitive ability have used both types of comparisons. Both have been shown to produce variability in the validity of cognitive ability estimates. In the context of informants reporting on a child, this involves the respondent making within-subject comparisons, in that the respondent is instructed to compare the child’s ability in one area, with their abilities in other areas.

Inter-individual assessment in the context of informant-reports of children’s cognitive abilities requires the respondent to make between-subject comparisons; that is, to compare the child to his or her similarly-aged peers when making a judgment. Studies within the parent- and teacher-report domain have differed in their methods of operationalising inter-individual comparisons. The majority of studies have chosen to simply specify in either the instructions or the item stem of the questionnaire to compare the child with other children (e.g., “he is not capable of keeping up with children of his age”; Delgado-hachey & Miller, 1993), or displayed descriptive labels that imply the use of social comparison (e.g., "below average", "above average"; Sommer et al., 2008). Another method is to display a range of IQ scores as a normal distribution (Figure 2.1), as was first developed and used by Furnham and Gasson (1998), and subsequently by Furnham and colleagues in their empirical work relating to both self-estimates and other-estimates of intelligence (e.g,. Chamorro-Premuzic et al., 2009; Furnham & Budhani, 2002; Furnham & Mansi, 2014; Furnham & Valgeirsson, 2007; Kirkcaldy, Noack, Furnham, & Siefen, 2007). Respondents are typically presented with a bell curve of IQ scores, and a brief description about the distribution of IQ scores in the general population. Descriptive labels for each standard score at the middle (mean standard score of 100) and scores associated with up to three standard deviations to the left and right of the mean are also provided. For example, a score of 85 would be labelled *low average*, and a score of 130 would be labelled *superior* (Furnham, 2000). The respondents are asked to estimate the standard score they believe the child or student would obtain if assessed for that particular cognitive ability area.

According to Festinger's (1954) theory of social comparison processes, estimates that are based on inter-individual comparisons are likely to be more accurate than those made on intra-individual comparisons. This is because inter-individual comparisons require participants to respond based on a frame of social reference, while scales utilising the former comparison do not, and may only display labels with absolute terms such as *low ability* or *high ability* (Freund & Kasten, 2012). While studies within the domain of parent- and teacher-report have employed both methods of comparisons, the effect that this variable has on informant accuracy has not yet been explicitly investigated.

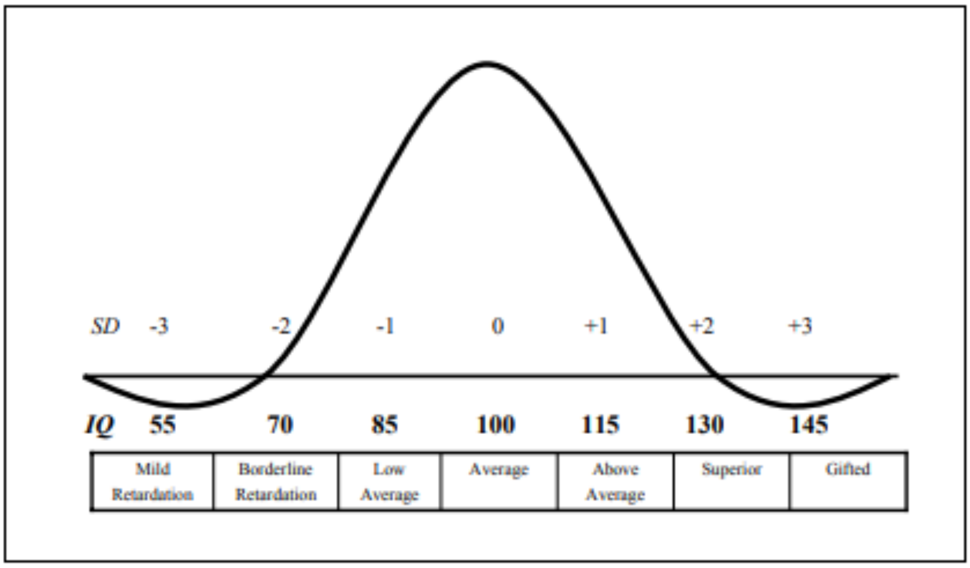


Figure 2.1 Normal distribution (bell curve) method commonly used in extant research to obtain parent- and/or teacher-reports of children's cognitive ability.

From Chamorro-Premuzic and Furnham (2006). Published by Taylor & Francis, London, England.

Looking instead to the self-report domain, it has been pointed out that scales providing social comparisons possess higher validity than scales that do not (Mabe & West, 1982). Likewise, it is anticipated that estimations made on the basis of relative terms will be more valid than those made on the basis of absolute terms. In the area of ability evaluation specifically, the issue of interest typically is not how much ability a person has in the absolute sense (i.e., intra-individual comparison), but rather, how much ability he or she has in comparison with other people with similar characteristics (i.e., inter-individual comparison). Although the identification of the sources and functions of intra-individual variation in cognitive abilities is important for understanding the child's development and learning profile (Siegler, 2007) and allows for subsequent planning of appropriate educational interventions, issues of inaccurate judgements in terms of underestimation (or overestimation) may arise if the parent or teacher perceives the child to have a weakness (or strength) in one ability area, despite their ability still being higher (or lower) than their peers (Jacobs, 2012). The use of social comparisons in measures used to obtain informant-estimates of children's cognitive ability is therefore proposed to be an important measurement condition that facilitates valid reports from parents and teachers.

2.4.3 Number of items (single vs. multi)

In addition to distinguishing between the types of comparison used, the parent- and teacher-report measures of children's intelligence developed and used in the present literature can be categorised according to whether the study authors have employed single- or multi-item measures in their design.

The single-item measure approach is appealing due to its shortened survey length and less complex development process, and it is less time-consuming and monotonous for research participants which may result in greater survey effectiveness especially when used with clinical populations (Hoeppner, Kelly, Urbanoski, & Slaymaker, 2011). From a psychometric perspective, single-item measures can also reduce the occurrence of common method variance, whereby seemingly valid correlations are found due to same response format rather than item content (Hoeppner et al., 2011). Use of single-items has been argued to be better suited to the measurement of unidimensional constructs that are narrow in scope (Sackett & Larson, 1990), because if the construct can be conceptualised as singular and concrete, it does not require multiple items to adequately represent it in the measure (Rossiter, 2002).

Despite the fact that cognitive ability has long since been determined to be a multi-faceted and complex construct, as theorised by many different cognitive ability frameworks including CHC theory, many informant-report measures developed or used by study authors employ the single-item design, when a multi-item measure may have been more appropriate. For example, the broad ability of comprehension-knowledge is comprised of several narrow abilities such as lexical knowledge, general verbal information, and communication ability

(Schneider & McGrew, 2012). In asking parents and teachers to estimate a child's level of verbal intelligence, the measure used by the study authors simply provided a brief description of the ability as a single item representing the construct (Chamorro-Premuzic et al., 2009). Although single-items may be favoured due to brevity and therefore practicality in reducing respondent refusal, they are also notorious for being unreliable, and are likely to produce construct under-representation as they are narrower in scope (Epstein, 1983). The use of single-items should thus be restricted to when the construct being measured is conceptually narrow or is unambiguous to the respondent, as it is only under such circumstances where they, in regards to reliability and construct validity, perform comparably to multi-item measures (Diamantopoulos, Sarstedt, Fuchs, Wilczynski, & Kaiser, 2012; Sarstedt & Wilczynski, 2009).

Thus, multi-item measures would seem to be the more appropriate tool to use in the measurement of cognitive abilities. Multi-item measures are known for increased reliability and validity as they help to establish the internal consistency of all the items in representing the underlying attribute, and average out errors and specificities that are inherent in single items (Bergkvist & Rossiter, 2007; DeVellis, 2003). Additionally, they also enable aggregation, which reduces the error variance associated with the inability of single-item measures to adequately represent a broad construct. By using multiple items per construct in a measure, its validity is consequently increased (Epstein, 1983; Jacobs, 2012). Moreover, in their assessment of the utility of single-item and multi-item measures, Sarstedt and Wilczynski (2009) reported that the latter generally outperforms the former with regard to reliability and validity, suggesting that the interpretation of findings from studies using single-item measures could be problematic. Since variability in results could arise due to this methodological factor, its potential influence on parent and teacher accuracy when reporting on children's intelligence also warrants investigation.

2.4.4 Item reference (cognitive construct vs. cognitive task)

The beginning of any scale development process is marked by the generation of items to assess a construct of interest (Hinkin, Tracey, & Enz, 1997). This process can be either inductive, by creating items first from which scales are then developed, or deductive, initially starting with a theoretical framework from which items are then produced (Schwab, 1980). Should researchers choose to employ the deductive method, it can then be inferred that the definition and presentation of cognitive ability items on parent- and teacher-report questionnaires is dependent on the specific type of theoretical framework of intelligence the researcher has chosen. It can be seen in the literature that the presentation of cognitive ability items on these questionnaires have differed in two ways. One way has been to describe the ability as a psychological or cognitive construct (e.g., verbal ability – the ability to use words). Another way is to word the item such that it reflects a behaviour that is indicative of the cognitive ability area (e.g., visual processing – seems very slow when using handwriting to copy material).

Most studies that have presented a "construct" representation of cognitive abilities as scale items to their participants have simply used the theoretical definition of the cognitive ability term(s) of interest. For example, in the questionnaire for parents and teachers to estimate students' various types of intelligences, items were presented as how they were technically defined, such as, "logical-mathematical (ability to detect patterns, reason deductively, and think logically)" (Hernández-Torrano et al., 2014). Findings from such studies have mostly indicated positive relationships between informant ratings and children's measured performance, suggesting that parents and teachers are accurate reporters. However, many researchers did not choose to follow a theoretical model in the construction of their scale items, and instead described abilities rather ambiguously. For example, Pesu, Viljaranta, and Aunola (2016) used a questionnaire consisting of items intended to measure informant expectations concerning the child's reading at school, such as "*How well do you think your* *child is doing at reading?*" Knowing that the ability to read is dependent on various types of basic skills (i.e., reading decoding, reading comprehension, and reading fluency), this scale item is ambiguous in that it could be interpreted in different ways by different reporters.

Moreover, if and when a theory has been used to guide item development, the most popular choice has been Gardner's (1983) Multiple Intelligences (MI) theory; a non-psychometric model. What is potentially most concerning about using this framework to investigate the validity of parent- and teacher-reports is that it does not possess standardised psychometric measures of the constructs contained within the theory (Visser, Ashton, & Vernon, 2006). The fact that the accuracy of parents' and teachers' estimates of children's cognitive abilities was determined from tests that were not specifically developed to measure the multiple intelligences suggests that the results of such studies is questionable in terms of its validity.

There exists a limited number of studies that have developed informant-report scales using a psychometric model of cognitive abilities as a guide. A teacher-report scale that is commercially available is the Children's Psychological Processing Scale (CPPS; Dehn, 2012), developed using CHC theory and neuropsychological theories such as Luria's (1970) theory of human processing. It is a measure of cognitive and meta-cognitive processes related to academic learning in children aged 5 to 12 years, and is designed to facilitate the identification of psychological processing weaknesses in children referred for learning disability evaluation. The CPPS is a worthwhile addition to the research domain of cognitive ability measurement tools, as it adequately addresses the need for the assessment of psychological processes to determine client eligibility for learning disability services.

Fiorello, Thurman, Zavertnik, Sher, and Coleman (2009) developed a scale intended for teachers that was based on tasks believed to be representative of seven CHC broad cognitive abilities believed to be important for academic achievement. However, instead of comparing teacher ratings to student performance to obtain a gauge of teacher-report accuracy, the researchers were interested in establishing the ecological validity of CHC abilities in daily classroom activities by comparing teachers' and school psychologists' perceptions of the importance of CHC abilities in the classroom context. Hence, this measure cannot be truly considered as a teacher-report scale of children's cognitive abilities.

Looking toward the parent-report domain, the scale developed by Waschbusch et al. (2000) utilised a "specific task" item reference, where the behaviourally-specific scale items were hypothesised to be indicators of the ability constructs contained within the extended *Gf-Gc* (Cattell, 1971; Horn & Cattell, 1982) differential theory of cognitive abilities. The reported correlations between the developed scale and psychometric measures of the corresponding cognitive ability factors ranged from .29 to .45, with the authors concluding that parents were able to report on their child's cognitive abilities with some degree of specificity (Waschbusch et al., 2000).

2.5 Theoretical and Methodological Limitations of the Current Literature

It is clear that the child-level, informant-level, and perhaps more importantly, scale item-level factors as mentioned above could all play a part in moderating the accuracy with which parents and teachers estimate children's cognitive abilities. Although it is important to be aware of the human-related characteristics that impact informant accuracy when reporting on children's abilities, these factors are a part of human nature, and are thus unchangeable. It is thus critical to determine the measurement conditions (i.e., the item-level factors) that have an influence on how accurate reporters are when responding to a scale, as such variables are controllable by the scale developer.

Systematic reviews are an efficient scientific activity that provide valuable insight to the field in which they are conducted, as they are able to establish whether findings across studies spanning across different contexts and time-points can be replicated and generalised, or whether findings significantly differ by particular subsets (Mulrow, 1994). The issue of validly interpreting the individual results of studies using small ungeneralisable sample sizes (e.g., Arciuli, Gurisik, & Munro, 2010) can be avoided by way of a meta-analysis, which increases the power and precision of studies' results by calculating a summary effect size of all studies within the same domain. To the authors' knowledge, there has been few attempts at synthesising the current literature regarding the accuracy of informants in reporting children's abilities (none for parent-reports, and only three for teacher-reports), a significant gap in the research domain.

Additionally, many studies (e.g., Phillipson & Phillipson, 2007; Pomerantz & Wei, 2006; Sink, Barnett, & Pool, 1993; Upadyaya & Eccles, 2014; Zakharov & Carnoy, 2015) within the parent- and teacher-report domain have used performance-based measures that are not standardised as a basis for comparison for informant accuracy, such as teacher-made tests or scholastic grades. Instruments that have been standardised, normed, and designed specifically for the measurement of cognitive abilities allow for uniform administration and scoring of responses by different examiners to minimise any influence different examiners may have on the skills being measured (Millman & Greene, 1993). Although teacher-made tests may effectively evaluate different skills that children learn in school, they (and subsequently, school grades) are susceptible to subjectivity and may not necessarily reflect the cognitive ability construct that is intended to be measured.

2.5.1 Theories of cognitive ability

An aspect of the scale development process that is also determined by the researcher is the theoretical framework that is used to guide the item-writing process, and is especially important when the deductive method for item-generation is applied. If the researcher chooses a framework that is not well-established and has inadequate empirical evidence to support the validity of the constructs contained within the theory, the developed items may be lacking in a solid theoretical foundation, which in turn, indicates that the content validity of the final scales may be questionable. Despite the prominence and extensive empirical evidence available in support of psychometric theories of cognitive abilities, much of the research within the domain of parent- and teacher-reports have not been based on any particular psychometric theory. In fact, it was often the case that no theory was explicated at all, instead having raters directly predict the child's performance on objective measures (e.g., Miller et al., 1991), or to make judgments on overall cognitive ability or academic achievement alone (e.g., Li et al., 2008).

One issue with having raters specifically predict how well a child would perform on a task is that the subtests in intelligence batteries are likely to be different to behaviours displayed and activities faced in everyday situations, and so could be unfamiliar to the respondents, thereby leading to an increased likelihood of inaccurate judgments. As for making judgments on overall cognitive ability, the construct is arguably a highly ambiguous one that could be interpreted in a number of different ways. Research that has explored laypersons' perspective and understanding of the concept of general intelligence has shown that people could vary widely on what they consider it to be – for example, some believe intelligence to be a primarily nonverbal reasoning and problem-solving skills, while others see verbal ability as a marker of intelligence (Horn, 1968; Sternberg, Conway, Ketron, & Bernstein, 1981). In the recent years, studies have also found that there are fundamental differences between people's conception of intelligence in non-Western cultures and those that have shaped Western intelligence tests (Benson, 2003; Cocodia, 2014). The variance in definitions of intelligence that exist within individuals thus indicates that simply asking people to rate a child's overall cognitive ability would likely be problematic.

As mentioned previously, the FSIQ score has been purported to be a measure of general intelligence for over a century, and with this assumption comes a prominent role in modern society. The FSIQ score has traditionally been utilised as an index of typical development, facilitating clinical diagnoses, used to explain differences in achievement and as a predictor of future success in important life outcomes, and for exploring causes of individual differences in cognitive ability (Richardson & Norgate, 2015). However, the research domain is fraught with perpetual debate and controversy about the clinical utility of the IQ score, with some claiming that the FSIQ score is still valid in the presence of variability (i.e., strengths and weaknesses) in specific cognitive abilities (Daniel, 2007; Watkins, Glutting, & Lei, 2007), while others assert that the FSIQ is a less viable predictor of achievement as this variability increase (Fiorello et al., 2007). Although the moderate correlations that have been reported within the extant literature (e.g., Jensen, 1998; Mackintosh, 1998; Naglieri & Bornstein, 2003) is indicative of the reasonably good predictor that FSIQ is of academic achievement, 50-75% of the variance in academic achievement is not accounted for by FSIQ score (Rohde & Thompson, 2007), indicating that there needs to be a shift in focus from general intelligence to specific aspects of cognitive processing in order to better explain differences in academic achievement.

Additionally, much has been said about the irrelevance of the FSIQ score to the definition and diagnosis of learning disabilities (e.g., Klassen, Neufeld, & Munro, 2005; Naglieri & Reardon, 1993; Siegel, 1989), with many agreeing that the cause of learning disabilities is likely the result of a deficit in a specific narrow ability area, such as phonological processing for dyslexia, instead of a simple discrepancy between FSIQ and achievement scores. This indicates that in the context of obtaining parent and teacher-reports of children's functioning, it would be more worthwhile for researchers to develop items that are reflective of specific cognitive abilities instead of simply asking informants for their opinion on the child's general intelligence.

2.5.1.1 Non-psychometric models of cognitive abilities

Theories of intelligence have evolved through a series of paradigms that have been proposed to clarify our understanding of the phenomenon. Although most of these have been data-driven (e.g., psychometric approaches that seek to understand the structure of intelligence have been based on and tested by the use of data obtained by ability tests), others such as MI theory (Gardner, 1983) have been derived in non-psychometric ways, and there are a sizeable number of studies that have used this theory in their parent- and teacher-report studies (e.g,. Furnham, 2000; Furnham & Bunclark, 2006; Furnham & Petrides, 2004; Furnham, Reeves, et al., 2002; Hernández-Torrano et al., 2014). MI theory proposes the existence of seven distinct intelligences, namely: linguistic, logical-mathematical, spatial, musical, bodily-kinesthetic, interpersonal, and intrapersonal intelligence (Gardner, 1983). MI theory emphasises an inherent system of interacting abilities that are combined into patterns which represent people's different forms of expertise (McGrew, 1993). The development of this theory was guided by Gardner's observations of people in their environment (Gilman, 2001), a similar approach utilised in the development of Freudian personality theories, and is in sharp contrast to how intelligence has traditionally been conceptualised and delineated.

The conception of MI theory was brought about from the desire to explain the variety of performance outcomes individuals can produce and to establish the ecological validity of human intelligences, and this framework has been found to have considerable appeal, particularly within the educational setting (Campbell, 1997; Chen, 2004; Gardner & Hatch, 1989). However, there has been no empirical evaluation of the validity of the theory as a whole (Allix, 2000; Sternberg & Kaufman, 1998), and it has been found to have inadequate empirical support (Waterhouse, 2006). Further, MI theory does not possess a set of testable psychological subcomponents for each of the seven intelligences (Allix, 2000), indicating that there is no standardised performance-based assessment tool that measures the constructs represented in MI theory. Consequently, the validity of the results of previous research within the parent- and teacher-report domain that have utilised this model are questionable. On the other hand, MI theory does highlight that instead of reducing intelligence to a single number (i.e., FSIQ score), a more complex model is required to fully capture the full spectrum of human cognitive abilities.

2.5.1.2 Cattell-Horn-Carroll theory of cognitive abilities

Fulfilling the need for a cognitive model that encompasses a wide range of human abilities is the empirically derived Cattell-Horn-Carroll (CHC) theory of cognitive abilities, currently the most well-established psychometric theory of cognitive abilities (Flanagan & Dixon, 2014; Flanagan & Ortiz, 2001; McGrew, 2009; Schneider & McGrew, 2012). The reader is directed to Chapter 1 (section 1.2) for a visual representation of the model and definitions of broad cognitive ability areas. Due to the presence of an impressive body of empirical research, including developmental, heritability, and neurocognitive evidence, in support of the model (Flanagan, Ortiz, & Alfonso, 2007), CHC theory is currently used as the basis for selecting, organising, and interpreting tests of cognitive abilities (Alfonso, Flanagan, & Radwan, 2005).

Since its conception, CHC theory has been applied extensively within the educational realm, with research pertaining to cognitive-academic relations highlighting the significance of seven of the 16 CHC broad abilities currently contained within the model in the successful acquisition of reading, writing, and mathematics skills (Jacobs, Watt, & Roodenburg, 2013). Recent revisions of the theory (Schneider & McGrew, in press) include the replacement of what was known as long-term storage and retrieval (*Glr*) with two separate broad ability constructs: learning efficiency (*Gl*) and retrieval fluency (*Gr*), thus bringing that number up to eight. Identifying a person's strengths and weaknesses in these important broad abilities (e.g., auditory processing, fluid reasoning, learning efficiency, retrieval fluency), and narrow abilities (e.g., phonemic awareness, induction, associative memory, and rapid automatised naming) would add significant explanatory power to overall IQ measures when predicting achievement (Fiorello et al., 2009), and facilitate the fine-tuning of interventions to address academic difficulties accordingly.

Further, behaviourally-specific correlates of the CHC broad abilities that are important for academic achievement have been explored through the work of Flanagan, Ortiz, et al. (2013), and Fiorello et al. (2009). It is argued that although psychologists would likely understand the ability areas in a way that is consistent with how they are delineated as hypothetical psychological constructs, parents and teachers would be more inclined with wanting to comprehend the practical implications of these constructs for children's learning processes (Fiorello et al., 2009). The operationalisation of the CHC abilities through observable daily tasks would, in part, begin to determine the ecological validity of the theory, as well as play a role in parents' and teachers' willingness to accept and implement CHC-based educational strategies in the home and school setting (Fiorello et al., 2009).

Thus, the development of CHC theory has subsequently brought about the expansion of our current understanding of the full range of human cognitive abilities that are important for academic achievement. Despite CHC theory being generally regarded as the most robust psychometric theory of cognitive abilities currently in existence (Alfonso et al., 2005; McGrew, 2009), its application within the research domain of parent- and teacher-reports of children's cognitive abilities has been extremely limited, including the development of adequate, reliable and valid informant-report scales based on CHC theory. Considering the strong theoretical reputation of the framework, this gap in the literature needs to be addressed.

2.5.2 Translating psychological jargon to layperson language

Another methodological issue often faced by scale developers, especially those in highly specialised fields such as psychology, is the translation of technical jargon to layperson language. Laypeople do not come across terms such as “fluid reasoning” and “retrieval fluency” in everyday situations, so it is likely that their first encounter with such technical terms would be in the survey they are responding to. Additionally, many words in the English language have numerous meanings, or are interpreted differently by people in various contexts and cultures. Pretesting scale items with a select group of participants who are representative of the target audience of the measure could be useful for addressing such quality issues, and of the various pretesting techniques (e.g., focus groups, usability testing, etc.) available, cognitive interviews have been touted as an effective technique that allows each item to be examined individually in an in-depth way that reveals the respondent’s thought processes.

Cognitive interviewing is the administration of drafted survey questions to respondents who are part of the target population, in order to collect supplementary qualitative information about their cognitive processes while answering the question, and to gain insight into the quality of the questions in order to improve them (Beatty & Willis, 2007; DeMaio & Landreth, 2004). The technique has been described as an essential tool to explore specific aspects of survey questions, to find out how respondents interpret the items, and how well they comprehend the task involved in answering the item. As such, the process can reveal valuable data regarding the quality of the responses collected (in terms of whether the participant is answering the question in the way that was intended by the researchers), and any difficulties the participant encountered when answering the items (Beatty, 2004). As a semi-structured interview, cognitive interviewing can be used as a means of pre-testing and modifying survey items, and allows scale developers to create items that are comprehensible to the target population.

The methodology is known to be effective in remedying common threats to survey validity, most of which stem from the complexity of phenomena that the researcher seeks to capture in a survey instrument, and the possibility that respondents may answer in a socially desirable way (Desimone & Le Floch, 2004). The background theory underlying cognitive interviewing is attributable to Tourangeau (1984), and its basic tenet is that responding to survey questions requires a complex set of cognitive processes (Tourangeau, Rips, & Rasinki, 2000). These include comprehension of the question, the retrieval of relevant information in memory, decision processes used while answering the question, and the matching of an internally generated answer to the response categories provided. Any one of these four aspects could cause difficulties for a respondent, and the cognitive interview exercise allows researchers to determine which scale items are causing disruptions to the respondents' cognitive processes.

Examples of the common issues that result from the disruption of the cognitive processes required from the survey respondent are nonresponse and non-completion of questionnaires, which leads to the collection of incomplete data and may affect the generalisability of survey findings (Drennan, 2003). Central to cognitive interviewing is the verbal probing technique that can help researchers uncover the specific problems in scale items that are causing such issues. Open-ended questions such as "*what does the term 'solve logic puzzles' mean to you?*" are used to "probe" further into the basis for the response (Willis, 2005). They are designed by the researcher prior to the interview, and aim to assess item comprehension and clarity, assess item relevance to the intended construct, and to seek elaboration the respondents' thought processes while answering the scale items (Willis & Miller, 2011).

Given the utility of cognitive interviewing and the potential it has to improve the quality of scale items, it is surprising that its application to scales developed to obtain parent and teacher reports of children's cognitive abilities is scarce. Looking to the measure used in Furnham's (2000) study, an example of the definition of a specific ability that is presented to respondents to rate reads, "*verbal or linguistic intelligence (the ability to use words)*". Although the language used may easy enough to understand, a child's ability to use words in speech and to use words in writing could differ, and it is not immediately obvious in the above definition which "ability" they are to estimate. The additional step of conducting a cognitive interview with a member of the target audience to pre-test the quality of this scale item may have helped to avoid this issue in the development phase.

The content validity of a newly developed scale is typically established through the use of an expert panel review of the scale items. However, there is an abundance of evidence in other research domains showing the feasibility of cognitive interviewing as a pre-testing technique integrated within the scale development process to ensure the minimisation of response error by understanding the cognitive processes respondents go through when answering questionnaires. An example is the Hierarchical Personality Inventory for Children (HiPIC; Mervielde & De Fruyt, 2002), which was developed using a "bottom-up" strategy involving the researchers interviewing parents by using neutral prompts (akin to verbal probing) to elicit descriptions of behaviour that was characteristic of their child. The transcribed interviews were further segmented into a personality descriptive category system that guided the creation of HiPIC items. Subsequent studies investigating the replicability and validity of the measure showed that the HiPIC as "a most comprehensive personality inventory today assessing individual differences in children" (Mervielde & De Fruyt, 2002, p. 142). There is thus a need for future research looking to develop a reliable and valid parent- and teacher-report measure of children's cognitive abilities to apply pretesting techniques such as cognitive interviews to address this gap in the research domain.

2.6 Conclusion

The traditional method of measuring human cognitive abilities using tests of maximal performance and the prominent role that it plays in modern society has remained largely unchanged for the past century. However, there is a large segment of society who are likely the most in need of assessment that is unable to obtain cognitive ability assessments due to its tendency of being a lengthy and expensive process. These barriers to the maximal performance-based method indicate that it is not one without flaws. Additionally, there has also been an over-reliance on the FSIQ score to make diagnostic conclusions for the cognition-related disorders of intellectual disability, giftedness, and learning disabilities, and the subsequent educational placement and intervention for those diagnosed. The clinical validity and utility of the FSIQ score as the sole deciding factor in making such decisions that impact children's future life outcomes has been questioned for years by both researchers and practitioners alike, signalling that the transition to intelligent intelligence testing that captures a more holistic view of human abilities is required.

Since parents and teachers of children being referred to psychologists are already used as reliable informants of children's personality and emotional and behavioural functioning, it makes sense to suggest that they could be utilised as reporters on their cognitive functioning as well. The information that these important adults could provide about the day-to-day performance of cognitive tasks and how this translates to the child's academic life could be a valuable complement to the results from the performance-based method. However, the question remains as to whether these adults are able to provide reliable and valid information about a child's cognitive strengths and weakness. The relationship between parent and teacher reports of children's cognitive abilities and children's performance on standardised cognitive ability assessment has been investigated extensively, with generally consistent results. Though positive and significant relationships have been found repeatedly for both types of informant-report, indicating that parents and teachers are typically above chance in their estimates, the moderate correlations signify that they are far from perfect in assessing a child's cognitive strengths and weaknesses. In fact, they have the tendency to overestimate what the child is capable of, and such erroneous perceptions of a child's intellectual capacity could result in the child not being able to meet the unrealistic expectations set by the adult, which could have a negative impact on the relationship between them and the child's self-confidence.

In an attempt to explain what makes an informant under-estimate, over-estimate, or accurately estimate a child's cognitive abilities, researchers have looked at factors such as the age of the child, the gender of the parent or the teacher's professional experience. A much smaller percentage of the literature, however, has considered the methodological decisions made by researchers in how they obtain informant-reports. It is important to remain cognisant of the fact that the child and informant characteristics are invariably unchangeable. Thus, it is critical to determine the measurement conditions that are conducive to accurate estimations, since factors such as the number of items per construct that are represented in the scale, whether respondents are to use a frame of reference in their comparisons, and how items are presented to respondents, are within the researcher's control.

Once the ideal measurement conditions are determined via a meta-analytic review, this information could be used to develop a parent-report scale that reduces the influence of the informant and child characteristics. One way to do so is through the development of scales that are comprised of behaviourally-specific items that are manifestations of the cognitive abilities in question. The operationalisation of cognitive constructs into observable tasks could provide the parent and teacher informants with clarity on what exactly they are reporting on, via a specific framework on which to base their estimations. Further research should therefore endeavour to use psychometrically validated theories of cognitive ability, such as CHC theory, that contain reliable and validated performance-based operationalisations of the constructs contained within the theory. Such a move would likely produce more reliable and valid results when investigating the accuracy of parent-and teacher-reports of children's cognitive abilities.

Further, the practical implications of developing a valid parent- and teacher-report scale are numerous. The results of the measure also could be used by teachers to obtain an overview of the cognitive profiles of students and subsequently develop appropriate individual learning plans for them, or by the students themselves when wanting to develop study strategies that take into account the peaks and troughs of their cognitive profile. Additionally, giving parents the active role as an informant of their child's cognitive functioning could not only empower them as active agents in the assessment process, but they would also be closer to understanding how their child's cognitive abilities are manifested in their academic life.

Chapter 3

Meta-Analyses: Methodology

3.1 Rationale for Conducting Meta-Analysis

The literature review presented in the previous chapter outlined the human characteristics that have been shown to affect the accuracy of parent- and teacher-reports of children's cognitive abilities. Much less attention has been paid, however, to the measurement conditions that facilitate valid estimations from parents and teachers, a significant gap in the literature given that these methodological decisions are controlled by the researcher. Additionally, the limited number of systematic reviews of the parent- and teacher-report research domains indicates that there is little integration of existing information and thus, whether or not findings are consistent and generalisable across populations, contexts, and time is not clearly established (Mulrow, 1994).

To address these gaps in the current research, the rigorous and structured approach of meta-analysis was used in this dissertation. Meta-analysis was first defined by Glass (1976) as "the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings" (p. 3). It is a subset of systematic reviews that quantitatively and systematically assesses the results of two or more comparable studies to obtain an overall estimate of the effect, and so allows for the derivation of conclusions about that body of research (Cumming, 2012; Fagard, Staessen, & Thijs, 1996; Haidich, 2010). Additionally, it is postulated by Schmidt and Oh (2016) that meta-analysis is an effective solution to the supposed replication crisis evident in research, since by definition, its methodology is based on multiple replications of studies.

Lipsey and Wilson (2001) discuss several advantages that meta-analysis has over other types of systematic literature reviews such as narrative and scoping reviews. First, identifying important effects and the study conditions that lead to them is more likely with the meta-analysis approach, which includes the ability to pool effect sizes across studies and the systematic coding of study characteristics. The former indicates that the statistical power of a meta-analysis is considerably greater than that of any individual study; the latter allows for the statistical identification of reasons (e.g., participant and/or method characteristics) behind observed variation in effect sizes between studies. Additionally, the meta-analysis approach requires the researcher to state clearly the method of locating studies to be included in the review and coding of study characteristics. The explicit and systematic approach that is imposed by meta-analytic techniques thus facilitates structure in the evaluation of the results obtained, indicating that the validity of the procedure, evidence, and conclusions can be determined by the reader. Finally, a representation of the aggregated effects across studies is possible from coding the magnitude as well as the direction of the effects in each individual study. Important variations in the strength of the statistical relationships reported are therefore taken into account instead of the exclusive focus that qualitative systematic reviews tend to have on the statistical significance of results. Drawing conclusions based solely on whether the null hypothesis was rejected can mislead conclusions; likewise, ignoring the effect sizes can lead to inappropriate recommendations if a statistically significant result is not practically significant.

3.2 Overview and Hypothesis

This chapter presents the processes and methods used to conduct comprehensive meta-analytic reviews of the literature concerning informant-reports of children's cognitive and academic abilities. Specifically, the ability areas under investigation are cognitive constructs contained within CHC theory, namely: Fluid Reasoning (*Gf*), Comprehension-Knowledge (*Gc*), Short-Term Working Memory (*Gwm*), Learning Efficiency (*Gl*), Retrieval Fluency (*Gr*), Visual-Spatial Processing (*Gv*), Auditory Processing (*Ga*), and Processing

Speed (*Gs*), as well as general intelligence *g,* and the academic abilities of reading (*Grw-R*), writing (*Grw-W*) and mathematics (*Gq*). The reasons for focusing on these ability domains in this dissertation were outlined in a previous chapter (section 1.2). The validity of parent- and teacher-reports of *g* was also investigated to allow for the comparison of the validity of informant-reports of general versus specific cognitive ability. In addition to synthesising previous study findings to arrive at psychometrically sound estimates of the overall validity of informant-reports, potential child-level, informant-level, and item-level moderators that could influence the size of the correlation between informant-reports and children's actual performance were also evaluated.

3.2.1 Hypotheses

The aim of the meta-analyses was to investigate and compare the validity of parent-reports and teacher-reports of children's cognitive and academic abilities. Based on the research reviewed in the previous chapter, the following were hypothesised:

H1: The summary effect size for the relationship between parent-reports (teacher-reports) of children's ability and performance measures of children's ability will be statistically significant and positive.

H2: Significant heterogeneity in the effect sizes will be found, thereby necessitating the investigation of potential moderators.

H3: The validity of parent-reports (teacher-reports) of children's ability that are correlated with parallel criterion measures will be greater than those correlated with non-parallel criterion measures

H4: The validity of parent-reports (teacher-reports) of older children's ability will be greater than those of younger children

H5: The validity of teacher-reports of children's ability from more experienced teachers will be greater than those from less experienced teachers.

H6: The validity of parent-reports (teacher-reports) of female children's ability will be greater than those of male children's ability.

H7: The validity of parent-reports of children's ability from mothers will be greater than those from fathers.

H8: The validity of parent-reports (teacher-reports) of children's ability obtained from multi-item measures will be greater than those obtained from single-item measures.

H9: The validity of parent-reports (teacher-reports) of children's ability obtained from measures utilising inter-individual comparisons will be greater than those obtained from measures utilising intra-individual comparisons.

H10: Parent-reports (teacher-reports) of children's ability based on items that describe cognitive test tasks will have greater validity than those based on items referring only to cognitive constructs.

The following section outlines the methods used to conduct the meta-analyses. It begins with a description of the systematic search strategy for studies with relevant data, followed by the eligibility criteria applied for inclusion and exclusion of studies. Next, the procedure for data extraction and analytical issues pertaining to the calculation and extraction of effect sizes are outlined. Finally, the process of the meta-analysis and the associated statistical decisions made are described.

3.3 Systematic Search Strategy

A critical feature of the proper application of meta-analytic techniques is the development of systematic and explicit procedures for the identification of relevant studies to be included for the review; in being systematic, the procedures reduce bias (Pettiti, 2000). Comprehensive literature searches for the two research domains were conducted using three electronic databases in the fields of psychology and education: Scopus, PsycINFO, and Web of Science.

Relevant research was searched for using pairwise combinations of informant terms (parent or teacher), report terms, cognitive ability terms, and academic ability terms. Details on how these search terms were combined and entered into the relevant databases is elaborated on in a subsequent chapter of this dissertation. Since there has been no meta-analysis on studies pertaining to the accuracy of parent-reports, and to obtain comparability of findings to the published teacher-report meta-analyses of Hoge and Coladarci (1989), Südkamp et al. (2012), and Machts, Kaiser, Schmidt, and Möller (2017), studies that were published after 1980 were included. Several of the search terms were truncated to allow for the identification of additional terms, marked by an asterisk. For example, the search term *rate* was truncated to allow for identification of *rated* and *ratings*, and the search term "intell" was truncated to allow for identification of *intelligence, intellectual*, and *intellect.* The term "ability" was used only in conjunction with "cognitive" or "academic", as entering "ability" on its own primarily resulted in the identification of topics that were too broad, such as reports of children's executive functioning, metacognitive skills, and social ability. The informant term was also only used with the report terms, marked by quotation marks (e.g., "parent perce*", "teacher rat*") because entering each term on its own resulted in the return of studies that were not relevant, such as the impact of parental involvement on children's school performance, and teacher judgments of students' school adjustment.

Each article identified during the search procedure was subjected to an initial screening, where the title and abstract was examined for indications that both parent-reports or teacher-reports and performance measures of cognitive ability were obtained in the study. What was considered an adequate measure of *cognitive ability* was loosely defined at this stage of the study identification process, so that studies were not prematurely excluded as a result of specific definitions of cognitive ability. Whether or not the informant-report item or measure could be considered as one of cognitive ability was determined via a classification system used by two separate coders working independently. This system and process is outlined subsequently. When the initial screening of the abstract suggested the possibility of inclusion as data, the entire article was obtained for a more detailed evaluation.

To ensure the literature search was as comprehensive as possible, other additional search strategies were used. The ancestry method was used to identify any additional articles through a search of the reference lists of the retained articles. Further, the recently published meta-analytic reviews of the accuracy of teacher-reports of students' academic achievement (Südkamp et al., 2012) and students' cognitive abilities (Machts et al., 2017) were scanned for additional teacher-report articles that had not been located using the pairwise combinations of search terms.

3.4 Inclusion and Exclusion Criteria

After studies with relevant information have been identified, the next step is to define the eligibility criteria for inclusion in the meta-analysis. The goals for defining clear criteria is to form an operational definition of the research problem (Abrami, Cohen, & D'apollonia, 1988), thus ensuring the reproducibility of the meta-analysis and minimising bias in selection of studies for the review (McDonagh, Peterson, Raina, Chang, & Shekelle, 2013).

3.4.1 General characteristics of the study

For this dissertation, any study published in English that investigated the relationship between parent reports and/or teacher-reports and children's actual performance on a standardised cognitive and/or academic achievement test was included. Studies that used other psychological or psychosocial constructs (e.g., motivation, anxiety, etc.) for comparison were excluded. Studies that were conducted in the field (i.e., not case descriptions or computer simulations), and explicitly instructed the informants to rate a child's abilities were retained for inclusion. Further, studies that only asked the informants to rate their confidence in their estimations were excluded from the analysis.

3.4.1.1 Child sample

Given the focus on teacher-reports in this dissertation, only studies that recruited samples comprising school-aged children (i.e., five to 18 years of age) were included. Studies that recruited university students or kindergarten children were excluded. Studies were also excluded if child participants were drawn from disordered or clinical populations (e.g., intellectually disabled, acquired brain injury), as the intention of this meta-analysis was to draw general conclusions regarding the validity of parents and teachers to accurately report on typically developing children.

3.4.1.2 Criterion measure

The measurement of children's cognitive and academic abilities had to have been assessed using standardised tests. That is, tests that were administered in a standardised way, were objectively scored (manually or electronically), and offered norms that allow comparison of the test-taker's score with that of a normative sample. These test characteristics are similar to those defined by Freund and Kasten (2012) and Jacobs (2012) in their meta-analytic reviews of self-reports of cognitive abilities. Accordingly, studies that compared informant-reports to non-objective and/or non-standardised measure of ability, such as school grades or school examination results, were excluded. Selected studies also needed to provide effect sizes of the comparison between informant rating and criterion measure as a correlation coefficient, or at least be transformable to one. Lastly, the time between parent and/or teacher ratings and student testing had to be within one school term since the aim was to assess informant accuracy in relation to children's current level of cognitive functioning, not historical reports of children's past performance, or parental or teacher expectations of children's prospective abilities.

3.4.1.3 Publication source

In meta-analyses, the issue of publication bias is often addressed by the inclusion of unpublished studies (e.g., dissertations, conferences papers, etc.). In the present study, the focus was on articles published in scientific journals. This is because the findings of low correlations between parent and/or teacher-reports and children's performance do not usually prevent publication as they can still provide meaningful results (i.e., parents and/or teachers are not accurate reporters of children's cognitive functioning). The wide range of correlations presented in the Results sections of the reviewed papers provides evidence to suggest publication bias does not appear to be present in the current dataset. Methods for empirically identifying and assessing the impact of publication bias in the current review are discussed subsequently.

3.4.2 Classification of CHC ability areas

As the focus of the meta-analyses was on the validity of parent-report and teacher-reports of *g* and the cognitive abilities of *Gf, Gc, Gwm, Gl, Gr, Gv, Ga* and *Gs*, and academic abilities of *Grw-R, Grw-W* and *Gq*, as delineated in CHC theory, it was necessary to determine if the informant-reports obtained in the included articles could be considered as providing a measure of the respective ability areas. Since the existing studies throughout the research domains under review have used a variety of theoretical frameworks of cognitive ability as a reference point, with some studies not explicating any theory at all, it was not always immediately clear that a reported ability was theoretically analogous to abilities contained within the CHC framework. However, some cognitive ability constructs contained within the psychometrically-derived CHC model can exist in non-psychometric frameworks. For example, Gardner's (1983) Logical-mathematical, Verbal/Linguistic, and Spatial intelligences were considered by Carroll (1993) to be parallel to that of *Gf, Gc,* and *Gs* respectively (Jacobs, 2012). Since determining whether the content of various items or scales is theoretically parallel can involve subjective judgments, having multiple raters evaluate the comparability of the variables is recommended to ensure the reliability of the classification system (Hedges & Becker, 1986; Jacobs, 2012). Hence, this stage of the data coding procedure was conducted independently by the author and one other PhD student who was familiar with CHC theory and experienced in cognitive ability assessment.

Parent-report and teacher-report items and scales were classified as providing a measure of either *g*, *Gf, Gc, Gwm, Gl, Gr, Ga, Gv, Gs*, *Grw-R, Grw-W,* and *Gq*. The aim of this classification process was to retain studies that provided as pure a measure of any one of the 12 ability areas as possible. Therefore, reports of abilities that were determined to be a measure of a mixture of ability areas, or were viewed as measuring both cognitive and non-cognitive factors, were to be coded as *other*. Further, it was important to recognise the difference between *Gf* and *Gq*. One of the narrow abilities subsumed under *Gf* is quantitative reasoning, a skill that necessitates being able to display inductive and/or deductive reasoning when solving quantitative problems (e.g., knowing what comes next in a number series task; McGrew, 2005). In contrast, *Gq* is primarily evident when a task requires mathematical skills (e.g., numerical operations of addition and subtraction) and/or general mathematical knowledge (e.g., knowing how to calculate the circumference of a circle). Hence, informant-reports that did not have a clear indication of whether the ability to be rated was either reasoning with mathematical concepts or ability to solve math tasks was classified as *other*.

3.5 Data extraction

The third step of a meta-analysis involves establishing an explicit data coding system to extract pertinent information about the participant and methodological characteristics of the included studies (Russo, 2007). The following information was coded for each of the meta-analyses: (a) gender and age of child participants; (b) country in which data was collected; (c) theoretical framework of cognitive ability used; (d) whether parallel measures were used for report and criterion tests; (e) type of response format used to obtain informant-reports; (f) whether item/s on informant-reports refer to a cognitive construct or cognitive task. All coding categories are summarised in Table 3.1. The informant-level variable that was coded was the parent gender for the parent-report meta-analysis, and the years of teaching experience (in terms of the mean and standard deviation) for the teacher-report meta-analysis. Teacher gender was not coded as this information was not often reported in the extant literature. Additionally, when information about the sample size of child, parent, and/or teacher was available, this was coded as a continuous variable. When insufficient information was provided in an article to be able to code for that particular aspect of the study, *not reported* was recorded.

In determining whether the informant-report measure was parallel to the criterion test (i.e., the informants were asked to estimate the same ability the child was tested on), relevant literature and research (e.g,. Flanagan, Ortiz, et al., 2013; Schneider & McGrew, 2012) was consulted to determine whether criterion tests used in studies could be regarded as providing a measure of one of the CHC ability areas. However, in some instances such as when the relevant literature was not available in English, or when the study used an outdated test, it was not possible to locate past research that could help to decide which of the CHC abilities a particular assessment measured. When this issue arose, face validity indicators such as test names and descriptions of the tasks involved were used to inform a decision. Each study was coded separately by the author and the same PhD student referred to previously.

Table 3.1

Categorical Coding Categories for Meta-Analyses of Validity of Parent-Reports and Teacher-Reports of Children's Cognitive and Academic Abilities

Country of origin
1 = North America
2 = UK
3 = Europe (excluding UK)
4 = Australia
5 = Asia
Child gender
1 = Male
2 = Female
3 = Mixed
Parent gender
1 = Mother
2 = Father
3 = Mixed
Criterion test and informant-report match
1 = Parallel
2 = Non-parallel
Number of items
1 = Single
2 = Multi
Type of comparison
1 = Intra-individual
2 = Inter-individual
Item reference
1 = Cognitive construct
2 = Cognitive task

3.6 Effect sizes

The majority of studies used the correlation coefficient to report the effect size for the relationship between informant-report and children's performances on tests of the same ability. When results were reported only as t-tests or beta coefficients, these were converted to *r* values using the relevant formulas (Peterson & Brown, 2005). If there was inadequate information provided in the article to compute an effect size, the article was discarded from the analysis.

Multiple effect sizes were regularly noted in the one study as a result of most studies obtaining parent-/teacher-report of more than one of the 10 cognitive ability areas under review, as well as using more than one outcome measure (e.g., informant-report of *g* correlated with performance-based measures of *g* and reading ability). The resulting effect sizes are considered to be dependent because the same participant was measured twice. As outlined by Jacobs (2012), this was an issue as most meta-analytic methods assume independence of effect sizes (Hedges & Olkin, 1980). The inclusion of statistically dependent effect sizes can compromise the validity of the meta-analytic results by artificially reducing estimates of variance, subsequently inflating Type 1 error (Borenstein, Hedges, Higgins, & Rothstein, 2009; Scammacca, Roberts, & Stuebing, 2014).

Hence, the following methods were employed during the coding and analysis stages to ensure the independence of effect sizes. First, only results for present informant-reports (as opposed to prediction of future academic or other life outcomes) were recorded, as these were considered to be most relevant to the purposes of a parent- and teacher-report measure designed to assist practitioners with assessment decisions. Additionally, if informant-reports were obtained both before and after the performance test was administered, only the pre-test report was retained. This is because using pre-test reports would ensure that the informant-report would be free from the potential bias of knowing how the child performed on the criterion measure. Additionally, an informant-report measure of abilities would be more useful to clinicians as a supplement, by using information from the parent or teacher to guide the focus of the assessment process with performance-based measures. The second method was to, when possible, run separate analyses for each cognitive ability area being investigated. When neither of these two methods were able to be implemented, then multiple correlations from the one study were averaged.

3.7 Statistical Analysis Plan

3.7.1 Publication bias

Publication bias, also known as the 'file drawer problem'(Rosenthal, 1979), arises from an increased likelihood of publication in journal articles for studies with statistically significant or clinically favourable results, compared to studies with nonsignificant or unfavourable results (Ahmed, Sutton, & Riley, 2012). This leads to the synthesis of an incomplete set of the available evidence within the research domain, and the production of summary results that are potentially biased towards favourable treatment effects. The results of meta-analyses are commonly used to make clinical recommendations for practical interventions, so problematic consequences may occur if an inappropriate conclusion is drawn from over-estimated effect sizes. Ways to reduce publication bias include undertaking searches of the grey literature which refer to material that is not formally published (Martin, Pérez, Sacristán, & Álvarez, 2005), as well as statistical methods for detecting and correcting for it (Vevea & Woods, 2005).

Funnel plots of the effect sizes and their standard errors of the included studies were generated and visually inspected to see if there was a skew indicating publication bias. Funnel plots are scatterplots of the effect sizes obtained from individual studies against a measure of study size, and are primarily used as a visual aid for detecting publication bias in meta-analysis (Sterne & Harbord, 2004). In the absence of bias, the plot will resemble a symmetrical, inverted funnel. If publication bias is present, the funnel plot will be skewed and asymmetrical, with observable gaps at the bottom area of the graph attributable to smaller studies that are missing from the meta-analysis (Borenstein et al., 2009). In this scenario, Orwin's fail-safe N (Orwin, 1983) will be calculated to determine the number of studies it would take to bring the effect size down to the smallest effect deemed to be of clinical significance. A relatively small N would indicate a reason for publication bias (Borenstein et al., 2009). Hence, both funnel plots and fail-safe N was used as indicators for the presence and extent of publication bias in the current meta-analyses.

3.7.2 Choice of model for within-groups analysis

In conducting a meta-analysis, the researcher has the option to choose one of two statistical models to calculate the summary effect size; the fixed-effects model or the random-effects model. According to Borenstein et al. (2009), the difference between the two lies in the assumption of effect sizes. In the fixed-effects model, the assumption is that there is one true effect size underlying all included studies and all differences in the observed effects are due to sampling error, so that the summary effect size is the estimate of the common effect size that is shared in all included studies of the meta-analysis (Borenstein et al., 2009). In contrast, the random-effects model accounts for the potential of varying population parameter values from one study to the next (Hunter & Schmidt, 2000), such that the summary effect size is the "estimate of the mean of the distribution of effect sizes" (Borenstein et al., 2009, p. 6). Choosing which statistical model depends on the sampling frame used to select the included studies, and the models lead to different significance tests and confidence intervals for mean effect sizes (Hunter & Schmidt, 2000). The choice of statistical model is important because fundamentally, the goals of the analysis vary with the choice of the model, and serious errors can occur when the assumptions underlying these models are violated.

Following recommendations made by Borenstein et al. (2009) and Lipsey and Wilson (2001), the reasons for using a random-effects model for the current study were two-fold. Firstly, since the studies included in the meta-analytic reviews of this dissertation did vary on participant and method characteristics, and because moderators of the observed effect sizes were hypothesised *a priori*, the use of a random-effects model was considered to be the most appropriate in this case to adjust for unexplained heterogeneity. When there are variations in study characteristics of the included studies, incorrectly assuming that there is one true effect size across all studies (i.e., adopting a fixed-effects model) could increase the risk of Type I error occurring and confidence intervals that are too narrow, thereby resulting in a substantial over-statement of the precision of the meta-analysis results (Hunter & Schmidt, 2000). Additionally, the inferences made from a random-effects model can be applied to a population of studies larger than the sample, while inferences made from a fixed-effect model can only be extended to those studies included in the analysis. Since the goal to extend findings to subsequent research is typical of social science researchers, a random-effects model was deemed to be the most appropriate in calculating the overall effect size for the current meta-analyses (Cohn & Becker, 2003; Hunter & Schmidt, 2000).

3.7.3 Assessment of heterogeneity

An underlying goal of conducting a meta-analysis is to assess for heterogeneity among the included studies, and to identify the possible variables (i.e., moderators) that contribute to it (Higgins, 2008). Heterogeneity refers to the variation in study outcomes and comes from either variability due to sampling error (within-study) or true variability among the population effect sizes estimated by the individual studies (between-study). Between-study variability is due to the effect of an unspecified number of characteristics that vary amongst studies such as those related to sample-specific characteristics, variations in design quality, and so on (Brockwell & Gordon, 2001; Erez, Bloom, & Wells, 1996; Hunter & Schmidt, 2000). Assessing for heterogeneity in a meta-analysis is crucial because the presence or absence of true heterogeneity can impact which statistical model the researcher decides to choose. Should the studies' results only vary due to sampling error (a homogenous case), a fixed-effects model can be utilised to obtain a summary estimate of effect size. In contrast, should the studies' results differ by more than just the sampling error (a heterogeneous case), then the random-effects model should be chosen to take into consideration both within- and between-study variability, or to investigate moderators using a fixed-effects model (Field, 2001; Hedges & Olkin, 1985; Hedges & Vevea, 1998). Additionally, understanding the possible causes of variability can increase both the scientific value and clinical relevance of results from a meta-analysis (Thompson & Sharp, 1999).

The assessment of heterogeneity in the observed effect sizes for the current meta-analyses were carried out by calculating the Q-statistic (Cochran, 1954) and I^2 index. The presence of heterogeneity is indicated by a significant Q-statistic (p <.05). Although the Q-statistic is a commonly used assessment of heterogeneity, it has been argued that its ability to detect true heterogeneity is reduced when the meta-analysis contains a small number of studies, and an increased risk of detecting negligible variability with a high number of studies (Huedo-Medina, Sanchez-Meca, Marin-Martinez, & Botella, 2006). Thus, a non-significant result for the Q-test with a small number of studies can lead the researcher to mistakenly assume a fixed-effects model when there is true heterogeneity, and vice versa. Hence, the I^2 value, ranging from 0% to 100%, was inspected concurrently to quantify heterogeneity, and is interpreted as the percentage of the total variability in effect sizes due to between-study variability (i.e., true heterogeneity). The following classifications proposed by Higgins and Thompson (2002) were used, with higher values indicating greater heterogeneity: High (I^2 = 75%), Medium (I^2 = 50%) and Low (I^2 = 25%).

3.7.4 Choice of model for between-groups analysis

Since it was expected that significant heterogeneity would be found in the observed effect sizes, thereby necessitating the investigation of moderators, a plan for performing subgroup analyses was required. According to Borenstein et al. (2009), the researcher needs to decide if the common estimate is applied across all studies (i.e., the within-group variance is pooled), or each subgroup's variance is applied only to the studies within that subgroup (i.e., the within-group variance is kept separate). Since the participant sample type (i.e., typically developing children) and methods used across studies were largely similar across studies in the majority of cases, it was expected that the true study-to-study dispersion would be the same across subgroups. Therefore, the within-group estimates of variance were pooled and applied to all studies.

Although a random-effects statistical model was used to compute the summary effect size within groups, because the meaning of both the fixed-effect and random-effects models changes when used in the between-groups context, a decision needs to be made on which statistical model (fixed-effects or random-effects) should be used in subgroup analysis (Jacobs, 2012). A fixed-effect model for between-groups analysis should be applied when the subgroups are unchangeable in the sense that the same subgroups have to be used if someone else is performing the analysis at another time. For example, in comparing a subgroup of studies with only female participants, and a subgroup of studies with only male participants, the appropriate option is with the fixed-effect model because there is no other way in which the gender variable can be grouped (Borenstein et al., 2009). As for choosing the random-effects model in between-groups analysis, this is appropriate when the subgroups used are a random sample of all potential subgroups, and thus could change from one meta-analysis to the next (Jacobs, 2012). An example of this is grouping studies by their country of origin – one meta-analysis could sample studies from U. S., Australia, and Germany, while another could have studies from Australia, U. S., and U.K. (Borenstein et al., 2009; Jacobs, 2012). For the current meta-analyses, the fixed-effect model was used because the moderator variables tested are unchangeable. For example, whether a study's informant-report measure had single or multiple items could only be grouped in one condition (i.e., either single or multiple). Hence, a mixed-effects model was used in the current meta-analyses, since a random-effects model was used for within-group variance, and a fixed-effects model was used for between-group variance (Borenstein et al., 2009).

3.7.5 Moderator analysis

The nature of synthesising a large amount of research inevitably involves the analysis of numerous studies that have vary widely in their methodologies (Öst, 2008). This introduces into the analysis a range of variables that can affect the overall effect size, and these factors are commonly referred to as moderators because of their potential influence on the strength and direction of the effect sizes. The current study hypothesised several human and methodological factors as moderators. These were: gender of child, age of child, gender of parent, professional experience of teacher, if the criterion test used as the comparison matched the informant-report (parallel or non-parallel), number of items on the informant measure (single or multiple), type of comparison used in the informant measure (intra-individual or inter-individual), and item reference (cognitive construct or cognitive test task).

The number of moderator analyses tested in a meta-analysis needs to be as limited as possible, because continuous probing can increase the probability of Type I error rates, and can also reduce the validity and generalisability of the conclusions drawn (Matt & Cook, 2009). It was thus decided that the moderator variables would only be tested for all cognitive ability areas combined to help counteract the problem of a lack of power to detect moderator effects (Jacobs, 2012). If moderators had been investigated for each cognitive ability area separately, this would have resulted in some subgroups containing a very small, and sometimes even non-existent, number of studies. The effect of the investigated moderator variables is described in the results section and explored later in the discussion section.

Chapter 4

Meta-Analyses: Results & Discussion

This chapter presents the results of meta-analytic reviews of the current literature pertaining to children's cognitive and academic abilities as reported by parents and teachers with the intention of identifying the human characteristics and, more importantly, the measurement conditions that are conducive to valid estimations of children's cognitive functioning from these informants. It begins with a detailed account of the outcome from the literature search and is followed by details of interrater reliability of CHC ability classification and data coding. Subsequently, the findings of the meta-analyses are presented beginning with the calculation of an overall effect size, followed by findings of the moderator analysis to investigate the potential effect of the variables under investigation on the accuracy of parent-reports and teacher-reports of children's cognitive and academic abilities.

4.1 Literature Search Outcome

Figures 4.1 (parent-report) and 4.2 (teacher-report) depict the process of the literature search and study selection for inclusion in the meta-analyses. The first phase of the search was to enter all combinations of search terms (as listed in table 4.1 and table 4.2) in each of the three databases, resulting in a total of 13,030 potentially relevant references for parent-reports, and 11,011 for teacher-reports. The initial screening of study titles resulted in the identification of a total of 572 parent-report and 464 teacher-report references as potentially including relevant information. After the elimination of article duplicates across the three databases, and reading article abstracts as supplied from the databases, 97 parent-report and 151 teacher-report full-text articles were obtained for further review.

Table 4.1

Parent-Report Search Terms used in Systematic Literature Search of Online Databases

No.	Search terms	Results		
		Scopus	Web of Science	PsycINFO
1.	parent* belie* and academic abilit*	58	86	5
2.	parent* belie* and academic perf*	99	32	5
3.	parent* belie* and cogniti* abilit*	77	136	10
4.	parent* belie* and intell*	181	159	42
5.	parent* belie* and IQ	42	64	15
6.	parent* estimat* and academic abilit*	27	37	0
7.	parent* estimat* and academic perf*	106	160	0
8.	parent* estimat* and cogniti* abilit*	93	151	3
9.	parent* estimat* and intell*	208	22	31
10.	parent* estimat* and IQ	114	18	14
11.	parent* evaluat* and academic abilit*	91	135	0
12.	parent* evaluat* and academic perf*	221	167	5
13.	parent* evaluat* and cogniti* abilit*	144	124	1
14.	parent* evaluat* and intell*	523	314	35
15.	parent* evaluat* and IQ	158	119	4
16.	parent* judg* and academic abilit*	11	22	0
17.	parent* judg* and academic perf*	26	37	2
18.	parent* judg* and cogniti* abilit*	45	36	4
19.	parent* judg * and intell*	59	43	6
20.	parent* judg* and IQ	42	33	0
21.	parent* perce* and academic abilit*	122	202	5
22.	parent* perce* and academic perf*	342	52	48
23.	parent* perce* and cogniti* abilit*	224	174	26
24.	parent* perce* and intell*	443	87	165
25.	parent* perce* and IQ	129	274	21
26.	parent* rat* and academic abilit*	133	27	3
27.	parent* rat* and academic perf*	386	79	54
28.	parent* rat* and cogniti* abilit*	317	95	214
29.	parent* rat* and intell*	171	210	829
30.	parent* rat* and IQ	72	117	194
31.	parent* report* and academic abilit*	145	218	10
32.	parent* report* and academic perf*	101	173	103
33.	parent* report* and cogniti* abilit*	364	280	230
34.	parent* report* and intell*	276	192	825
35.	parent* report* and IQ	108	198	190
	Total	5658	4273	3099

Table 4.2

Teacher-Report Search Terms used in Systematic Literature Search of Online Databases

No.	Search terms	Results (Retained)		
		Scopus	Web of Science	PsycINFO
1.	teacher* belie* and academic abilit*	144	158	1
2.	teacher* belie* and academic perf*	204	10	15
3.	teacher* belie* and cogniti* abilit*	80	120	4
4.	teacher* belie* and intell*	30	29	71
5.	teacher* belie* and IQ	13	18	5
6.	teacher* estimat* and academic abilit*	54	48	2
7.	teacher* estimat* and academic perf*	124	1	2
8.	teacher* estimat* and cogniti* abilit*	32	47	9
9.	teacher* estimat* and intell*	99	101	35
10.	teacher* estimat* and IQ	39	46	20
11.	teacher* evaluat* and academic abilit*	216	188	8
12.	teacher* evaluat* and academic perf*	24	13	32
13.	teacher* evaluat* and cogniti* abilit*	165	136	20
14.	teacher* evaluat* and intell*	29	353	68
15.	teacher* evaluat* and IQ	87	97	21
16.	teacher* judg* and academic abilit*	57	62	11
17.	teacher* judg* and academic perf*	22	6	15
18.	teacher* judg* and cogniti* abilit*	43	44	18
19.	teacher* judg * and intell*	184	133	72
20.	teacher* judg* and IQ	23	18	26
21.	teacher* perce* and academic abilit*	70	320	40
22.	teacher* perce* and academic perf*	87	22	71
23.	teacher* perce* and cogniti* abilit*	202	216	20
24.	teacher* perce* and intell*	71	69	268
25.	teacher* perce* and IQ	10	12	31
26.	teacher* rat* and academic abilit*	141	123	44
27.	teacher* rat* and academic perf*	192	169	222
28.	teacher* rat* and cogniti* abilit*	151	56	265
29.	teacher* rat* and intell*	299	161	849
30.	teacher* rat* and IQ	215	124	190
31.	teacher* report* and academic abilit*	236	6	8
32.	teacher* report* and academic perf*	120	72	136
33.	teacher* report* and cogniti* abilit*	189	29	144
34.	teacher* report* and intell*	535	123	584
35.	teacher* report* and IQ	136	83	148
	Total	4323	3213	3478

4.1.1 Parent-report

In the next step, the selected studies were read and the inclusion/exclusion criteria was applied. Among the 97 studies ascertained to be potentially relevant, 32 studies were excluded due to the absence of comparison between parent reports and children's actual performance. Seven studies did not use a standardised cognitive or academic ability test as the criterion measure, the child sample of 16 studies was either age-inappropriate or clinical, and parents were not explicitly asked to rate their child's ability in four studies. These studies, including others that were irrelevant (e.g., parents providing a developmental or mental age estimate of the child's functioning), not available in English, or did not contain enough information to be coded, were also excluded.

One study (Hedges, Drysdale, & Levick, 2015) explicitly stated that non-significant results were not reported, therefore five non-significant effect sizes (3.5%) were identified as missing. Missing effect sizes are regarded as "the most pervasive problem in meta-analyses" (Hedges, 1992, p. 292), as studies with missing effect sizes are unable to contribute to the estimation of the summary effects (Pigott, 2009). Missing effect sizes that are a direct result of selective reporting within individual studies due to nonsignificant results cannot be considered as missing at random (Norris et al., 2012; Page et al., 2014), and are likely to have a detrimental effect on the generalisability of results (Tabachnick & Fidell, 2007). Additionally, they can cause an inflation of the summary effect size calculated from the meta-analysis, thereby impacting on the validity of drawn conclusions (Matt & Cook, 2009). Various methods in dealing with missing data should produce similar results in cases where the number of missing effect sizes relative to the total number of effect sizes is small (i.e., < 5%; Tabachnick & Fidell, 2007), which was the case here. It was thus decided that this study (R. Hedges et al., 2015) would be excluded from the parent-report meta-analysis. Although this resulted in also deleting reported effect sizes that were statistically significant, its biasing effects are likely to be smaller than the effects of recording the missing effect sizes as zero (Kuncel, Hezlett, & Ones, 2001).

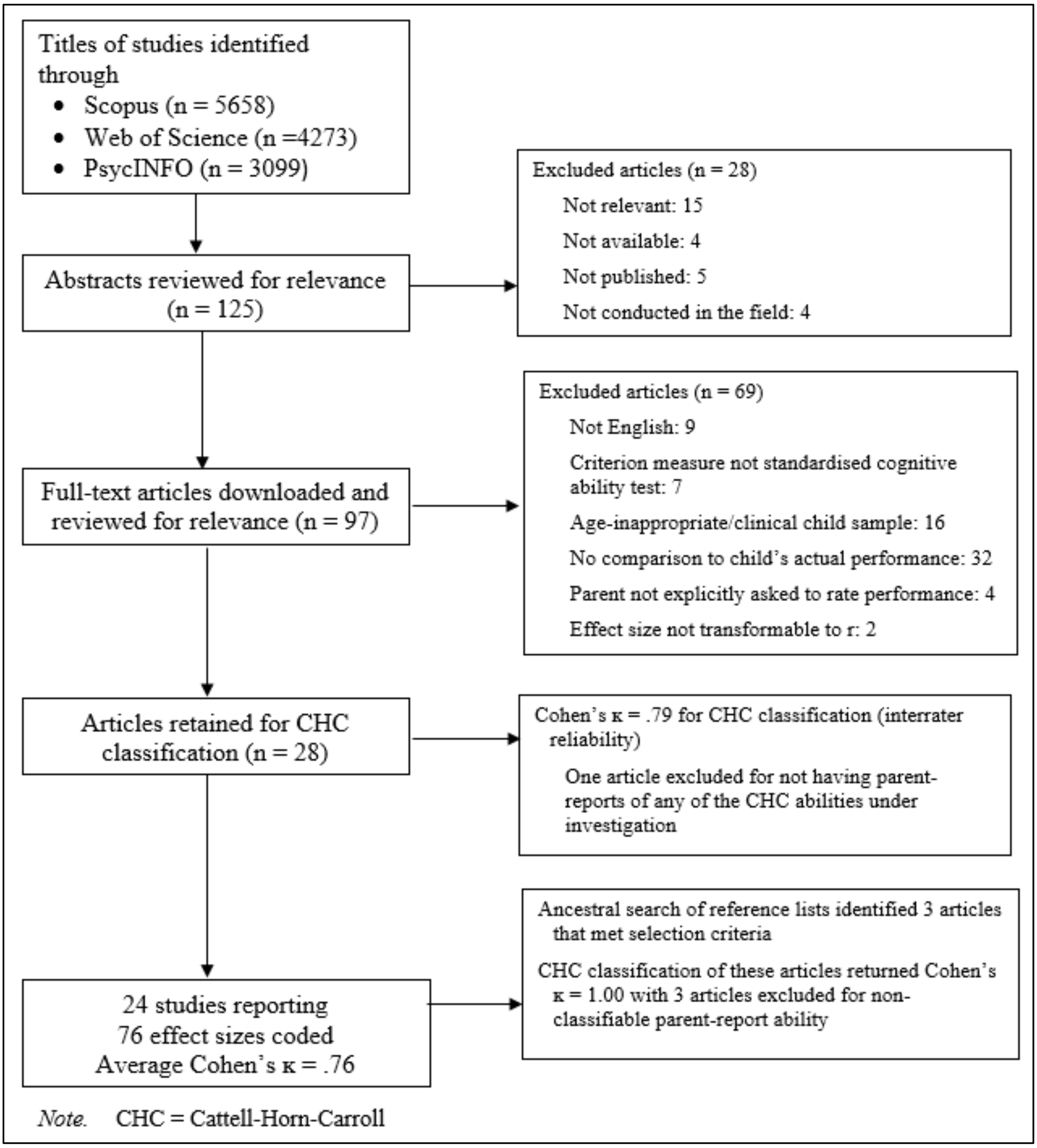


Figure 4.1 Flowchart of literature and study selection for inclusion in parent-report meta-analysis.

4.1.2 Teacher-report

The same process was followed for the selection of studies. Among the 151 studies determined to be potentially relevant, 28 studies were excluded as there was no comparison of teacher reports to children's actual performance. Seventeen studies did not use a standardised cognitive or academic ability test as the criterion measure, the student sample used by 17 studies was either age-inappropriate or clinical, and 10 studies were discarded as the effect size reported was either not a correlation coefficient, or was not transformable to a correlation coefficient (e.g., percentage hits). These studies, including others that were irrelevant, not available in English, or did not contain enough information to be coded, were also excluded. Two studies (Krkovic, Greiff, Kupiainen, Vainikainen, & Hautamäki, 2014; Südkamp, Praetorius, & Spinath, 2017) reported beta coefficients, which were successfully converted to r values using the Peterson and Brown (2005) formula. There were no studies that reported systematic exclusion of non-significant results, so the issue of missing effect sizes was not relevant in this meta-analysis.

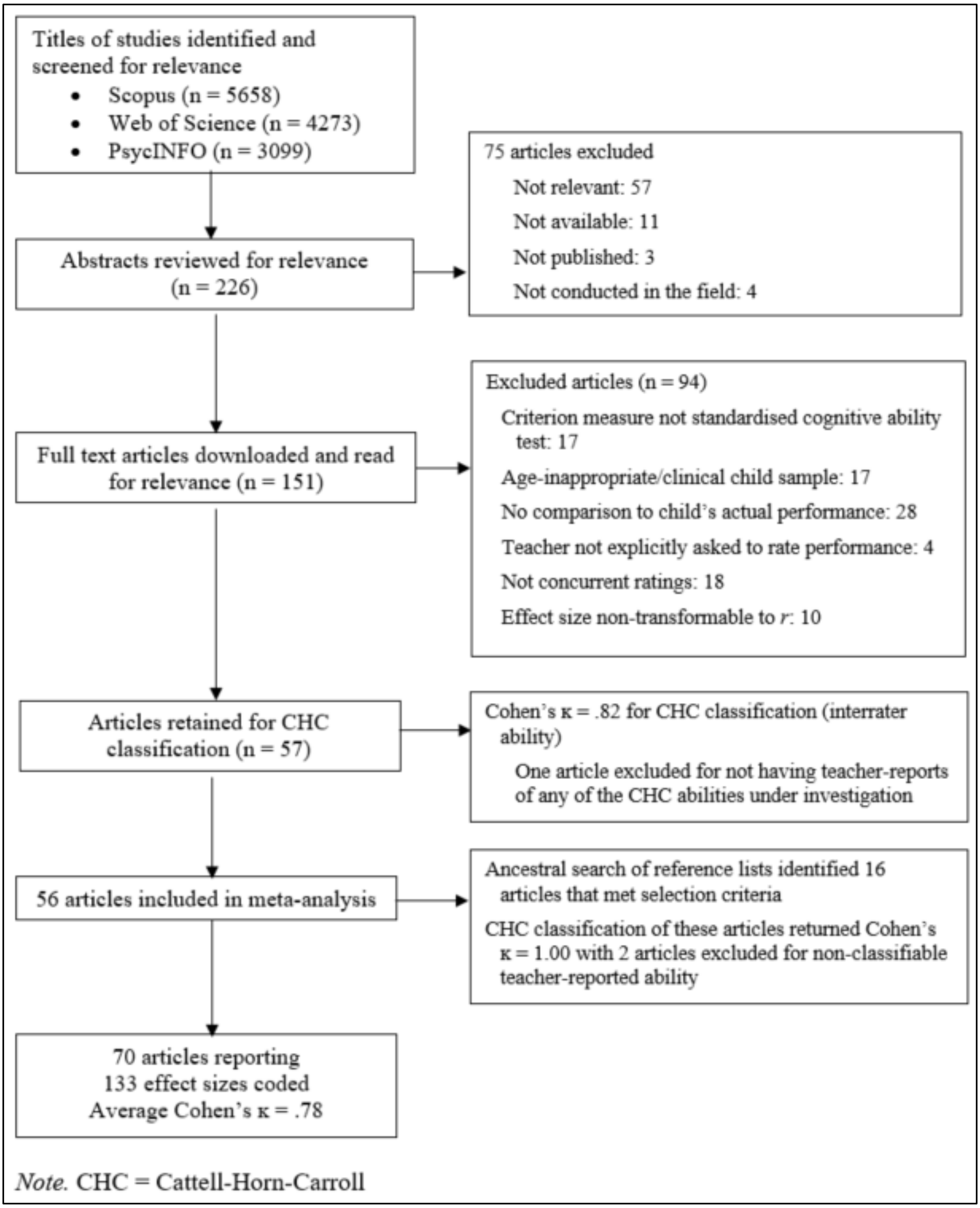


Figure 4.2 Flowchart of literature search and study selection for inclusion in the teacher-report meta-analysis

4.2 Interrater Reliability

Two raters were used as part of the process for establishing the eligibility criteria for studies and extraction of data via coding. Interrater agreement was calculated using Cohen's (1960) κ, which measures interrater agreement for qualitative items and takes into account the possibility of the agreement occurring by chance. All discrepancies were discussed until agreement was reached.

For classification of the CHC ability areas, the obtained values of κ =. 79 (parent-report) and .82 (teacher-report) were regarded to have obtained substantial agreement (Landis & Koch, 1977). After the retrieval of articles through ancestral searches of retained articles, a total of four parent-report articles and three teacher-report articles were excluded at this stage of the study selection procedure for not obtaining a report of either one of the twelve ability areas under investigation. Data from the retained parent-report and teacher-report articles were subsequently coded, with the average interrater reliability deemed to be substantial at κ = .76 and .78, respectively (Landis & Koch, 1977).

Therefore, following the implementation of these procedures, the total number of effect sizes that were included for each meta-analytic review were as follows: 76 effect sizes reported from 24 parent-report studies, and 133 effect sizes reported from 70 teacher-report studies.

4.3 Findings from the Meta-Analyses

Effect sizes were analysed using the *Comprehensive Meta-Analysis* software package (Borenstein et al., 2009). The 24 parent-report and 70 teacher-report studies included in the analysis of effect sizes are documented in Tables 4.3 and 4.4, respectively. The correlation between informant-reports and children's performance (*r*) and the size of the children sample (*N*) is reported for each study. For the parent-report meta-analysis, the 76 effect sizes ranged from *r* = .01 to .71 (*M* = .34, *SD* = .15). Study sample sizes ranged from 40 to 1,106, and the accumulated sample sizes for all studies included was 8,336. A majority (80%) of the included studies did not explicitly state a theoretical framework on which the developed parent-report measure was based. For the teacher-report meta-analysis, the 133 effect sizes ranged from *r* = -.37 to .86 (*M* = .54, *SD* = .19). Student sample sizes ranged from 14 to 11,675, and the accumulated sample sizes for all studies included in the review was 55,853. A majority (89.74%) of studies did not explicitly state a theoretical framework on which the teacher-report measure used was based.

4.3.1 Publication bias

If publication bias is present in a meta-analysis, results may over-estimate the effect size due to being based on a biased sample of the target population of studies (Borenstein et al., 2009). The presence of publication bias in the meta-analytic review was assessed by inspecting the funnel plots of both meta-analytic reviews (Figure 1 and Figure 2), which revealed moderately symmetrical relationships between the magnitude of effect sizes and their standard errors. Orwin's fail-safe *N* was also calculated, with an estimated number of 71 parent-report studies and 464 teacher-report studies needed to bring the effect sizes to substantive importance. These findings indicate that publication bias is not present in the studies selected for inclusion in the meta-analyses.

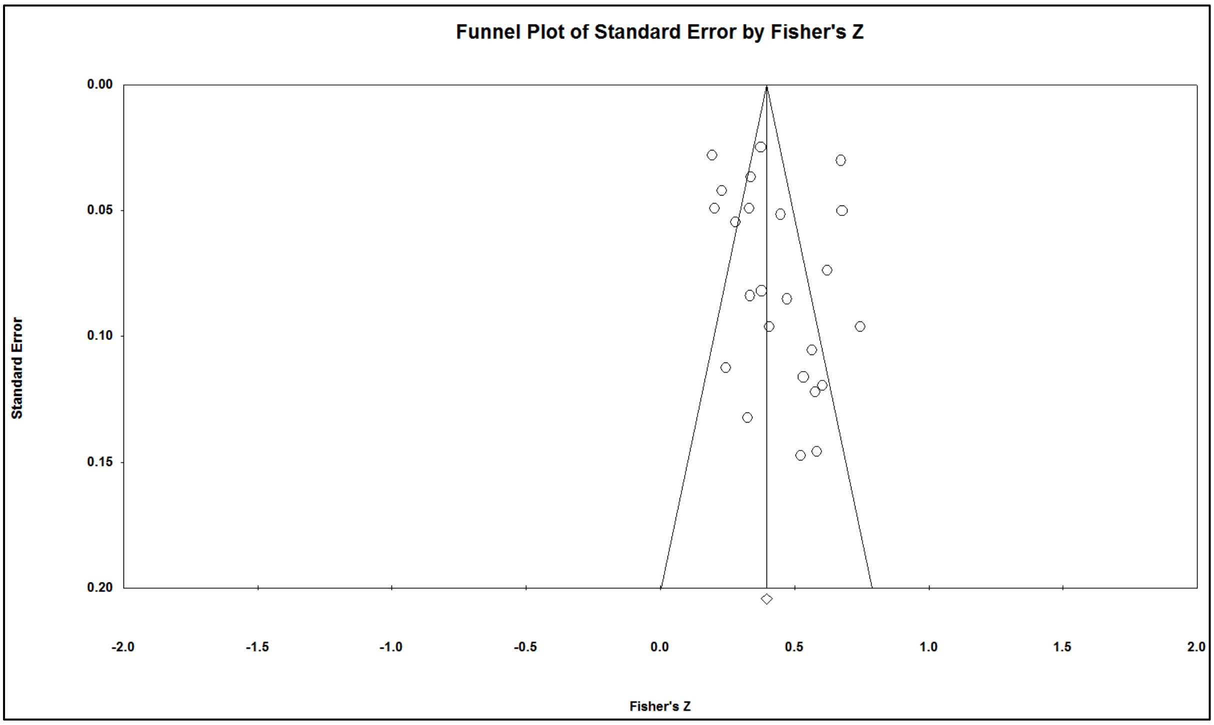


Figure 4.3 Parent meta-analysis: Funnel plot of standard error by Fisher's z

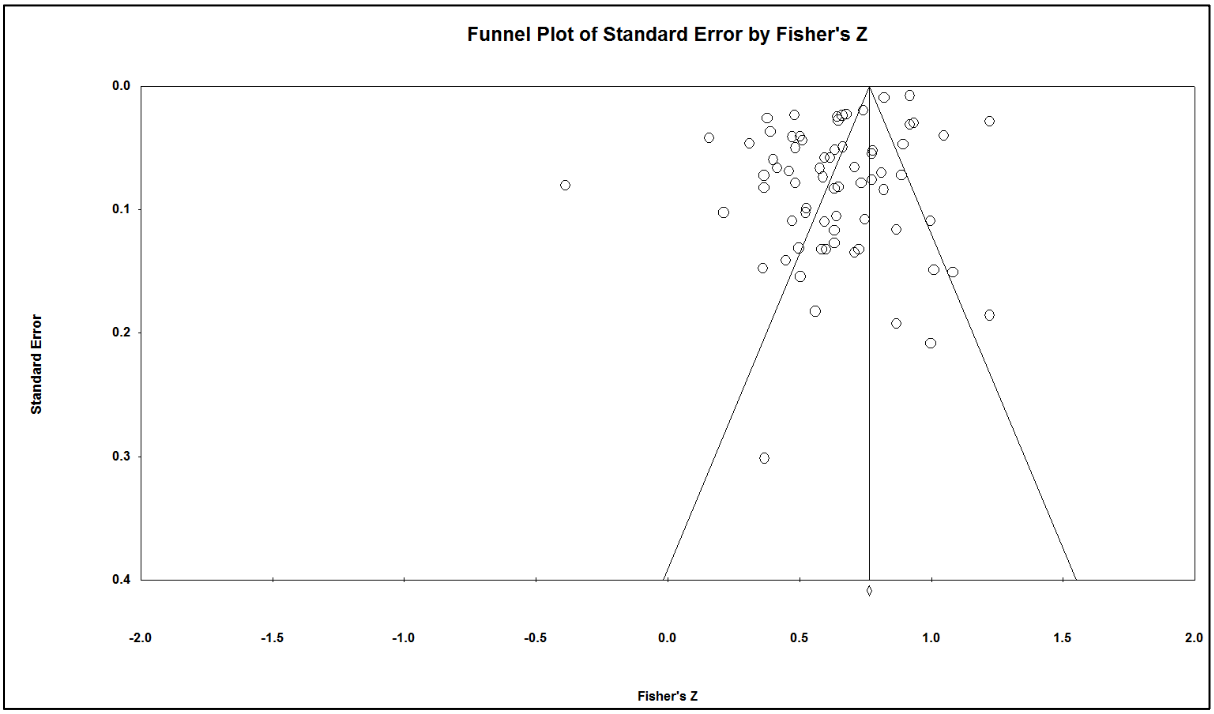


Figure 4.4 Teacher meta-analysis: Funnel plot of standard error by Fisher's z

Table 4.3

Overview of Studies Included in Meta-Analysis of the Validity of Parent-Reports of Children's Cognitive and Academic Abilities

Study author/s	Year pub.	Country	Theoretical Framework	Child N	Child gender	Parent gender	Child age M (SD)	Number of items	Type of comparison	Item reference	Parent-reported ability	Performance-based ability	Match of criterion test	Effect size (r)
Aunola, Nurmi, Lerkkanen et al.	2003	EU	None stated	52	Female	Mother	7.3 (0.32)	Multi	Intra	Con	Gq	Gq	Parallel	.47
						Father								.28
				59	Male	Mother								.43
						Father								.35
Aunola, Nurmi, Niemi et al.	2003	EU	None stated	52	Female	Mother	7.3 (0.32)	Multi	Intra	Con	Grw-R	Grw-R	Parallel	.70
						Father								.57
				59	Male	Mother								.61
						Father								.64
Baudson	2013	EU	CHC theory	533	Mixed	Mixed	7.25 (0.42)	Multi	Intra	Task	g	g	Parallel	.27
				531			8.33 (0.45)							.39
				565			9.42 (0.45)							.41
Bouffard	2005	USA	None stated	77	Mixed	Mother	6.22 (0.41)	Multi	Intra	Task	Gq	Gq	Parallel	.44
											Grw-R	Grw-R	Parallel	.53
Brummelman	2014	Europe	None stated	82	Mixed	Mixed	12.53 (0.45)	Multi	Intra	Con	g	g	Parallel	.24

Table 4.3 continued

Chamorro-Premuzic	2006	U.K.	None stated	187	Mixed	Mother	14.33 (0.32)	Sing	Inter	Con	g	g	Parallel	.65
											Gq	Gq	Parallel	.63
											Gv	Gv	Parallel	.27
											Gc	Gc	Parallel	.70
						Father					g	g	Parallel	.58
											Gq	Gq	Parallel	.54
											Gv	Gv	Parallel	.39
											Gc	Gc	Parallel	.56
Delgadohachey	1993	USA	None stated	70	Mixed	Mother	9.5	Sing	Inter	Con	g	g	Parallel	.52
Furnham	2006	U.K.	MI Theory	141	Mixed	Mixed	11.47 (2.21)	Sing	Inter	Con	g	g	Parallel	.44
Goforth	2014	USA	None stated	747	Mixed	Mixed	NR	Multi	Inter	Task	Grw-R	Grw-R	Parallel	.34
											Gq	Gq	Parallel	.31
Gut	2013	Europe	None stated	402	Mixed	Mixed	6.22 (0.65)	Sing	Inter	Con	Mixed	Gf	Non-parallel	.59
Herbert	2005	USA	None stated	378	Mixed	Mixed	6.46 (0.58)	Sing	Inter	Con	Grw-R	Grw-R	Parallel	.44
											Gq	Gq	Parallel	.40
Hernandez-Torrano	2014	Europe	MI Theory	566	Mixed	Mixed	14.85 (1.08)	Multi	Intra	Task	Gc	Gc	Parallel	.29
											Gq	Gq	Parallel	.22
											Gf	Gf	Parallel	.15
											Gv	Gv	Parallel	.24
Kirkcaldy	2007	EU	Sternberg	415	Mixed	Mother Father	10.6 (0.58)	Sing	Inter	Con	Gf	Gf	Parallel	.19 .21
Massa	2008	USA	None stated	73	Mixed	Mixed	NR	Multi	Intra	Task	Grw-R	Grw-R	Parallel	.54*

Table 4.3 continued

Magui	1996	USA	None stated	764	Male	Mother	NR	Sing	Inter	Con	Grw-R Gq	Grw-R Gq	Parallel Parallel	.61 .56
Miller	1986	USA	Piaget	49	Mixed	Mother	6.92	Multi	Intra	Task	g Gc	g Gc	Parallel Parallel	.49** .47*
Miller & Davis	1992	USA	None stated	60	Mixed	Mother	9.85	Multi	Intra	Task	Gq Gc Gsm Gv	Gq Gc Gsm Gv	Parallel Parallel Parallel Parallel	.43 .19 .28 .35
Miller et al.	1991	USA	None stated	26 24	Mixed	Mother Father Mother Father	8.16 11.25	Multi	Intra	Task	g	g	Parallel	.71 .50 .44 .40
Phillipson	2012	Asia	Baddely & Hitch (1974) WM model	1279	Mixed	Mixed	P1-P6	Multi	Intra	Task	Gwm	g	Non-parallel	.19**
Pesu	2016	EU	None stated	152	Mixed	Mother Father	7.5	Sing	Intra	Con	Gq	Gq	Parallel	.41** .31*
Sommer	2008	Europe	None stated	47 46	Female Male	Mixed	9.10	Multi	Inter	Task	g	Mixed	Non-parallel Non-parallel	.54 .48
Spinath	2005	EU	None stated	416	Mixed	Mixed	8.7 (1.2)	Multi	Intra	Task	g	g	Parallel	.32

Table 4.3 continued

Steinmayr	2009	Europe	PMA theory	203	Female	Mixed	16.95 (7.20)	Sing	Intra	Task	Gc	Gc	Parallel	.35
											Gq	Gq	Parallel	.36
											Gv	Gv	Parallel	.34
											Gf	Gf	Parallel	.34
				136	Male						Gc	Gc	Parallel	.26
											Gq	Gq	Parallel	.34
											Gv	Gv	Parallel	.28
											Gf	Gf	Parallel	.26
Waschbusch	2000	USA	G*f*-G*c* Theory	89	Male	Mixed	8.04 (2.06)	Multi	Intra	Task	g	g	Parallel	.50
											Gf	Gf	Parallel	.45
											Gc	Gc	Parallel	.37
											Gv	Gv	Parallel	.32
											Ga	Ga	Parallel	.29
											Gl/Gr	Gl/Gr	Parallel	.33
											Gwm	Gwm	Parallel	.19
				56	Female						g	g	Parallel	.47
											Gf	Gf	Parallel	.39
											Gc	Gc	Parallel	.16
											Gv	Gv	Parallel	.51
											Ga	Ga	Parallel	.35
											Gl/Gr	Gl/Gr	Parallel	.08
											Gwm	Gwm	Parallel	.01

Note. N = sample size; Mixed = both male and female children, both mothers and fathers NR = not reported; con = cognitive construct; sing = single-item; multi = multi-item; intra = intra-individual comparison; inter = inter-individual comparison; MI theory = Gardners' Multiple Intelligence Theory; PMA = Thurstone's Primary Mental Abilities theory; g = general intelligence; Gc = crystallised knowledge; Gf = fluid reasoning; Gl = learning efficiency; Gr = retrieval fluency; Gwm = short-term working memory; Ga = auditory processing; Gv = visual processing

Table 4.4

Overview of Studies Included in Meta-Analysis of the Validity of Teacher-Reports of Children's Cognitive and Academic Abilities

Study author/s	Year pub.	Country	Theoretical Framework	Student N	Student gender	Student age M (SD)	Teacher years of experience	Number of items	Type of comparison	Item reference	Parent-reported ability	Performance -based ability	Match of criterion test	Effect size (r)
Alloway	2009	UK	WM model	65	Mixed	8.65	NR	Multi	Intra	Task	Gwm	Gwm	Parallel	.56***
Alvidrez	1999	USA	None stated	61	Mixed	11 (NR)	NR	Sing	Intra	Con	g	g	Parallel	.46
Barry	2015	AU	None stated	49	Mixed	9.62 (1.53)	NR	Multi	Inter	Task	Ga	Gf Gwm Grw-R	Non-parallel	.35* .28* .41*
Bates	2001	AU	None stated	77	Mixed	7.32 (0.80)	1.86 (9.58)	Multi	Inter	Con	Grw-R	Grw-R	Parallel	0.70***
Baudson & Preckel	2013	EU	CHC theory	533 531 565	Mixed	7.25 (0.42) 8.33 (0.45) 9.42 (0.45)	NR	Multi	Intra	Task	g	g	Parallel	.54*** .56*** .60***
Baudson, Fischbach & Preckel	2016	EU	None stated	1774	Mixed	NR	17.55 (11.92)	Multi	Inter	Task	g	g	Parallel	.58***
Begeny, Eckert, Montarello et al.	2008	USA	None stated	87	Mixed	8.25 (0.825)	18.5 (10.9)	Sing	Intra	Con	Grw-R	Grw-R	Parallel	.76**
Begeny, Krouse & Brown	2011	USA	None stated	212	Mixed	NR	15.6 (9.2)	Sing	Intra	Con	Grw-R	Grw-R	Parallel	.47**

Table 4.4 continued

Benner	2007	USA	Ecological theory	522	Mixed	12.5 (2.1)	NR	Sing	Inter	Con	g	g	Parallel	.47**
Beswick	2005	Canada	None stated	205	Mixed	6.08 (0.34)	20.33	Multi	Intra	Task	Grw-R	Grw-R	Parallel	.67**
Bohlmann	2013	USA	None stated	193	Mixed	NR	NR	Sing	Inter	Task	Gq	Gq	Parallel	.35**
Chamorro-Premuzic	2009	UK	Verbal, mathematical, spatial intelligence	187	Mixed	14.33 (.32)	NR	Sing	Inter	Con	g	g	Parallel	.57**
											Gq	Gq		.57**
											Gv	Gv		.41**
											Gc	Gc		.50**
Coladarci	1986	USA	None stated	48	Mixed	NR	12.81	Multi	Intra	Task	Grw-R	Grw-R	Parallel	.79*
											Gq	Gq		.74*
Commodari	2005	EU	None stated	98	Mixed	12.6	NR	Sing	Intra	Con	g	Grw-R	Non-parallel	.48**
Dean	1981	USA	None stated	60	Mixed	8.38 (2.10)	NR	Multi	Inter	Task	g	g	Parallel	.62
Demaray	1998	USA	None stated	47	Mixed	7.18	10-30	Multi	Inter	Task	Grw-R	Grw-R	Parallel	.82***
											Grw-W	Grw-W	Parallel	.86***
											Gq	Gq	Parallel	.66***
DiPerna	1999	USA	None stated	32	Mixed	NR	NR	Multi	Inter	Task	g	g	Parallel	.84**
Doherty	1985	UK	None stated	145	Mixed	NR	NR	Sing	Intra	Task	Gq	Gq	Parallel	.67**
											Grw-R	Grw-R		.68**
DuPaul	2004	USA	None stated	53	Mixed	8.5	NR	Multi	Inter	Task	Gq	Gq	Parallel	.38*
											Grw-R	Grw-R	Parallel	.46*

Table 4.4 continued														
Eaves, Campbell-whatley, Dunn et al.	1994	USA	None stated	89	Mixed	9.67 (2.67)	NR	Multi	Inter	Con	g Grw-R Gq Grw-W	g Grw-R Gq Grw-W	Parallel	.66** .73** .48** .59**
Eaves, Williams, Winchester et al.	1994	USA	None stated	45	Mixed	8.92 (1.67)	NR	Multi	Inter	Con	g Gq Grw-R	g Gq Grw-R	Parallel	.49** .47** .46**
Eckert	2006	USA	None stated	33	Mixed	7.3	NR	Multi Multi	Intra Intra	Task Con	Gq Grw-R	Gq Grw-R	Parallel	.19 .73***
Feinberg & Shapiro	2003	USA	None stated	30	Mixed	NR	12.8 (9.1)	Sing	Intra	Task	Grw-R	Grw-R	Parallel	.70**
Feinberg & Shapiro	2009	USA	None stated	148	Mixed	NR	12.9 (10.8)	Multi	Inter	Con	Grw-R	Grw-R	Parallel	.56**
Gabriele	2016	USA	None stated	150	Mixed	NR	10.6 (9.5)	Multi	Intra	Task	Gq	Gq	Parallel	.35*
Gut	2013	EU	None stated	402	Mixed	6.22 (0.65)	NR	Sing	Inter	Con	Gq/Grw-R/Gc	Gf	Non-parallel	.45**
Hauser-Cram	2003	USA	None stated	105	Mixed	5.5	22.8	Sing	Inter	Con	Grw-R Gq	Grw-R Gq	Parallel	.36*** .59***
Helwig	2001	USA	None stated	512	Mixed	Grade 3 (268) Grade 5 (144)	12 (NR)	Sing	Intra	Con	Gq	Gq	Parallel	.54** .62**

Table 4.4 continued

Herbert	2005	USA	None stated	378	Mixed	Grade 1	16 (NR)	Sing	Inter	Con	Grw-R	Grw-R	Parallel	.59**
											Gq	Gq		.53**
				355		Grade 3					Grw-R	Grw-R		.51**
											Gq	Gq		.65**
				345		Grade 5					Grw-R	Grw-R		.36**
											Gq	Gq		.44**
Hernandez-Torran.	2015	EU	MI Theory	566	Mixed	14.06 (1.08)	NR	Multi	Intra	Task	Gc	Gc	Parallel	.24**
											Gq	Gq		.24**
											Gf	Gf		.06
											Gv	Gv		.08
Hoge	1984	USA	None stated	298	Mixed	NR	NR	Sing	Inter	Task	Grw-R	Grw-R	Parallel	.76**
								Sing	Inter	Con	g	Gc	Non-parallel	.23**
Hodge	2006	AU	None stated	14	Mixed	5.42	NR	Sing	Inter	Con	g	Gc	Non-parallel	.76*
												Gf		.18
												Grw-R		.35
												Grw-W		-.07
Hopkins	1985	USA	None stated	1032	Mixed	NR	NR	Sing	Inter	Con	Grw-R	Grw-R	Parallel	.73**
											Gq	Gq		.72**
Hughes	2005	USA	None stated	607	Mixed	6.57 (.39)	NR	Sing	Intra	Task	Grw-R	Grw-R	Parallel	.53**
											Gq	Gq		.34**
Imbrosciano	2003	AU	None stated	87	Mixed	7.5	NR	Sing	Inter	Con	g	g	Parallel	.50*
Johansson	2012	EU	None stated	5271	Mixed	NR	Grade 3	Multi	Intra	Task	Grw-R	Grw-R	Parallel	.70***
				6044			Grade 4							.65***
Johnson	2012	UK	None stated	313	Mixed	NR	10.75	Sing	Inter	Con	Grw-R	Grw-R	Parallel	.76***
				310							Gq	Gq		.80***

Table 4.4 continued

Kenny	1993	AU	None stated	81 84	Mixed	6.83 (NR) 7.83 (NR)	6.3 (NR)	Multi	Intra	Task	Grw-R	Grw-R	Parallel	.50** .57**
Kikas	2015	EU	None stated	334	Mixed	NR	NR	Sing	Intra	Con	Grw-R	Grw-R	Parallel	.65**
Krkovic	2014	EU	None stated	1979	Mixed	12.67 (.43)	NR	Sing	Intra	Con	Gq	Gq	Parallel	.59***
Limbos	2001	NR	None stated	369	Mixed	6.43 (0.34)	15.83 (9.83)	Multi	Inter	Con	Grw-R	Gr Ga Grw-R	Non-parallel Non-parallel Parallel	.63 .61 .71
Lockl,	2017	EU	None stated	86	Mixed	6.8 (.17) 7.8 (.18)	NR	Multi	Intra	Task	Gq Grw-R Gq Grw-R	Gq Grw-R Gq Grw-R	Parallel	.43** .56** .51** .56**
Maguin	1996	USA	None stated	369 395	Male	Young NR Middle NR	NR	Sing	Inter	Con	Grw-R Gq Grw-R Gq	Grw-R Gq Grw-R Gq	Parallel	.82* .68* .77* .71*
Martin	2011	USA	None stated	76	Mixed	NR	17.6	Sing	Intra	Task	Ga	Ga	Parallel	.56*
Meissel	2017	NZ	None stated	4771 11675	Mixed	NR	NR	Sing	Inter	Con	Grw-R Grw-W	Grw-R Grw-W	Parallel	.73*** .72***
Miller	1992	USA	None stated	60	Mixed	9.85	NR	Multi	Intra	Task	Gq Gv Gc Gsm	Gq Gv Gc Gsm	Parallel	.70** .53** .57** .29*
Paleczek	2017	EU	None stated	679 621	Mixed	8.02 9.06	NR	Sing	Intra	Con	Grw-R	Grw-R	Parallel	.53 .62

Table 4.4 continued

Pedulla	1980	US	None stated	2617	Mixed	NR	NR	Sing	NR	Con	g Grw-R Gq	g Grw-R Gq	Parallel Parallel Parallel	.61 .65 .63
Pesu	2016	EU	None stated	152	Mixed	7.5 (NR)	16.0 (10.5)	Sing	Intra	Con	Gq	Gq	Parallel	.57**
Plucker	1996	USA	MI theory	1813	Mixed	NR	NR	Multi	Intra	Task	Gv Gq Gc	Gv Gq Gc	Parallel Parallel Parallel	.47** .41** .46**
Prawat	1980	USA	None stated	284	Mixed	NR	NR	Multi	Inter	Con	g	g	Parallel	.38**
Pretzlik	2003	UK	None stated	58	Mixed	10.55 (NR)	NR	Sing	Inter	Con	g	g	Parallel	.61**
Politmou	2015	EU	WM model	98	Mixed	9.0 (1.8)	NR	Multi	Intra	Task	Gwm	Gwm	Parallel	.21**
Reilly	1985	USA	None stated	26	Mixed	9.0	NR	Sing	Inter	Con	Grw-R Grw-W Gq	Grw-R Grw-W Gq	Parallel Parallel Parallel	.73 .86 .65
Rothenbusch	2016	EU	None stated	1468	Mixed	8.0 (1.0)	NR	Sing	Inter	Con	g	Gf Gc	Non-parallel	.39*** .33***
Shapiro	2017	USA	None stated	177	Mixed	Grades 3-5	12.0 (NR)	Multi	Intra	Task	Grw-R	Grw-R	Parallel	.65***
Sharpley	1986	AU	None stated	120 110	Male Female	NR	NR	Sing	Inter	Con	Grw-R Gq Gc Grw-R Gq Gc	Grw-R Gq Gc Grw-R Gq Gc	Parallel Parallel Parallel Parallel Parallel Parallel	.46*** .45*** .41*** .50*** .38*** .15***
Simone	2017	USA	None stated	157	Mixed	8	NR	Sing	Inter	Con	g	Gwm	Non-parallel	-.37***

Table 4.4 continued

Speece	2010	USA	None stated	228	Mixed	9.45	NR	Sing	Inter	Con	Grw-R	Grw-R	Parallel	.52*
Spinath	2005	USA	None stated	416	Mixed	8.8 (1.0)	NR	Multi	Intra	Con	Gq	Gf	Non-parallel	.44***
Sommer	2008	EU	None stated	47 46	Female Male	9.10 (.41)	NR	Multi	Inter	Task	g	g	Parallel	.70* .39*
Souroulla	2009	EU	None stated	165	Mixed	9.17 (NR)	NR	Multi	Intra	Task	Grw-R	Grw-R	Parallel	.45**
Südkamp	2017	EU	None stated	743	Mixed	8.42 (1.30)	NR	Sing	Inter	Con	g	Gf	Non-parallel	.37***
Triga	2004	EU	None stated	506 500	Female Male	NR	NR	Sing	Inter	Con	Grw-R	Grw-R	Parallel	.81** .84**
Uppal	2016	USA	WM model	462	Mixed	Grades 1-3	12.38 (5.52)	Multi	Intra	Task	Gwm	Gwm	Parallel	.30***
Urhahne	2011	EU	None stated	235	Mixed	9.92 (.51)	20.14 (8.87)	Multi	Intra	Task	Gq	Gq	Parallel	.61***
Vilhena	2016	EU	None stated	96 356	Mixed	7.59 (.49) 9.58 (.96)	NR	Multi	Intra	Task	Grw-R	Grw-R	Parallel	.74** .69**
Wilson	1993	USA	None stated	293	Mixed	Grade 8-12	NR	Sing	Intra	Con	g	Gc Gq	Non-parallel	.59 .47
Wright	1988	USA	None stated	195	Mixed	NR	NR	Sing	Inter	Con	Grw-R Gq	Grw-R Gq	Parallel	.71*** .71***
Zsoldos	2003	EU	None stated	60	Mixed	8.67	NR	Sing	Inter	Con	Gq Grw-R Grw-W	Gq Grw-R Grw-W	Parallel	.67** .48*** .38***

Note. MI Theory = Gardner's (1983) Multiple Intelligences theory; WM model = Baddeley & Hitch (1974) working memory model; N = sample size; Mixed = both female and male students; NR = not reported; sing = single-item; multi = multi-item; inter = inter-individual comparison; intra = intra-individual comparison; con = cognitive construct; task = cognitive task or behaviour; g = general intelligence or academic ability; Gf = fluid reasoning; Gc = crystallised knowledge; Gwm = short-term working memory; Ga = auditory processing; Gv = visual processing; Grw-R = reading; Grw-W = writing; Gq = quantitative knowledge
*p<.05, **p<.01, ***p<.00

4.3.2 Summary effect sizes

Hypothesis 1 states that the summary effect size for the relationship found between parent-reports (teacher-reports) of children's ability and performance-measures of children's ability will be statistically significant and positive. Hypothesis 2 states that there would be significant heterogeneity in the effect sizes, thereby necessitating the investigation of potential moderators.

To test these hypotheses, a summary effect size was first computed for all parent-reported cognitive ability areas combined. Multiple effect sizes from the one study were averaged when running this analysis in order to retain independence. This analysis was based on the results of 24 studies with a total sample size of 8,336. The summary effect size was $r = .41$ with a 95% confidence interval of .35-.47 ($p < .001$), and heterogeneity of results was present as evidenced by a significant Q-statistic ($Q = 248.10$, $p < .001$). The measure of I^2 was 90.73, indicating that approximately 90% of the dispersion in effect sizes was the result of true variance as opposed to error. The standard deviation of the summary effect size (T) was 0.17, indicating the presence of dispersion in the observed effect sizes.

Similarly, a summary effect size was computed for all teacher-reported cognitive ability areas combined, with multiple effect sizes from the one study averaged when running the analysis to retain independence. This analysis was based on the results of 39 studies with a total sample size of 55,853. The summary effect size was $r = .56$ with a 95% confidence interval of .52-.60 ($p < .001$). Heterogeneity between studies was present as evidenced by a significant Q-statistic ($Q = 2304.97$, $p < .001$). The measure of I^2 was 97.01, indicating that approximately 97% of the dispersion in effect sizes was a result of true variance as opposed to error. The value of 0.23 as indicted by the T measure, 0.23, suggested that dispersion was present in the observed effect sizes.

Therefore, Hypotheses 1 and 2 were supported. The finding of significant heterogeneity in the summary effect sizes of both parent- and teacher-reports indicated the need to investigate potential moderators of the observed effect sizes, outlined subsequently. Additionally, the null hypothesis that all of the dispersion between studies was due to random error within each study could be rejected. Rather results indicated that at least some of the dispersion was due to real differences in effect sizes.

4.3.3 Moderator analyses

4.3.3.1 Criterion test match (parallel vs. non-parallel)

Hypothesis 3 stated that parent-reports (teacher-reports) of children's ability would be more valid when correlated with a parallel criterion measure than when correlated with a non-parallel criterion measure. This was investigated as the first moderator for both the parent-report and teacher-report meta-analyses, and was tested first for all ability areas combined. Two studies (Hoge & Butcher, 1984; Limbos & Geva, 2001) that reported both parallel and non-parallel effect sizes were observed for the teacher-report meta-analysis, thereby violating the assumption of independent effect sizes. The two parallel effect sizes were subsequently removed from the study to retain independence of effect sizes and to ensure that tests of between groups differences could be run for both parent-report and teacher-report analyses.

Table 4.5 details the moderator analysis results of the match between informant-reports and criterion measures of ability areas. No significant moderating effect of this variable was observed for parent-reports, indicating that the use of parallel criterion measures did not significantly impact on parent accuracy when reporting on children's abilities. As for teacher-reports, significant heterogeneity observed in the parallel and non-parallel effect sizes of all teacher-reports of ability combined indicated a need for examination of this variable for each cognitive ability area. Only the ability areas of *g* and *Gq* were reported on in both groups with more than one study (e.g., there were no effect sizes reported for non-parallel measures of *Gf*, only one effect size was reported for parallel *Ga* and non-parallel *Ga*), thus greatly reducing the number of ability areas that could be assessed for between-group differences. Significant moderating effects were found for teacher-reports of *g* and *Gq*, indicating that teacher-reports of *g* and *Gq* that were correlated with parallel criterion measures of *g* and *Gq* respectively were significantly more valid than teacher-reports of these two ability areas that were correlated with non-parallel performance-based measures.

Table 4.5

Moderator Analysis Results of the Match between Parent-Reports/Teacher-Reports and Criterion Tests of Abilities

Group	*N*	*k*	*R*	95% C.I.	*Q*(df)
Match between parent-report and criterion measure (all ability areas combined)	8336	41	.42***	.37-.46	0.1(1), p= .75
Parallel	6562	37	0.39***	0.37-0.41	
Non-parallel	1774	4	0.36***	0.31-0.40	
Match between teacher-report and criterion measure (all ability areas combined)	56030	81	.55**	.52-.59	14.52(1)***
Parallel	51696	70	.58	.54-.61	
Non-parallel	4334	11	.36	.24-.47	
Teacher-reports of *g*	7992	23	.53***	.48-.57	11.22(1)***
Parallel	4921	16	.55***	.51-.59	
Non-parallel	3071	7	.29***	.12-.44	
Teacher-reports of *Gq*	11618	33	.49***	.45-.53	8.56(1)**
Parallel	10800	31	.56***	.50-.61	
Non-parallel	818	2	.44***	.39-.49	

Note. *N* = total sample; *k* = number of studies; *r* = correlation coefficient; CI = confidence interval; *Q* = measure of heterogeneity in effect sizes analysed; *g* = general cognitive ability; *Gq* = quantitative knowledge
**p <.01; ***p <.001

4.3.3.2 Child age and years of teaching experience

Hypothesis 4 states that parent-reports (teacher-reports) of older children would be more valid than those of younger children, while Hypothesis 5 states that teacher-reports from more experienced teachers would be more valid than those from less experienced teachers. A mixed-effects regression using method of moments was carried out to assess if the continuous covariates of the age of the child that was being rated (by parent and teacher) and the teacher's years of professional experience had an impact on effect sizes. Non-significant Q-statistics for all three models were obtained (see table 4.6), indicating that neither the age of the child nor how experienced the teacher was had a significant impact on parent and teacher accuracy when these adults were reporting on their children's/student's cognitive abilities. Thus, Hypotheses 4 and 5 were not supported.

Table 4.6

Mixed-Effects Meta-Regression of Child's Age and Years of Teaching Experience

Group	Q(df), p
Parent-report	
Child's age (all effect sizes combined)	2.53(1), p= 0.11
Child's age (parallel effect sizes only)	1.58(1), p= 0.21
Teacher-report	
Child's age (all effect sizes combined)	1.15(1), p= 0.28
Child's age (parallel effect sizes only)	2.17(1), p= 0.14
Years of teaching experience (parallel effect sizes only; all effect sizes combined)	1.51(1), p= 0.22

Note. Q = measure of variation in observed effect sizes explained by the covariates

4.3.3.3 Categorical moderators for parent-report

Given that no significant difference was found between parallel and non-parallel measures, moderator analyses for the remaining categorical variables were tested with all effect sizes combined. Table 4.7 presents the results for the categorical moderators tested in the parent-report meta-analysis, which were based on parent-reports of all cognitive ability areas combined. Multiple effect sizes from the one study were averaged so as to not violate the assumption of independence of effect sizes. No significant moderating effects were found for the variables of child gender or parent gender, indicating that the accuracy of parent-reports of children's abilities were not influenced by either variable. Thus Hypotheses 6 and 7 were not supported.

No significant moderating effects were found for the number of items used (Hypothesis 8) or type of comparison (Hypothesis 9), therefore these hypotheses were not supported. A significant moderating effect was found for parent-report item reference, but the significantly stronger correlation found between parent-reports and children's performance when obtained via reference to cognitive constructs was in contrast to expectations. That is, parent-report items that were based on definitions of abilities (i.e., cognitive construct) were found to result in more valid estimations than parent-report items that were based on behavioural expressions of abilities (i.e., cognitive test task). Hence, Hypothesis 10 was not supported.

Table 4.7

Moderator Analysis Results using Parent-Reports of All Cognitive Ability Areas Combined

Group/subgroups	N	k	R	95% CI	Q_b(df)
Child gender	1563	15	.47***	.38-.54	0.08(1), p = .78
Male	1153	8	.48***	.36-.58	
Female	410	7	.45***	.33-.57	
Parent gender	3072	23	.46***	0.38-0.53	0.96(1), p = .33
Mother	2046	14	.49***	.39-.58	
Father	1026	9	.42***	.30-.53	
Number of items	8366	41	.41***	.36-.45	0.18(1), p = .70
Single	2878	13	.40***	.33-.51	
Multi	5488	28	.43***	.35-.45	
Type of comparison	8336	41	.40***	.36-.45	1.34(1), p = .25
Intra-individual	5139	29	.39***	.34-.44	
Inter-individual	3197	12	.46***	.35-.54	
Parent-report item(s) based on	8336	41	0.38***	.34-.42	4.82(1)*, p =.03
Cognitive construct	2813	20	0.44***	.38-.53	
Cognitive test task	5523	21	0.36***	.31-.40	

Note. *N* = total sample size; *k* = number of studies; *r* = correlation coefficient; CI = confidence interval; Q_b = measure of heterogeneity in effect sizes analysed between studies.

**p* <.05, ****p* <.001

4.3.3.4 Categorical moderators for teacher-report

Given the initial support found for Hypothesis 3 that teacher-reports of children's cognitive and academic ability are more valid when correlated with parallel criterion tests, moderator analyses (table 4.8) were based on parallel criterion test results only. The number of CHC abilities under investigation was thus reduced to four (*g, Grw-R Grw-W* and *Gq*). Multiple effect sizes from the one study were averaged so as to not violate the assumption of independence of effect sizes. No significant moderating effects were found for student gender ($Q_b(1) = 0.003$, $p = .95$), hence Hypothesis 5 was not supported, indicating that the gender of the student being reported on for the various abilities did not affect teacher accuracy. The absence of significant moderating effects indicated that no support was found for Hypotheses 8 to 10 for teacher-reports of *g* and *Grw-R* abilities.

For teacher-reports of *Grw-W*, significant moderating effects were found which indicated that reports from teachers of children's writing ability were more valid when they were asked to compare the child to others (i.e., inter-individual comparison), than when asked to rate without a reference group (i.e., intra-individual comparison). Hence, Hypothesis 9 was supported. Results approached conventional significance ($p = .07$) for the moderators of number of items and item reference, providing partial support for Hypotheses 8 and 10. These findings indicate that teacher-report measures that have multiple items that refer to behavioural expressions of writing skills showed a trend towards stronger correlations between teacher-reports and children's measured performance of writing, than teacher-report measures with single items based on abstract definitions of writing ability.

For teacher-reports of *Gq*, significant moderating effects were found for number of items and item reference but these findings were in contrast to expectations. That is, teacher-report measures of math ability that used single items referring to cognitive constructs were found to be more valid than measures using multiple items referring to cognitive tasks. Hence, since there were no significant moderating effects found for type of comparison used, Hypothesis 8 to 10 were not supported.

Table 4.8

Moderator Analysis Results using Teacher-Reports Correlated with Parallel Criterion Tests

Group/subgroups	N	k	r	95% CI	Q_b(df)
Parallel reports of *g*					
Number of items	7561	17	.56***	.52-.60	0.24(1), *p*= .63
Single	3592	7	.55***	.41-.61	
Multi	3969	10	.57***	.51-.62	
Type of comparison	4944	16	.57***	.55-.59	0.12(1), *p*= .73
Intra-individual	3484	5	.57***	.55-.59	
Inter-individual	1460	11	.56***	.48-.63	
Item(s) based on	7561	17	.57***	.53-.61	1.80(1), *p*= .18
Cognitive construct	3950	9	.53***	.45-.60	
Cognitive test task	3611	8	.59***	.54-.63	
Parallel reports of *Grw-R*					
Number of items	26447	44	.64***	.60-.67	1.07(1), *p*= .30
Single	11751	23	.66***	.60-.70	
Multi	14696	21	.62***	.56-.67	
Type of comparison	26508	44	.64***	.60-.68	2.79(1)*
Intra-individual	14636	20	.60***	.54-.66	
Inter-individual	11872	24	.67***	.61-.71	
Item(s) based on	26673	44	.65***	.62-.69	0.64(1), *p*= .42
Cognitive construct	24028	27	.66***	.62-.69	
Cognitive test task	2645	17	.61***	.49-.71	
Parallel reports of *Grw-W*					
Number of items	11808	4	.81***	.71-.88	3.24(1), *p*= .07
Single	11761	3	.69***	.43-.84	
Multi	47	1	.86***	.76-.92	
Type of comparison	11808	4	.60***	.46-.71	12.15(1)***
Intra-individual	60	1	.38***	.14-.58	
Inter-individual	11748	3	.81***	.67-.89	
Item(s) based on	11808	4	.81***	.71-.88	3.24(1), *p*= .07
Cognitive construct	11761	3	.69***	.43-.84	
Cognitive test task	47	1	.86***	.76-.92	
Parallel reports of *Gq*					
Number of items	11113	31	.56***	.51-.60	4.83(1)*
Single	7888	18	.60***	.54-.65	
Multi	3225	13	.49***	.40-.57	
Type of comparison	11113	31	.56***	.50-.61	0.38(1), *p*= .54
Intra-individual	5839	15	.54***	.46-.61	
Inter-individual	5274	16	.58***	.49-.65	
Item(s) based on	11113	31	.56***	.52-.61	3.78(1)*
Cognitive construct	7270	18	.60***	.54-.65	
Cognitive test task	3843	13	.50***	.41-.58	

4.4 Discussion

The overarching aim in conducting the meta-analyses was to identify the human characteristics and measurement conditions that led to greater validity in parent-reports and teacher-reports of children's cognitive ability. This was assessed by the relationship (correlation coefficient) of parent- and teacher-reports with traditional performance methods of intelligence and academic achievement. Potential moderators that could influence the accuracy of informant-reports were also investigated. In total, 24 parent-report studies reporting 76 independent effect sizes, taken from a total of 8,336 participants, and 79 teacher-report studies reporting 133 independent effect sizes, taken from a total of 55,853 participants, were meta-analysed. In this section, an integration of findings from both parent-report and teacher-report meta-analyses will be presented in relation to the research hypotheses. The theoretical and practical implications of the research findings will also be delineated. Finally, the methodological strengths and limitations of the study will be discussed, accompanied with suggestions in the directions future research can take to extend these findings.

The finding of statistically significant and positive summary effect sizes between parent-reports, teacher-reports, and actual performance of children's cognitive and/or academic ability indicated that parents and teachers are capable of providing moderately accurate estimations of children's abilities (Hypothesis 1). Although there has been no previous meta-analysis of parent-reports with which to compare the current findings, the effect size found is consistent with results from empirical studies (e.g., Delgado-hachey & Miller, 1993; Hunt & Paraskevopoulos, 1980; Miller, 1986) that have come to the conclusion of parents having a reasonable level of insight on their children's cognitive functioning. This finding is reassuring since parents are inherently an important source of information about their child's functioning in various areas (e.g., behaviour, emotional functioning, etc.) for many professionals, and highlights that their utilisation as valid and reliable informants could be extended to the psychological construct of cognitive abilities as well.

The finding for the teacher-report meta-analysis is consistent with the results of previous meta-analyses conducted by Südkamp et al. (2012) and Machts et al. (2017). There are differences observed in the summary effect sizes found between the current study and these two meta-analyses. The overall effect size obtained in the current meta-analysis is comparably higher than that found in Machts et al. (2017) meta-analysis of teacher judgments of cognitive abilities ($r = .43$), but lower than that of Südkamp et al.'s (2012) review of teacher judgments of academic abilities ($r = .63$). These observed discrepancies in teacher judgments of cognitive ability and academic ability could have been due to the fact that judgment of the former requires stronger inferences than the latter (Machts et al., 2017). The higher level of familiarity that teachers have with their student's functioning in the various academic areas that involve literacy and numeracy skills, as compared to knowledge of how cognitive abilities manifest in academic tasks, could be a reasonable explanation for these differences.

The finding of significant heterogeneity between effect sizes in both meta-analyses indicated that investigating potential moderators for the strength of the relationship between parent-/teacher-reports and children's performance was warranted (Hypothesis 2). Differences in accuracy of parents and teachers when estimating a child's level of cognitive functioning may be due to individual difference factors such as child and parent gender, the teacher's professional experience, and/or methodological limitations of the informant-report measure used in the extant research. The latter may result in greater measurement error due to ambiguously worded scale items, an inadequate number of items representing the cognitive construct in question, or lack of an adequate frame of reference to accurately judge the child's cognitive performance (Fenson et al., 1994). Therefore, both individual difference and methodological factors were investigated as potential moderators in both meta-analytic reviews.

In contrast to previous research in which individual difference variables, such as the gender of the child and the parent informant, the age of the child, professional experience of the teacher, have been found to influence the accuracy of parents and teachers in their estimations of children's cognitive functioning, none of these variables were found to be significant moderators in either the parent-report or teacher-report meta-analyses. Hence, Hypotheses 4 through to 7 were not supported.

Gender stereotypical judgments made by parents and teachers are prevalent within individual studies in the existing literature, such that male children have consistently obtained higher ratings from these informants compared to female children (Furnham, 2000; Furnham, Reeves, et al., 2002; Hinnant et al., 2009; Tiedemann, 2000), and that fathers tend to provide more inaccurate estimations of children's ability than mothers (Bird & Berman, 1985; Chamorro-Premuzic et al., 2009). Findings from these individual studies may have led to conclusions that perceptions of gender stereotypes regarding cognitive ability exist, and that mothers are more aware than fathers of their child's academic progress, and thus may be more influential in determining the child's developmental pathway (Bird & Berman, 1985). These gender-related assumptions, however, were not supported by the present data. Moreover, although Rost (2009; as cited in Machts et al., 2017) proposed that teachers' judgment accuracy was higher for older students than younger students in his investigation of students' schooling level as a moderator, the current study did not find children's age to have significant effects on teacher accuracy. This is in line with results found in Machts et al.'s (2017) meta-analysis. Similarly, although some research (e.g., Babad, 1985; Hofer, 2015) has indicated that teachers who have more experience provide more accurate estimations of their students' cognitive functioning compared to less experienced teachers, this was not found in the present review.

The current study's findings hence suggest that idiosyncrasies of individual studies within the existing parent-report and teacher-report literature may have caused significant effects of gender, age, and years of teaching experience to be found at times. When applying a more robust method of inquiry such as meta-analysis, however, these differences disappear. The actual intelligence of an individual child as measured by performance-based assessment is not influenced by the gender of the child or the parent, or how experienced the teacher is. Likewise, these extraneous variables that naturally occur amongst humans should also have minimal influences on the measurement process of cognitive abilities via the parent- and teacher-report method.

Whether or not a parallel criterion test of cognitive ability was used in conjunction with a parent-/teacher-report measure of cognitive ability was assessed as a potential moderator in both analyses. Results were on either ends of the spectrum, with this condition found to be non-significant for parent-reports, but significant for teacher-reports. Partial support was thus found for Hypothesis 3. This finding suggests that using a performance-based measure of an ability area that is theoretically congruent with a teacher-report measure of the same ability facilitates more valid estimations from teachers. No significant moderating effects were found for this measurement condition for parent-reports of children's abilities, which is not consistent with results from the previous meta-analyses from the teacher-report (Südkamp et al., 2012) and self-report (Jacobs, 2012) domains. Out of all the effect sizes included in the current parent-report review, only seven (9%) were coded as using non-parallel measures. Although this could indicate that the issue of using parallel versus non-parallel measures is unfounded, significant differences may not have been detected due to a lack of power. Further empirical studies are thus needed to ascertain the effect of this variable on parent report accuracy.

That teacher-reports are more valid when compared with criterion tests measuring parallel abilities is consistent with a previous meta-analysis of teacher judgments of students' academic achievement (Südkamp et al., 2012), and that of self-reports of cognitive abilities (Jacobs, 2012). It must be highlighted that for the current study, comparisons of this condition could only be made for academic abilities (i.e., reading, writing and math) and general ability (*g*), due to a small number of studies that have investigated teachers' capacity to estimate children's functioning in specific cognitive abilities. Although there has been a previous meta-analysis (Machts et al., 2017) of teacher judgments of cognitive ability, its focus was on general intelligence, creativity and giftedness only, thus the influence that criterion test match has on how accurate teachers are when reporting on specific cognitive abilities of students is still unknown. The current study's results nevertheless provide some support for the argument that ensuring there is congruence between the ability reported by informants and the ability measured by criterion tests is important when assessing the validity of estimations from parents and teachers. Thus, researchers should endeavour to select appropriately parallel performance-based measures when investigating the convergent validity of informant-reports of children's abilities, lest erroneous conclusions about the capacity of the parent- and teacher-report methods be reached.

Unexpectedly, there was no evidence to show that the number of items (single vs. multiple) per cognitive ability construct significantly influenced the accuracy of parent-reports of children's abilities. As for teacher-reports, only significant moderating effects of this variable were found for teacher-reports of writing and mathematics ability, and not for general ability or reading ability. Hence, Hypothesis 8 was only partially supported. Multi-item measures have consistently been found to be more reliable and valid than single-item measures in a number of ways, including the capacity for computing correlations between items to establish internal consistency, as well as being more likely to tap all facets of the construct of interest (Baumgartner & Homburg, 1996; Bergkvist & Rossiter, 2007). Further, meta-analyses conducted in the self-estimates of intelligence research domain have shown that reports obtained via multi-item measures have greater validity than single-item measures

(Jacobs, 2012). Since cognitive abilities are framed as multi-faceted constructs within CHC theory, as exemplified by the hierarchical structure of broad cognitive abilities being subsumed by narrow cognitive abilities (Schneider & McGrew, 2012), this suggests that adequate construct representation of CHC cognitive abilities would be best ensured by multi-item measures.

However, that single item measures were found to be as valid as multi-item measures could have resulted from the fact that respondents were reporting on someone other than themselves (in this instance, their child or student). When individuals are asked to rate themselves using psychological scales, it is well-known that social desirability can influence responses, such that results could turn out to be overly positive (van de Mortel, 2008). This type of response bias refers to the propensity of survey respondents to select responses to questions that will be perceived in a positive manner by others, and can be reflected in the inclination to over-report "good behaviour" and under-report "bad behaviour" (Furnham, 1986). Therefore findings of the meta-analytic reviews seem to indicate that protections against the effects of socially desirable responding are not required when asking parents or teachers to report on their children. However, given that causality cannot be inferred from the findings of significant moderation in meta-analyses (Hall & Rosenthal, 1991), which is explored further below, this finding needs to be investigated further in future primary studies.

The type of comparison used in parent-report measures was not found to be a significant moderator of parent accuracy when reporting on their child's abilities. As for teacher-reports, the use of inter-individual comparisons was found to be a significant moderator for only teacher-reports of writing ability, and not for general cognitive ability, reading, or math ability. Thus, only partial support was found for Hypothesis 9, such that providing teachers with a frame of reference leads to more accurate judgments when rating children's writing ability, but not for any other ability, or any parent-reported ability. The finding of non-significance for the type of comparison used is in contrast to previous results found within the self-report domain (Freund & Kasten, 2012; Jacobs, 2012; Mabe & West, 1982), where self-reports of cognitive ability based on inter-individual comparisons were found to be more valid than intra-individual comparisons. Similarly, Macht et al. (2017) found in their meta-analytic review of teacher judgments of students' cognitive ability that accuracy was significantly lower for unreferenced judgment (i.e., teacher questionnaires that did not utilise reference norms) than for referenced judgment. The authors concluded that providing teachers with frames of reference that enforce sensitivity to comparative social norms was important for facilitating accurate teacher judgments. Thus, the well-documented finding of significant moderating effects that the measurement condition of using items with inter-individual comparison has on the accuracy in self- and teacher-reports of ability was unexpectedly, not replicated in the current study. Even so, there is some evidence that the use of social comparison when making judgments is important in facilitating valid estimations from teachers that was found in the current study. Therefore, future studies should endeavour to implement this in their informant-report design to avoid reaching erroneous conclusions about the capacity of parents and teachers to provide accurate judgments of a child's ability.

Although results approached statistical significance for teacher-report items based on cognitive test tasks in reference to writing ability to provide more valid estimations, items that referred to cognitive constructs were found to provide more valid estimations compared to cognitive test tasks in parent-reports, and for teacher-reports of math ability. Thus, Hypothesis 10 was largely unsupported. These findings are inconsistent with research within the self-report domain that has found that providing respondents with concrete behavioural examples of cognitive abilities leads to greater accuracy in reports than an abstract definition of a construct (Jacobs, 2012).

There is currently a limited number of studies using behavioural expressions of cognitive tasks in their scale design, with the parent-report scale developed by Waschbusch et al. (2000) using item wordings that describe specific cognitive tasks, based on *Gf-Gc* theory (a predecessor of CHC theory). The authors found that parents could report the cognitive strengths and weaknesses of their children with some degree of specificity, suggesting that they are able to provide equally valid estimations of both general and specific ability. Since parents, unlike teachers and professionals, could find it challenging to verbally describe their children's cognitive abilities, having them estimate their child's ability to perform cognitive test tasks that are behaviourally operationalised may assist in more valid parent-reports (Waschbusch et al., 2000). However, this result was not found, suggesting that measures using behavioural descriptions do not reduce ambiguity of item interpretation for parents. Failure to find significant evidence for greater validity of parent-reports and teacher-reports obtained via item reference to cognitive test tasks could be due several limitations of the current study, and is outlined subsequently.

4.4.1 Limitations and directions for future research

Although it can be argued the use of multi-item inter-individual parent- and teacher-report measures that use scale items referring to behavioural expressions of cognitive ability via test tasks can produce valid estimations of children's abilities from a theoretical perspective, conclusions regarding the effect of these methodological conditions on validity cannot currently be reached. The results of the current meta-analytic reviews appear to vouch for the sufficiency of measures using single items based on cognitive constructs in validly measuring children's cognitive ability via parent- and teacher-report methods. However, it is important to be cognisant of the fact that a causal role cannot be inferred for moderators identified via meta-analysis, nor from correlational data that also preclude cause and effect conclusions. This is because investigation of moderator variables in meta-analyses do not include random allocation of studies to different conditions, resulting in the possible existence of other systematic differences between these groups other than the type of response format, with these additional differences being indistinguishable from those brought about by the moderator variable (Cooper & Lemke, 1991).

Additionally, the majority of included studies for both parent- and teacher-report meta-analyses did not use an established, well-validated theoretical framework of cognitive abilities. In these studies, raters were either asked to predict a child's performance on objective measures (e.g., Miller et al., 1991), or to make judgments on overall cognitive or academic ability alone (e.g,. Aunola et al., 2003; Aunola et al., 2002). Moreover, parent-report and teacher-report scales that have been built upon the theoretical foundation of less well-validated cognitive ability theories such as Gardner's (1983) theory of Multiple Intelligence (MI theory) have been featured extensively in the extant informant-report literature (e.g., Chamorro-Premuzic et al., 2009; Furnham, 2000; Hernández-Torrano et al., 2014). Although this framework has had considerable appeal from an educational perspective (Chen, 2004; Gardner & Hatch, 1989), there has been no empirical evaluation of its validity as a whole (Allix, 2000), and it has been found to have inadequate support (Waterhouse, 2006). This indicates not only an uncertainty in the validity of conclusions made about the accuracy of parent-reports and teacher-reports from studies using these theories, but also that there is a significant limitation in the extant literature within this domain that items on informant-report measures used in the extant research may reflect inadequate coverage or irrelevant aspects of cognitive abilities. These conditions (i.e., construct under-representation and construct irrelevance) are known to be detrimental to the construct validity of an instrument (Hoyt, Warbasse, & Chu, 2006). Thus, the development of measures based on well-validated theoretical models, such as CHC theory, is needed to provide insight on the potential influence that item reference may have on the validity of informant-reports of children's cognitive ability.

Considering that different types of ability have been found to have a significant moderating effect on accuracy of self-reports (Jacobs, 2012), it would have been of great interest to investigate if a similar trend existed for parent- and teacher-reports of children's abilities. However, the majority of the previous research have focused its efforts on either general intelligence, overall academic achievement, or the specific academic abilities of reading and mathematics. Only a small number of studies (e.g., Chamorro-Premuzic et al., 2009; Hernández-Torrano et al., 2014; Miller & Davis, 1992; Miller et al., 1991) enquired about the specific abilities of fluid reasoning, crystallised knowledge, short-term working memory, and visual processing, and even fewer studies have examined processing speed, auditory processing, learning efficiency and retrieval fluency. Hence, the current study was unable to investigate if the type of ability that informants were asked to report on had moderating effects on for parent- and teacher-report accuracy due to inadequate studies available for comparison. Given what is now known about the importance of these abilities in the acquisition of literacy and numeracy skills (Flanagan, Ortiz, et al., 2013), this research domain will benefit from future studies exploring the potential variability in accuracy of parent- and teacher-reports of children's functioning when reporting on the different cognitive ability areas, instead of solely comparing specific abilities with general ability.

Although an abundance of empirical research has investigated parent- and teacher-reports of children's abilities, many studies did not compare reports from these informants with a criterion-based measure and as such, could not be included in the current meta-analyses (e.g., Furnham & Budhani, 2002; Furnham & Fukumoto, 2008; Furnham & Gasson, 1998; Furnham, Rakow, & Mak, 2002; Räty et al., 2004; Tiedemann, 2000; Tomasetto, Mirisola, Galdi, & Cadinu, 2015). Further, due to the lack of studies that included the targeted moderator variables, not all studies could be included in the moderator analyses. This was apparent for several of the human characteristics, most notably parent gender for the parent-reports, and student gender and years of teaching experience of the teacher for the teacher-reports. The systematic analysis was thus quite restricted by the lack of information on the participant samples in the primary studies. Future research on parent and teacher accuracy of children's abilities should therefore include elaborate and detailed descriptions of parent, teacher, and children characteristics.

The person variables and measurement conditions examined in this review thus do not comprise an exhaustive list of factors that may potentially influence the validity of parent- and teacher-reports of children's cognitive ability. Further, the majority of studies included in the current review were conducted in Western countries, thereby limiting the generalisability of the results to non-western populations. Studies have shown that laypersons' perceptions of what constitutes intelligence could vary across cultures and contexts (Cocodia, 2014; Sternberg & Grigorenko, 2004), with people in Western cultures being inclined to conceptualise intelligence as a means to engage in rational debate, whereas Eastern cultures perceive it as a way for members of a community to emphasise understanding and relating to others (Benson, 2003). Hence, behaviour that is considered as intelligent in one culture may be considered unintelligent in another. Future studies could investigate parent- and teacher-reports of children's CHC abilities in non-Western populations to determine the ecological validity of the theory in a different context, and if there are similarities and differences of these informant-reports to that of Western cultures.

4.5 Conclusion

The current study has provided important insights into the accuracy of informant-reports of children's cognitive abilities for both research and practical perspectives. The finding that parents and teachers are only moderately accurate informants of children's cognitive and academic functioning is likely to be a result of methodological conditions of the instrument used to obtain the information, rather than the individual differences of the child, parent, or teacher. The results bring attention to the importance of researchers choosing and/or developing methodologically appropriate instruments to use when assessing parent and teacher estimations, as this is likely to have an effect on the validity of the findings. Comparing information from parents and teachers with incongruent criterion measures could lead to drawing mistaken conclusions about the capacity of these adults to accurately report on their child's cognitive functioning. Further, future informant-report measures of children's cognitive ability should be developed based on well-validated theories that contain robust theoretical constructs such as the CHC model. Using such a theory as the basis for scale development helps to decrease construct-irrelevant variance by ensuring only items relevant to the construct are included, and that the likelihood of construct under-representation is reduced by adequate coverage of the key aspects of the cognitive ability construct (Clark & Watson, 1995; DeVellis, 2012).

From a practitioner's perspective, the results indicate that parents and teachers can be a relatively reliable source of information regarding a child's cognitive functioning. Information from these adults of children who are referred for cognitive assessment can be a valuable supplement to informing the results of the assessment process, and their involvement also provides them with the chance to divulge unique information about their child that may not otherwise have been discovered (Diamond & Squires, 1993). Since parents and teachers spend more time interacting with the child in different contexts, they also offer the benefit of providing unique data about specific types of behaviours, in a variety of settings, over a longer time than a professional would have experienced during the formal assessment process (Bagnato, Neisworth, & Munson, 1989). It is thus imperative that the development of future informant-report scales is grounded in a well-validated cognitive theory, with measurement conditions that are favourable to valid estimations of children's cognitive abilities.

Some valuable insight has thus been gained from conducting these meta-analytic reviews as presented in this chapter that pertain to the measurement conditions found to facilitate more accurate reports from parents and teachers of children's cognitive abilities. From this, efforts to develop a psychometrically robust informant-report measure of children's CHC cognitive abilities is detailed in the following chapter – the scale development process of the *Estimates of Children's Cognitive Abilities* (ECCA) instrument.

Chapter 5

Development of the ECCA

The development of a new psychological scale typically involves both qualitative and quantitative procedures. Qualitative procedures that are important components of scale development include extensive literature reviews of the construct(s) of interest, an expert review of preliminary items, and cognitive testing of items with members of the target population of the scale. Quantitative processes include descriptive univariate and bivariate statistical analyses, as well as exploratory and confirmatory factor analyses (EFA/CFA), both of which are based on Classical Test Theory (CTT). While other measurement paradigms such as item response theory (IRT) and Rasch methods are also commonly used in the quantitative aspects of the scale development process, they are beyond the scope of this thesis to explore.

This chapter reports the steps implemented for the development of the *Estimates of Children's Cognitive Abilities* (hereby referred to as ECCA) scale as a parent- and teacher-report indicator of children's cognitive functioning in eight cognitive ability areas: Fluid Reasoning (*Gf*), Comprehension-Knowledge (*Gc*), Short-Term Working Memory (*Gwm*), Learning Efficiency (*Gl*), Retrieval Fluency (*Gr*), Visual-Spatial Processing (*Gv*), Auditory Processing (*Ga*), and Processing Speed (*Gs*). These ability areas are in accordance with psychological constructs contained within the Cattell-Horn-Carroll (CHC) model of cognitive abilities, and are theorised to be important for the successful acquisition of numeracy and literacy skills. Insights gained from meta-analytic reviews of the current literature regarding the validity of parent- and teacher-reports of cognitive ability as presented in the previous chapters were applied. Specifically, a multi-item inter-individual response format was chosen for the ECCA, with items developed to reflect specific behavioural operationalisations and cognitive tasks of the constructs under investigation.

The aim of this chapter was to answer the third research question of this dissertation: do the developed ECCA scale items display adequate content relevance? To do so, this dissertation closely followed scale development processes as outlined by the Standards for Educational and Psychological Testing (American Psychological Association [APA], American Educational Research Association [AERA], & National Council on Measurement in Education [NCME], 2014), and by DeVellis (2017). The first step involved establishing clear definitions of the constructs the scale was intended to measure. This was achieved by strict adherence to CHC theory, the most comprehensive and well-validated theory of human cognitive abilities currently available (Schneider & McGrew, 2012). Existing parent- and/or teacher-report measures of constructs that are theoretically similar to the abovementioned cognitive ability areas were also evaluated. After the creation of items, the item pool was subsequently reviewed by an expert panel for relevance, clarity and conciseness. Cognitive interviews were also carried out with members of the relevant audience (i.e., teachers) for content validity evaluation. The final stage in the scale development process involved a pilot test of the newly developed measure with a small number of teachers. Figure 5.1 displays the steps and phases adhered to in the development of the ECCA scale.

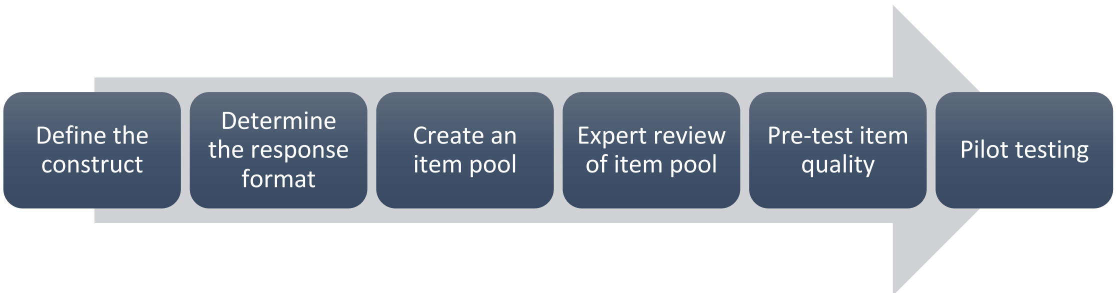


Figure 5.1 The scale development process

5.1 The Importance of Content Validity

The process of validation is the cornerstone of test development and application, since it is through this critical step that test scores take on meaning (Bensen, 1998). Validity in the context of scale development is concerned with the adequacy of a scale as a measure of a specific variable, that is, whether the latent variable shared by items in a scale is, in fact, the variable of interest to the researcher (DeVellis, 2017). The way a theoretical construct is operationalised in scale development is critical in ascertaining the validity of the resultant measure (Johnston et al., 2014). The type of validity that is most relevant to the initial stages of scale development is content validity, thus making it worthwhile to begin this chapter with a brief discussion of its importance.

Content validity refers to the degree to which the aspects of a measurement instrument is relevant to and representative of the targeted construct (Haynes, Richard, & Kubany, 1995) Streiner and Norman (1989) purport that content validity is typically established within the realm of subjective professional consensus, or require an clear process involving the establishment of a theoretical framework of the measure, and research and/or expert opinion on clinical observations. If a developed instrument contains within it a number of domains that represent a sample of the total population of domains that could be included, content validity considers how adequate this sample is, with regard to the expected functionality of the instrument (Stokes, 2011).

The measurement of an abstract psychological construct such as cognitive abilities indicate content validation could be challenging due to complex definitional boundaries (Murphy & Davidshofer, 1994). Additionally, since content validation provides an indication of the instrument's construct validity (Anastasi, 1988), and the use of a content-invalid instrument can degrade clinical inferences drawn from it (Haynes et al., 1995), the researcher should ensure that content validation of questionnaire items is determined in the initial stages of scale development. The content validation guidelines as outlined by Haynes et al. (1995) is reflected in several of the steps of the scale development process adhered to in this chapter, specifically in carefully defining the domain and facets of the construct of interest (Nunnally

& Bernstein, 1994), and using expert and population sampling for the initial item pool to establish relevance and clarity.

5.2 Defining the Constructs

Content relevance is intricately tied to the definition of the construct being examined – a scale's content should reflect the conceptual definition applied to that scale. Any scale developer that sets out to create a measure of a psychological construct should endeavour to familiarise themselves with detailed knowledge of the specific phenomenon of interest. Hence, the first step in the development of an instrument is to conduct a comprehensive review of the literature in order to develop clear conceptual definitions of the construct(s) of interest (DeVellis, 2012; McCoach, Gable, & Madura, 2013). The more researchers know about what it is they intend to measure, the better equipped they would be to develop reliable, valid, and usable scales. Messick (1995) highlights the dangers of "construct underrepresentation" that could happen should a researcher sample items too narrowly from a content domain, which would result in an inadequate reflection of what constitutes the construct. Another pitfall of "construct-irrelevant variance" could also arise from sampling items too widely that may reflect content not relevant to the specific construct. Therefore, it can be seen that how one chooses to define the construct is a critical first step that sets the tone for the subsequent phases of the scale development process – good definitions are fundamental to content validity (Delgado-Rico, Carretero-Dios, & Ruch, 2012; Haynes et al., 1995; Polit & Beck, 2006).

The deductive approach was considered to be more appropriate for the development of the ECCA, rather than the inductive method, which is preferable when investigating an unfamiliar phenomenon where little theory may exist (Schwab, 1980). Deductive scale development utilises a theoretical definition as a foundation for item generation, and requires the researcher to have a comprehensive understanding of the relevant research and literature to ensure content adequacy in the final scales (Schwab, 1980). Additionally, DeVellis (2012, 2017) strongly advocates for using "theory as an aid to clarity" (p.106), and argues for the importance of being well-grounded in the substantive theories related to the construct to be measured so as to prevent the content of the scale from inadvertently drifting to irrelevant domains.

Hence, CHC theory was chosen as the guiding theoretical framework for numerous reasons. Firstly, the conception of the ECCA began with the intention of addressing the research gap for the limited use of CHC theory as a theoretical foundation on which informant-report measures of cognitive abilities are developed within the parent- and teacher-report research domain. Although there is an abundance of cognitive ability models in existence (e.g., Gardner's (1987) MI theory, Thurstone's (1938) Primary Mental Abilities, Sternberg's (1985) Triarchic Theory of Intelligence), due to the extensive body of empirical research in support of CHC theory, it was deemed to be the most favourable. Thus, the definitions of the eight broad cognitive abilities within the model were reviewed (Flanagan, Ortiz, et al., 2013; Schneider & McGrew, 2012). Additionally, given that the ECCA instrument was designed for informants to estimate children's level of performance in cognitive ability areas, subtests from widely used intelligence batteries in cognitive ability assessments (i.e., Wechsler Intelligence Scale for Children and Woodcock Johnson Test of Cognitive Abilities) that measure children's performance in these abilities were examined as concrete operationalisations of these constructs.

Existing measures of the constructs should also be considered during this step (DeVellis, 2012). Through an extensive literature search, two multi-item informant-report measures of CHC abilities were identified and reviewed (i.e., Fiorello et al., 2009; Waschbusch et al., 2000). Several measures of intelligence constructs that were considered to be theoretically parallel to constructs contained within CHC theory were also reviewed. For example, the majority of Furnham and colleagues' work (e.g., Chamorro-Premuzic &

Furnham, 2006; Furnham, 2000; Furnham & Budhani, 2002; Furnham & Fukumoto, 2008; Furnham & Gasson, 1998; Furnham & Petrides, 2004; Furnham, Rakow, et al., 2002; Furnham & Valgeirsson, 2007) involved the creation of several self- and informant-report measures that were guided by Gardner's (1983) MI theory. As Carroll (1993) considered the constructs of Logical-Mathematical, Verbal and Spatial intelligences contained within this framework to be theoretically analogous to *Gf, Gc* and *Gv* respectively (Jacobs, 2012), these measures were located and evaluated for relevance. Appendix A lists the existing informant-report measures of cognitive ability that were reviewed at this stage of the scale development process.

5.3 Generating an Item Pool

After careful consideration of the extant literature and reviewing conceptual definitions, and the purpose of the scale has been clearly articulated, the next step in the scale development process is to develop operational definitions of the constructs to be used in the instrument. The item-writing guidelines adhered to in this dissertation was based on a combination of the work of DeVellis (2017), Clark and Watson (1995), and Hinkin, Tracey and Enz (1997), and are presented in Table 5.1.

Although items should not stray beyond the limits of the defining construct, it is recommended that all possibilities of types of items within those bounds should be exhausted, so as to ensure that the scale items accurately captures the essence of the construct (DeVellis, 2017). This implies that at this phase of the scale development process, redundancy (i.e., being more inclusive) is not necessarily a bad thing – developing seemingly repetitive and redundant items can help to reveal the phenomenon of interest in different ways. Hence, having a larger item pool is generally recommended, as the researcher has more freedom to eliminate items based on *a priori* criteria such as lack of clarity and questionable item relevance to guard against poor internal consistency of the included constructs.

Table 5.1

Item Writing Guidelines Used when Developing the Estimates of Children's Cognitive Abilities Measure

Choose items that reflect the scale's purpose
Be precise
Language should be simple and familiar to target audience (avoid jargon and colloquialisms)
Ensure reading level is appropriate for intended respondents
Avoid exceptionally lengthy items
Avoid items that represent two constructs (i.e., double-barrelled items)
Avoid irrelevant information

Note. Based on guidelines suggested by DeVellis (2017), Clark and Watson (1995) and Hinkin et al. (1997)

The current dissertation has excluded the use of both positively and negatively worded items in the creation of ECCA items, a guideline that was introduced in the initial conception of scale development processes in order to prevent response bias (van Sonderon, Sanderman, & Coyne, 2013). Weijters (2006) differentiates between two types of response bias, that being response set (bias pertaining to the content of the items) and response style (a predisposition to answer items regardless of their content), both of which can lead to a discrepancy between answers on a questionnaire and the respondents' actual state or opinion, thus posing a serious threat to the validity of the measure (Baumgartner & Steenkamp, 2001). van Sonderon et al. (2013) highlight the three types of response style that are often encountered by respondents when answering surveys: acquiescence (the inclination to agree or disagree with statements), inattention (not reading the item and response categories carefully) and confusion (inability to understand intended item content).

It has been argued by some advocates (Baumgartner & Steenkamp, 2001; Nunnally, 1978) of this practice that reversing a portion of scale items can help to minimise the effect of response bias arising from these response styles. Such items are designed to represent the reverse of the construct that is being measured, and these inverse items are created using phrases that introduce a directionality that is opposite to the construct, through the use of the word “no”, or other adjectives and adverbs that imply negative connotations (Salazar, 2015; Weijters & Baumgartner, 2012). Negatively-worded items are considered to limit the effects of bias from response style as they act to limit the tendency to agree with statements without regard for their actual content, decrease response speed, and facilitate cognitive reasoning in the respondents (Podsakoff, Mackenzie, Lee, & Podsakoff, 2003). However, studies have shown that these effects of response bias are not significant, and in fact are outweighed by the real effects of miscoding and misinterpretation by test-takers (Sauro & Lewis, 2011). Ultimately, the decision to not use positively and negatively worded items was made due to the chosen response format for the ECCA (outlined below), which did not allow for the ability to create reverse-oriented items. Based on the knowledge accumulated from reviewing previous literature, theoretical definitions, and existing CHC scales, the number of ECCA items written for each of the eight cognitive ability areas of *Gc, Gf, Gwm, Gl, Gr, Ga, Gv* and *Gs* ranged from seven to nine, resulting in 64 items in total for the initial version of the scale, presented in Table 5.2.

5.4 Determining the Measurement Format

Questionnaires are typically composed of two parts – the statement to which a participant responds (i.e., the scale item), and the response scale on which the participant expresses his or her response (Johnston et al., 2014). Both components affect content validity of the scale, indicating that determining the response format that would be used in the ECCA should be carefully considered. To ensure that the items and question format are compatible, this step occurred simultaneously with the generation of ECCA items.

The item stem is the beginning part of the scale item that presents the item as the question to be asked of the respondent, and is important to establish as it affects the wording of designed scale items. It was decided that respondents would be instructed to consider the rated child in comparison to others of the same age when responding to each item. This was intended to prompt inter-individual comparisons, as the findings of the meta-analyses detailed previously, and of research within the self-report domain revealed that accuracy of informant-reports of ability increase when scales employ inter-individual comparisons (Jacobs, 2012). The chosen item stem thus reads as follows: "*Compared to other children of the same age, the ability of the student to perform these tasks is…*" to reflect the inter-individual comparison to a specific reference group.

Next, the format of the response choices was considered. The initial response format selected was in terms of the ease or difficulty with which the child being rated could perform a particular cognitive task. After the process of cognitive interviews with members of the target audience (detailed subsequently) however, this terminology was amended to reflect one that was considered to be less subjective and more readily understood by laypersons and teachers; that being in terms of the relative ability of the child to complete a particular task. Traditional performance measures of cognitive ability typically describe the performance of test-taker relative to their same-aged peers, and use descriptive categories such as *below average*, *average*, and *above average*. Responses collected from the cognitive interviewees reflected a similar preference for using these terms. As a result, a 5-point Likert scale format was chosen, ranging from well below average to well above average, with an option for the respondent to select "not sure" to decrease the effects of uninformed response error (Dolnicar & Grün, 2014). The response anchors were thus as follows: 1 (*not sure*), 2 (*well below average*), 3 (*below average*), 4 (*average*), 5 (*above average*), and 6 (*well above average*).

5.5 Expert Review of Item Pool

The next step in the scale development process is a review of the initial generated item pool by a group of individuals who are experienced and knowledgeable in the constructs that the scale is designed to measure (DeVellis, 2012). The aim of this phase is to determine if the created items are relevant to the construct of interest, and to evaluate the clarity and conciseness of how items were worded. Inviting the panel to provide qualitative comments about each item as necessary can also provide the researcher with insight as to *why* the item may be ambiguously interpreted, and subsequently gain new perspectives on how the construct might otherwise be defined.

A group of 11 individuals (seven local and four international) who are experts in CHC theory were invited via email to participate in the expert panel by completing an online survey (included in Appendix B). They were provided with definitions of each broad ability, and asked to rate the relevance of each item to the construct it intended to measure on a 5-point scale ranging from 1 (*not at all relevant*) to 5 (*very relevant*). Conciseness and clarity ratings were measured on a 5-point scale ranging from 1 (*poor; did not understand*) to 5 (*excellent; no changes required*). For items to be included in the final scale they required ratings of at least 4 on both relevance and conciseness/clarity from at least 80% of the panel members. If an item satisfied the relevance condition but not the conciseness/clarity condition, then the item was re-written for the final version of the scale. Items that did not satisfy the relevance rating were discarded. The expert panel was also asked to propose rewordings to enhance relevance, conciseness, and clarity, as well as to suggest any additional items that they believe to be reflective of behavioural manifestations of the CHC broad ability area in question.

Table 5.2 presents the results of the expert review phase, with six members of the proposed panel returning survey responses (response rate of 54.5%). For the *Gf* subscale, five items received relevance ratings of four or five from less than 80% of the judges and so were discarded. Of the remaining items, one was retained exactly as originally written, and three were reworded based on comments from the expert panel. One additional item was suggested, “sort and classify objects according to categories”. For the *Gc* subscale, five items were discarded and of the remaining items, three were retained. Four items were reworded and one new item was introduced, “identify similarities between two common objects”.

Four items on the *Gsm* subscale were removed because of low relevance ratings. One item was retained exactly as it was written, and four items were re-worded. Only one item was retained on the *Gl* subscale; four items were reworded and four items were discarded. Four new items were written based on the expert panel’s comments. These were “use associative strategies to learn”, “remember previously unrelated information as having been paired”, “remember narratives and other forms of semantically related information”, and “associate novel information with existing information”.

Of the seven items on the *Gr* subscale, two items were retained exactly as written, while the remaining items were discarded. Based on the expert panel’s suggestions, five new items were written. These were “rapidly name examples from a given category within a time limit”, “quickly name words that share a semantic feature within a time limit”, “quickly name objects presented on a page”, “quickly retrieve information when presented with a sensory prime that relates to learnt concepts”, and “recall information related to a particular topic quickly”. Five items on the *Gv* subscale received relevance ratings of 4 or 5 from at least 80% of the expert panel, and four of these items were retained exactly as written. One was reworded, and one new item was introduced, “mentally rotate three-dimensional images in his/her mind”.

All eight items on the *Ga* subscale received relevance ratings of four or five from at least 80% of the panel members. Due to lower conciseness and clarity ratings, three items were re-worded according to the panel members' suggestions. Finally, four *Gs* items received relevance ratings from less than 80% of the panel members, so were discarded. Three items were retained exactly as written, two were re-worded, and two new items were included. These were "quickly recognise simple visual patterns" and "respond to questions in a timely manner".

Table 5.2

Relevance, Clarity, and Conciseness Ratings from the Expert panel (N = 6) of the Estimates of children's Cognitive Abilities Measure (v.1)

Item	Relevance ratings	Clarity/Conciseness ratings	Changes
Gf			
Demonstrate problem solving skills in new and everyday situations	100%	83.33%	Retain
Identify what comes next in a pattern	100%	66.67%	Identify what comes next in a visual pattern
Know what is missing from an incomplete logic puzzle	100%	50%	Select the correct missing piece to complete an incomplete logical puzzle
Know the number that comes next in a number series (e.g., 1, 2, 4, 7, 11, ...?)	83.33%	66.67%	Figure out the number that comes next in a number series
Select one of several pictures to complete a puzzle	50%	66.67%	Discard
Identify and perceive relationships (e.g., sun is to morning as moon is to ...?)	50%	50%	Identify and perceive relationships between concepts
Understand how parts fit together in a puzzle	33.33%	66.67%	Discard
Form and recognise concepts (e.g., how are a dog, cat and cow alike?)	33.33%	50%	Sort and classify objects according to categories
Make connections between new material and acquired knowledge	16.67%	66.67%	Make connections between new material and what he/she already knows
Gc			
Provide synonyms and antonyms of a word	100%	100%	Retain
Provide oral definitions for words	100%	100%	Retain
Provide correct responses to questions of fact	83.33%	100%	Retain
Point to pictures as orally instructed	83.33%	50%	Name familiar pictured objects
Understand verbal instructions	66.67%	66.67%	Discard
Understand conversations	66.67%	50%	Understand conversations of others
Display the extent of his/her general knowledge	66.67%	50%	Display general knowledge of social rules and concepts
Convey precisely what he/she is trying to say	50%	33.33%	Communicate effectively what he/she is trying to say
Participate in conversations with other children	16.67%	66.67%	Discard – identify similarities between two common objects

Table 5.2 continued

Gwm			
Repeat numbers or words orally in the same order as presented	100%	100%	Retain
Remember a phone number long enough to dial it	100%	83.33%	Retain – remember a verbally spoken phone number for long enough to write it down on paper
Remember a series of unrelated words	100%	66.67%	Listen to and repeat a list of unrelated words in correct order
Follow multi-step verbal instructions	83.33%	66.67%	Follow spoken directions or instructions to do one thing after another
Remember a series of related words	66.67%	50%	Listen to and repeat a list of related words in correct order
Retrieve all items on a list as instructed orally	50%	66.67%	Write down dictated sentences accurately
Writing down dictated information	16.67%	50%	Discard – Listen to a series of numbers and words, and repeat the words first in alphabetical order, then numbers in ascending order
Remember all materials needed for a task	16.67%	33.33%	Discard
Gl			
Absorb and retain information	83.33%	50%	Listen to and retell the principal components of an orally presented story
Accurately recall names paired with pictures of faces	66.67%	100%	Retell a story that he/she has read before
Acquire new skills	66.67%	83.33%	Remember the order in which events of the day have occurred
Relate and link learned information together	66.67%	66.67%	Link unrelated concepts to learn new information (e.g., red = stop, green = go)
Recall specific words or facts	50%	83.33%	Discard
Recall mathematical procedures	50%	83.33%	Discard
Relate new material to previous knowledge	50%	50%	Discard
Recall basic math facts	33.33%	83.33%	Discard

Table 5.2 continued

Gr			
Recall facts about what he/she has read in a book	100%	66.67%	Rapidly name examples from a given category within a time limit (e.g., things you can eat and drink)
Rapidly name letters presented on a page	83.33%	66.67%	Retain
Retell a story that he/she has learnt before	83.33%	66.67%	Quickly name words that share a phonological or semantic feature (e.g., foods that are vegetables, animals that have four legs) within a time limit
Retrieve needed information from long-term memory	66.67%	66.67%	Quickly name words that share a non-semantic feature (e.g., words that start with the letter "T") within a time limit
Remember previously learnt spelling words	66.67%	66.67%	Quickly name objects presented on a page
Recall facts for end-of-term tests and/or exams	66.67%	66.67%	Quickly retrieve information when presented with a sensory prime that relates to learnt concepts
Remember the order in which events have occurred	33.33%	66.67%	discard
Gv			
Build a model (with Lego or blocks) from a picture of the completed model	100%	100%	Retain
Put parts together to form a whole	100%	83.33%	Retain
Identify information from pictures, charts, graphs, maps, etc.	100%	66.67%	Understand information presented in a visual format (e.g., reading graphs, charts, and maps)
Differentiate between similarly shaped symbols	83.33%	83.33%	Retain
Accurately judge distances (e.g., placing objects not too close to an edge)	83.33%	83.33%	Retain
Find specific information on a printed page (e.g., getting a number out of a phone book)	66.67%	83.33%	Retain
Not be distracted by irrelevant visual information	66.67%	66.67%	Discard
Organise information from different sources into one document	16.67%	50%	Discard – mentally rotate three-dimensional images in his/her mind

Table 5.2 continued

Ga			
Understand what is being said when background noise is present	100%	83.33%	Retain
Hear the difference between similar sounding words (e.g., rhyming words)	100%	83.33%	Retain
Differentiate speech sounds (e.g. difference between /ch/ and /sh/)	100%	83.33%	Retain
Blend letter sounds together fluently to form words	100%	66.67%	Blend auditory presented letter sounds fluently to form words
Recognise a word only when parts of it are pronounced	100%	66.67%	Identify a word when only parts of it are pronounced
Hear the different sounds in words	100%	66.67%	Hear the internal structure of sound in words
Know which sounds go with which letters	83.33%	83.33%	Retain
Filter out background noise to listen to the teacher's voice in the classroom	83.33%	83.33%	Retain
Gs			
Quickly compare how different or similar two objects are	100%	66.67%	Retain - quickly compare or scan visual information (e.g., letters, words, numbers, symbols, etc.) for similarities and differences
Work quickly and efficiently on tasks already mastered	83.33%	83.33%	Retain – work quickly and efficiently on routine tasks
Quickly and accurately copy information from the whiteboard	83.33%	66.67%	Retain
Scan and quickly determine important information on a page	83.33%	66.67%	Retain
Complete a series of arithmetic problems within a time limit	83.33%	50%	Complete a series of basic mathematical calculations within a time limit
Solve known math questions automatically	66.67%	83.33%	Discard – quickly identify identical items (e.g., letters, numbers, objects) within a series of items
Complete mastered tasks in a timely manner	66.67%	66.67%	Discard
Display automaticity of rote information	66.67%	66.67%	Quickly recognise simple visual patterns
Complete classwork in a timely manner	50%	83.33%	Discard

Note. Only relevance and clarity/conciseness ratings of 4 or 5 were tabulated as percentages

5.6 Pre-testing Scale Items via Cognitive Interviewing

As established in the literature review chapter of this dissertation (Chapter 2), qualitative results obtained from pretesting scale items via cognitive interviewing can provide valuable information to improve the scale development process. This work presented within the following section progresses cognitive interviewing as a means to assessing the semantic and conceptual equivalence of the ECCA. Table 5.3 displays the guidelines as recommended by Patrick et al. (2011) that were followed for the cognitive interview.

Table 5.3

Guidelines for Cognitive Interviewing

Step	Key points
Determine context of interviews	Identify target population
Design the interview	Semi-structured interview allowing exploration of item clarity, the relationship between intended and comprehended meaning, difficulties with wording and instruction, alternate suggestions and identification of inappropriate wordings
Conduct cognitive interview	Allow respondent to verbalise (think-aloud) about each item (probe as required)
	Recording for transcription and analysis is recommended

Note. Based on guidelines suggested by Patrick et al. (2011)

Potentially problematic items as identified in the previous step were selected for pre-testing in the first round of cognitive interviews conducted with psychologists (see Table 5.4 for the complete list of items). Most of these items received high relevance ratings but low conciseness/clarity ratings from the expert panel. Two registered psychologists, one male and one female, were asked to participate in individual, semi-structured interviews lasting from approximately 45 to 120 minutes. The length of interview time was dependent on the nature of the answers provided. All interviews were audio-recorded with the permission of the respondent for ease of transcription. For each CHC ability area, they were asked to describe examples of behaviours that would be displayed if a child had high or low levels in that particular ability. They were then given the list of selected ECCA items and asked if that item was clear or unclear, and how well the item explored the intended construct (well or poor).

Table 5.4 provides a summary of the results from the cognitive interviews with the psychologists. A transcript of cognitive interviewing with one participant is included in Appendix C. Out of the 35 items that were tested, six were deemed to have satisfactory clarity and relevance ratings, and did not undergo any changes. Fifteen items required rewording due to technical jargon being used in the items (e.g., “logical puzzle”, “pre-determined categories”, “known prime”), and nine items were deemed to be acceptable if examples of the cognitive task that accompanied the scale item were either added or changed. A total of six items were discarded and replaced with suggestions of other behavioural manifestations of the ability area that the respondents believed were important to the construct.

Table 5.4

Summary of Responses from Cognitive Interviewing with Psychologists (N = 2)

ECCA scale items probed	Clarity	Relevance	P1 Responses	P2 Responses	Changes
Know what is missing from an incomplete logical puzzle	P1: Unclear P2: Unclear	P1: Well P2: Well	"Logical puzzle" might be tricky for respondents. Provide examples. Do you mean visual puzzle/pattern?	What is a logical puzzle? Provide concrete example. How would the respondent know what the child knows?	Select the correct missing piece to complete a logical puzzle (include pictorial example)
Identify and perceive relationships between new concepts	P1: Unclear P2: Clear	P1: Well P2: Well	"Concept" is quite a technical term	-	Add pictorial example
Select one of several pictures to complete a puzzle	P1: Unclear P2: Unclear	P1: Well P2: Well	This item can also mean jigsaw. Listed example is a verbal concept which younger children might not know	What do you mean by puzzle here? Seems to be the same item as the first.	Discard
Sort and classify objects according to pre-determined categories	P1: Clear P2: Clear	P1: Well P2: Well	-	Change pre-determined, sounds technical	Sort and classify objects according to a given category
Make connections between new material and what he/she already knows	P1: Clear P2: Clear	P1: Unsure P2: Well	Sounds like it is related to Gl	-	Use previously acquired knowledge to solve new problems
Understand conversations of others	P1: Clear P2: Unclear	P1: Well P2: Well	-	Peers or adults?	Understand others' conversations (peers and adults)
Provide correct responses to questions of fact	P1: Clear P2: Clear	P1: Well P2: Well	-	-	No change
Participate in conversations with other children	P1: Clear P2: Clear	P1: Well P2: Well	-	-	No change
Display the extent of his/her general knowledge	P1: Unclear P2: Clear	P1: Well P2: Well	General knowledge is quite broad	-	Display general knowledge of learnt social rules
Point to pictures as orally instructed	P1: Clear P2: Clear	P1: Unsure P2: Poor	More appropriate for younger children	Older children won't do this	Correctly name pictures of familiar objects
Communicate effectively what he/she is trying to say	P1: Clear P2: Clear	P1: Well P2: Well	-	-	No change
Remember a series of related words in the correct order	P1: Clear P2: Clear	P1: Poor P2: Well	Why does it have to be related words?	-	Repeat numbers or words orally in the same order as presented

Table 5.4 continued

Remember a verbally spoken phone number for long enough to write it down	P1: Clear P2: Clear	P1: Well P2: Well	People don't really memorise phone numbers anymore	-	No change
Follow spoken directions or instructions to do one thing after another	P1: Unclear P2: Clear	P1: Well P2: Well	How many directions do you mean?	-	Follow instructions to do two or more tasks (e.g., pick up your shoes and put your clothes in the basket)
Write down dictated sentences accurately	P1: Unclear P2: Clear	P1: Well P2: Clear	What do you mean by "accurate"? Verbatim is the word you're after I believe.	-	Write down dictated sentences accurately, ignoring spelling and punctuation errors
Listen to a series of numbers and words, and repeat the words first in alphabetical order, then numbers in ascending order	P1: Clear P2: Clear	P1: Poor P2: Poor	This is not a typical classroom task so teachers would not know how to answer this.	Parents and teachers wouldn't know how to answer this	Discard
Link unrelated concepts to learn new information (e.g., red = stop, green = go)	P1: Unclear P2: Clear	P1: Well P2: Well	The example there is confusing – this association is already so entrenched in our culture so it's not immediately obvious that these colours and actions are actually not related	Might want to come up with a better example	Link unrelated concepts to learn new information (e.g., linking language symbols and sounds to learn how to read)
Associate novel information with existing information (e.g., mnemonics)	P1: Unclear P2: Clear	P1: Well P2: Well	"Mnemonics" is technical. Provide a literal example and seems to be repetitive of another item	-	Discard
Listen to and retell the principal components of an orally presented story	P1: Clear P2: Clear	P1: Well P2: Well	Change "principal components"	-	Listen to and retell important parts of an orally presented story
Remember the order in which events of the day have occurred	P1: Unclear P2: Clear	P1: Well P2: Well	Is the sequence important? Events of the day seems quite broad to me	-	Remember the order of events that happened in a movie or TV show
Use associative strategies (e.g., mental imagery, acronyms) to learn new information	P1: Unclear P2: Clear	P1: Well P2: Well	Provide examples of mental imagery and acronyms	-	Use associative strategies to learn new information (e.g., acronym ROYGBIV for colours of the rainbow)
Quickly name objects presented on a page	P1: Clear P2: Clear	P1: Well P2: Well	-	Saying recognise rather than name implies ability to say the word if needed	Quickly recognise objects presented on a page
Rapidly name examples from a given category within a time limit (e.g., things you can eat and drink)	P1: Clear P2: Clear	P1: Well P2: Well	Change example, same as WJ III COG	Should emphasise quantity rather than time limit	Rapidly give numerous examples from the same category (e.g. fruit and vegetables)
Retrieve learnt information when presented with a known prime that relates to concepts	P1: Unclear P2: Unclear	P1: Well P2: Well	What is a "known prime"?	"Known prime" is quite technical	Quickly retrieve connected information when presented with a cue

Table 5.4 continued

Not be distracted by irrelevant visual information	P1: Clear P2: Clear	P1: Well P2: Well	Provide an example, like playing Where's Wally	-	Provide "Where's Wally" as an example
Put parts together to form a whole	P1: Unclear P2: Clear	P1: Poor S P2: Well	What does a whole mean? Do you mean putting a jigsaw puzzle together?	-	Put parts of an image together to form the whole image (e.g. jigsaw puzzle)
Understand information presented in a visual format (e.g., reading graphs, charts, and maps, etc.)	P1: Clear P2: Clear	P1: Well P2: Well	-	-	No change
Mentally rotate three-dimensional images in his/her mind	P1: Unclear P2: Unclear	P1: Well P2: Well	How would the rater know if the child is able to do that? Provide example?	It's relevant to Gv but the rater won't know if the child can "mentally rotate"	Discard
Find specific information on a printed page (e.g., getting a number out of a phonebook)	P1: Clear P2: Clear	P1: Well P2: Well	Not many people use phonebooks these days. Change example	Phonebooks aren't used these days. Change to finding something in a book or computer screen.	Find specific information on a printed page (e.g., finding a specific word in a passage of text)
Blend auditory presented letter sounds fluently to form words	P1: Unclear P2: Clear	P1: Well P2: Well	"auditory presented letter sounds" is quite technical	-	Hear letter sounds and blend them fluently to form words
Hear the internal structure of sounds in words	P1: Clear P2: Clear	P1: Well P2: Well	How will parents know if the child can hear it or not?	Use different sounds to make up or form words	Identify the beginning, middle, and ending sounds in words
Identify a word when only parts of the word is pronounced	P1: Clear P2: Clear	P1: Poor P2: Poor	Performance-based task that teachers don't really see in the classroom	Teachers don't do this with their students	Discard Sound out unfamiliar words
Scan and quickly determine important information on a page	P1: Clear P2: Clear	P1: Well P2: Well	-	-	No change
Complete a series of basic mathematical calculations within a time limit	P1: Clear P2: Clear	P1: Well P2: Well	Reword to "simple maths sums" and provide example	-	Quickly complete a series of simple maths sums (e.g., 3+2, 5+4, etc.)
Quickly recognise simple visual patterns	P1: Clear P2: Unclear	P1: Poor P2: Poor	Doesn't really happen in the classroom.	Don't really know what you mean by visual pattern	Discard Respond to simple questions in a timely fashion

Note. P1 = Participant 1; P2 = Participant 2

After the first round of cognitive interviewing with psychologists, the revised scale, ECCA v.3, had a total of 55 scale items. Verbal probes for the second round of cognitive interviewing were developed as per the guidelines recommended by Willis (1999). Table 5.5 contains the basic categories of cognitive probes, and an example of each that was developed for the cognitive interview. Two female educational specialists, one specialising in the field of literacy and the other in numeracy, were invited to participate by completing the online survey and subsequent cognitive interview, lasting from approximately 1 to 2 hours, and took place over a video call via Skype. The respondents were given a copy of their recorded responses to refer to during the interview. The definitions of each broad cognitive ability area with their corresponding narrow ability areas were provided (see Appendix E) and clarified as required.

Table 5.5

Examples of the Types of Verbal Probes Developed for Cognitive Interviewing

Type of verbal probe	Example
Comprehension/interpretation probe	What does the term “problem-solving skills” mean to you?
Paraphrasing	Can you repeat the question in your own words?
Confidence judgment	How sure are you that the student can display this behaviour?
Recall probe	When have you seen the student complete this type of task?
Specific probe	Do you think that being able to understand others’ conversation is important to the construct of crystallised knowledge?
General probe	Was this item easy or hard to answer? How did you arrive at your answer?

The respondents were asked to comment on each item with regard to comprehension, including specific words or phrases that made it hard to understand what the item was asking, and the conceptual nature of the item, which included its relevance to the intended construct it was measuring. Each item was read to the respondent and further explanation was given if required by the respondent to facilitate comprehension. Items found to be fully comprehensible and relevant to the construct were not explored further within the interview. Further semi-structured interviewing was conducted with items that were identified as problematic for the respondent, that is, if the item deemed to be unclear and/or a poor measurement of the intended construct. This focused on clarifying the difficulties with the item, identification of key words and phrases of concern and alternative wording for items (Hilton & Skrutkowski, 2002).

A summary of the results from the cognitive interviews with the educational specialists is displayed in Appendix E. Content analysis of the two interviews revealed that 30 out of the 55 items were easily understood by respondents and required no further clarification. They were consequently retained for the final version of the scale. There were issues with clarity in terms of how the item was worded and difficulties in comprehension due to the inclusion of technical jargon, requiring amendments to items by either adding clearer examples or minor re-wording. One *Gl* item was discarded ("retell narratives and other forms of semantically related information") as it was deemed to be too technical and repetitive of previous items related to *Gl*, and two items were added based on suggestions made from the respondents. After these revisions, the final version of the ECCA scale had a total of 58 items, with each broad cognitive ability construct containing six to nine items (see Appendix F for final version of the scale).

5.7 Pilot Testing

The last step in the scale development process involved pilot testing, as recommended by DeVellis (2012). Hence, a convenience sample was recruited consisting of five government primary school teachers (two male and three female), whose years of teaching experience ranged from six to more than 20 years. The five children (one male and four female) who were reported on ranged from 5 to 14 years old ($M = 9.6$, $SD = 3.29$). As such, the pilot sample was representative of the intended target audience of the ECCA. Participants were asked to complete the measure and note any difficulties they had when responding to the questions.

Instructions that were used in the Self-Report Measure of Cognitive Abilities (SRMCA) as developed by Jacobs (2012) were retained and adapted where necessary for use in the ECCA scale's instructions. Hence, after questions pertaining to demographic information of the child and the informant (teacher), the instructions presented to the pilot sample participants were:

> *The ECCA questionnaire is designed to measure a student's level of cognitive functioning in different cognitive ability areas. Cognitive abilities are the mental skills that people use to process information. There are several different areas of cognitive ability, and children tend to vary in their levels across the different areas (i.e., people tend to be high in some cognitive abilities and low in others, rather than high in all or low in all). One example of a cognitive ability area assessed in this survey is* ***learning efficiency****, which is the ability and efficiency with which an individual can learn, store, and consolidate new information in long-term memory.*
>
> *The statements below will ask you to rate, from well below average to well above average, the student's ability when performing certain tasks. When responding to each item please compare the student to* ***OTHER CHILDREN HIS/HER AGE.*** *It is important that you answer as honestly as possible. There are no right or wrong answers, you just need to pick the response that you believe best describes the student.*
>
> *If you believe that the student is capable but does not perform the task or behaviour any longer* ***due to having grown out of it,*** *award full credit for the item (i.e., "well above average"). After rating each cognitive task, you will then be asked if it was easy or difficult to provide a rating for that item. Please note down any difficulties*

that you may have had (e.g., wording of the item is unclear, have not observed the child performing the behaviour, etc.) in the text box provided.

Following the instructions, the 58 ECCA items were presented to respondents as one scale, ability by ability. No difficulties comprehending the instructions, items, or response format were reported, hence, this version of the ECCA was retained as the finalised scale to be used in the subsequent studies of this dissertation.

5.8 Conclusion

This chapter detailed the development of the ECCA describing each step that was involved in order to optimise the construct and content validity of the measure. Including the additional step of pre-testing scale items with cognitive interviewing techniques proved to be a worthwhile endeavour. By pre-testing with both psychologists (the test user) and teachers (the test-takers), this step indicated a progression towards the successful translation of psychological jargon into layperson language, an important issue that has remained unaddressed in the development of similar measures that currently exist within the domain of parent- and teacher-report research.

After a scale has been created and pilot tested, the next phase in the scale development process is to administer it to a development sample. This allows for an empirical determination of the underlying factor structure of the scale, as well as an evaluation of the performance of the individual items (DeVellis, 2012; Spector, 1992). The following chapter details the third study carried out as part of this dissertation, which sought to investigate the initial validity of the ECCA.

Chapter 6

Scale Dimensionality with EFA

Following on from the creation of items to comprise the Estimates of Children's Cognitive Abilities (ECCA) scale as detailed in the previous chapter, this chapter presents the first stage of data analysis for investigating the construct validity of the ECCA to answer the fourth research question of this dissertation: does the ECCA display internal validity? Specifically, the focus was on determining whether the factorial structure of the developed ECCA sufficiently replicated a priori expectations of eight factors corresponding to the CHC broad abilities of Fluid Reasoning (*Gf*), Comprehension-Knowledge (*Gc*), Short-Term Working Memory (*Gwm*), Learning Efficiency (*Gl*), Retrieval Fluency (*Gr*), Visual-Spatial Processing (*Gv*), Auditory Processing (*Ga*), and Processing Speed (*Gs*). Such confirmation would provide evidence in support of the structural and construct validity of the ECCA.

6.1 Investigating Construct Validity through Factor Analysis

The need to determine the construct validity of psychological theories and measures used to reflect the constructs contained within them was established by several authors (Campbell & Fiske, 1959; Cronbach & Meehl, 1955; Loevinger, 1957) in the middle of the 20th century. Their seminal work in the concept of construct validation included advancing the legitimacy of hypothetical psychological constructs (MacCorquodale & Meehl, 1948), developing evidence for the validity of measures of such constructs (Cronbach & Meehl, 1955), and identifying the construct validation procedure as a general framework for the development and testing of theories and their resultant measures (Loevinger, 1957). Currently, construct validity is generally considered to be the cornerstone of establishing the validity of psychological measurements (Strauss & Smith, 2009), with some (e.g., Landy, 1986) believing it subsumes both content and criterion validity.

Therefore, it can be inferred that construct validity forms the backbone of what gives a measure its practicality and clinical utility, and should be respected by scale developers and users alike. After establishing the content relevance of items in a newly developed scale, the next phase is to investigate its construct validity. Ascertaining the structural validity of scale items and discriminant validity between scale factors can be accomplished through the use of factor analyses. Factor analysis is a family of statistical techniques that is used widely in scale development research to simplify and reduce a large number of items into a more concise, reliable, and conceptually rigorous measurement instrument (Hooper, 2012). This is carried out by modelling observed variables (i.e., the items on the scale) and their covariance structure in terms of a smaller number of latent factors. Latent factors are typically viewed as broad ideas that may describe an observed phenomenon – in this dissertation, these concepts are the eight broad abilities as contained within CHC theory. Using factor analysis enables the researcher to determine if a series of factors exist in the data, and if they are interpretable in a theoretical sense. Factor analysis techniques have proved their utility in models where numerous correlated variables occur simultaneously, and where the goal of the study is to determine the underlying sources of covariance among the data (Trninić, Jelaska, & Štalec, 2013), and to ascertain the minimum number of latent dimensions required to accurately account for the common variance among items (Reise, Waller, & Comrey, 2000).

A central issue that scale developers face is the decision to use exploratory factor analysis (EFA) or confirmatory factor analysis (CFA) to investigate the structural validity of a newly developed scale. Traditionally, EFA has been used to investigate the possible underlying factor structure of a set of observed variables without forcing a predetermined structure on the outcome (Child, 1990), and the technique is useful where there are no prior examinations into the construct validity of a scale (Tabachnick & Fidell, 2013). However, several researchers have suggested that CFA is more appropriate than EFA in the initial stages of scale validation when the development of the measure has been theoretically driven, as CFA allows for the testing of the proposed theoretical expectations of the number or nature of the factors in question (Henson & Roberts, 2006; Williams, Brown, & Onsman, 2010). This was the case with the ECCA, as the development of the scale items was guided specifically by CHC theory.

In the case of scale development research using both EFA and CFA methods to investigate the construct validity of a measurement tool, it is recommended that researchers should use two separate samples, or split a large sample into two. This permits replication of the factor structure of the scale via cross-validation studies by applying EFA to the first sample, then confirming the structure by means of CFA on the other half (Brown, 2006). However, the participant sample size that was obtained for this dissertation precluded splitting the sample into two smaller samples which would have been large enough to reliably conduct the analysis techniques. Rules regarding the ideal sample size for factor analysis is not straightforward because an exact minimum cannot easily be found analytically (Pearson & Mundform, 2010), and the current literature is fraught with disagreements on how large the sample size, or the appropriate sample (N) to survey item (p) ratio, should be (MacCallum & Widaman, 1999). What is generally agreed upon amongst researchers though, is that factor analysis is subject to sampling error associated with inadequate sample size, since factor analysis is based on correlation matrices (Lingard & Rowlinson, 2006; MacCallum & Widaman, 1999). In empirically testing the impact of sample size on results of factor analyses, Costello and Osborne (2005) reported that larger samples tend to produce more accurate solutions, and only 10% of samples with small sample to survey item ratios (i.e., 2:1) produced correct solutions.

With a sample size of $N = 309$ and the number of ECCA scale items being $p = 58$, the N:p ratio was approximately only 5:1 for the current study. Therefore, to avoid the effects of sampling error, a two-step process was used to assess evidence of the structural validity in the newly developed ECCA scale. The first step was to conduct exploratory factor extraction to make an initial decision and prediction about the number of potential factors underlying the ECCA scale. Identifying the factor structure with exploratory methods provides an initial theoretical understanding of how the 58 items on the ECCA can be combined within groups to create several constructs within the instrument. The second step was to confirm the structure using CFA methods in a subsequent study, specifically, utilising structural equation modelling (SEM), in accordance with the eight cognitive ability constructs defined within CHC theory. Although EFA helps to determine the initial dimensionality of the measure, it only provides evidence of a theoretical structure (Besnoy, Dantzler, Besnoy, & Byrne, 2016), and the resulting structure is primarily data-driven, rather than by hypothesis (Ockey, 2014). Hence, CFA was used to endorse the proposed factor structure in accordance with CHC theory, and to provide further evidence of construct validity of the ECCA (Fabrigar & Wegener, 2012).

6.2 Method

6.2.1 Participant sample details

A comment must be made about the decision regarding participant recruitment for the current study. The original intention of this dissertation was to develop an informant-report measure of children's cognitive abilities that could be completed by the parent and teacher of the child that was being referred for psychoeducational assessment, which meant that both parent and teacher participants were to be recruited for scale validation purposes. However, time constraints often associated with the timely completion of a PhD did not allow for the recruitment of both types of participants, due to the integration of pretesting potentially problematic scale items via cognitive interviewing techniques into the scale development process, as was detailed in the previous chapter. Nonetheless, this additional qualitative step proved to be a meaningful endeavour, since it further improved content validity of the scale items before subjecting the test to more extensive tests of validity. It was therefore decided that only teachers would be recruited for scale validation purposes at this stage of the project, with the hope that future studies can conduct further validation studies of the ECCA scale with parent participants.

Hence, to be eligible for participation in the study, each participant needed to be a registered teacher, and to have taught a school-aged student (5 to 18 years old) within the past 12 months at the time of taking the survey. Participants consisted of a convenience sample recruited online via distribution of the survey link by school principals, and by snowball sampling.

Online participant recruitment for the purposes of data collection is an efficient and convenient alternative to the paper-and-pencil method, as they allow for a greater participant reach in a short time frame, thus saving on time and costs (Lefever, Dal, & Matthíasdóttir, 2007; Wright, 2005). It also guards against accidental data loss and facilitates an easy data transfer into a database for analytical purposes (Carbonaro & Bainbridge, 2000; Ilieva, Baron, & Healey, 2002). Online surveys are also beneficial for the respondent, as they are able to complete the survey at the time and place of their convenience, and the anonymity and sense of social distance created by the internet can help with the discussion of sensitive topics (Brindle, Douglas, Hundley, & van Teijlingen, 2005; Douglas, McGarty, Bliuc, & Lala, 2005). However, some argue that online data collection can also lead to the systematic exclusion of potential participants without internet access and who may not be computer literate (Brindle et al., 2005; Dillman, 2007; Jones, Murphy, Edwards, & James, 2008), making data from e-questionnaires less generalisable. Nonetheless, research indicates that the demographic profile of internet users is becoming more inclusive (Granello & Wheaton, 2004), suggesting that concerns over the potential biasing impact of this data collection method are less relevant. Additionally, such limitations of online recruitment are unlikely to apply for data collection that focuses on teacher participants.

Data collection commenced in October 2017 and concluded in January 2018, with a total of 497 participants having commenced the survey. Of these, 138 participants had not completed any items beyond demographic information, and a number had failed to complete the survey. All cases (*n* = 38) with 30% or more missing data across all ECCA items were excluded. Due to the large number of scale items, items that belonged to the same broad ability area were treated as a separate subscale, and all cases (*n* = 10) with 30% or more missing data in each subscale were deleted. This resulted in a final participant sample comprising 309 cases. The demographic characteristics of both the teacher sample and student sample can be found in Tables 6.1 and 6.2. Although this participant group is suitable for the purpose of initial scale validation, a final instrument would ultimately need to be validated on a larger and more carefully stratified sample.

Table 6.1

Demographic Details of the Participant Sample (Teachers) for Scale Validation Purposes (N= 309)

Characteristic	Statistic	Teacher sample
State	N (%)	309 (100%)
VIC		306 (99.0%)
NSW		1 (.3%)
QLD		1 (.3%)
SA		1 (.3%)
Years of teaching experience	N (%)	309 (100%)
1 to 2 years		24 (7.8%)
3 to 5 years		43 (13.9%)
6 to 9 years		55 (17.8%)
10 to 15 years		46 (14.9%)
16 to 20 years		35 (11.3%)
More than 20 years		106 (34.3%)
School type	N (%)	309 (100%)
Independent or private		42 (13.6%)
Government		196 (63.4%)
Catholic		63 (20.4%)
Other (specify)		7 (2.3%)
Not reported		1 (.3%)
Year level taught	N (%)	309 (100%)
Primary		194 (62.8%)
Secondary		105 (34.0%)
Primary & secondary		9 (2.9%)
Not reported		1 (.3%)
Role in the school	N (%)	309 (100%)
Classroom teacher		223 (72.2%)
Learning support teacher		27 (8.7%)
Other		58 (18.8%)
Not reported		1 (.3%)
Attended workshop on learning difficulties	N (%)	309 (100%)
No		22 (7.1%)
Yes		286 (92.6%)
Not reported		1 (.3%)
Recency of learning difficulties workshop	N (%)	286 (100%)
Less than 6 months		124 (43.4%)
7 to 12 months		80 (28.0%)
1 to 2 years		57 (19.9%)
3 to 5 years		18 (6.3%)
More than 5 years		7 (2.4%)
Gender	N (%)	309 (100%)
Male		47 (15.2%)
Female		260 (84.1%)
Unspecified		1 (.3%)
Not reported		1 (.3%)

Table 6.2

Demographic Details of the Participant Sample (Students) for Scale Validation Purposes (N= 309)

Characteristic	Statistic	Student sample
Gender	N (%)	309 (100%)
Male		206 (66.7%)
Female		102 (33.0%)
Unspecified		-
Not reported		1 (.3%)
Age (year)	N	309
	Mean (SD)	11.06 (3.53)
	Min – Max	5.0 – 18.0
Grade	N (%)	309 (100%)
Foundation/Prep		24 (7.8%)
Year 1		21 (6.8%)
Year 2		31 (10.0)
Year 3		36 (11.7%)
Year 4		30 (9.7%)
Year 5		28 (9.1%)
Year 6		27 (8.7%)
Year 7		17 (5.5%)
Year 8		22 (7.1%)
Year 9		22 (7.1%)
Year 10		16 (5.2%)
Year 11		17 (5.5%)
Year 12		24 (4.5%)
Senior		1 (.3%)
VCAL		1 (.3%)
Ungraded 5-8year old group		1 (.3%)
Current performance in academic area	N (%)	309 (100%)
Reading		
Below the standard expected		158 (51.1%)
At the standard expected		87 (28.2%)
Above the standard expected		63 (20.4%)
Missing		1 (.3%)
Writing		
Below the standard expected		196 (63.4%)
At the standard expected		70 (22.7%)
Above the standard expected		42 (13.6%)
Missing		1 (.3%)
Speaking & Listening		
Below the standard expected		128 (41.4%)
At the standard expected		130 (42.1%)
Above the standard expected		49 (15.9%)
Missing		2 (.6%)
Mathematics		
Below the standard expected		154 (49.8%)
At the standard expected		93 (30.1%)
Above the standard expected		57 (18.8%)
Missing		5 (1.6%)

6.2.2 Materials

Estimates of Children's Cognitive Abilities (ECCA). This 58-item measure was detailed extensively in the previous chapter and is included in Appendix F. It was presented to participants using the same instructions employed during pilot testing of the measure (refer to Section 5.7).

6.2.3 Procedures

Ethics approval for this study was received from the Monash University Human Research Ethics Committee (refer to Appendix G). The 58-item ECCA scale was placed online using Qualtrics Survey Software and a web-link to the survey was included in emails and posts on social networking sites such as Facebook. Permission to conduct research within schools was also obtained from the governing bodies of government and Catholic schools in Victoria, which were the Department of Education and Training Victoria (Appendix H) and Catholic Education Melbourne (Appendix I), respectively. Unlike their government and Catholic counterparts, there was no requirement to obtain permission from any governing body before contacting principals of private and/or independent schools.

After receiving letters of approval from the relevant governing bodies, a database of current school principals and contact information of schools in Victoria was compiled by the primary researcher. A total of 1,965 principals of primary and secondary schools were contacted via email to participate in the study (10.3% private, 16.7% Catholic, 73.0% government). Each school principal was sent a letter (Appendix J) explaining the survey and research project in greater detail, and an explanatory statement (Appendix K) about the project for participating teachers. Principals were asked to forward the link to the online survey to teaching staff at their school. Participation was anonymous, and as an incentive to partake in the research, participants were given the choice to enter into a raffle for a fifty-dollar gift card.

6.2.4 Data Analysis Protocol

Williams et al. (2010) recommend implementing a statistical analysis protocol prior to running EFA to prevent any potential oversights in the process. Figure 6.1 illustrates the Five-Step Exploratory Factor Analysis Protocol developed by Williams et al. (2010) that was a reference point for the current study.

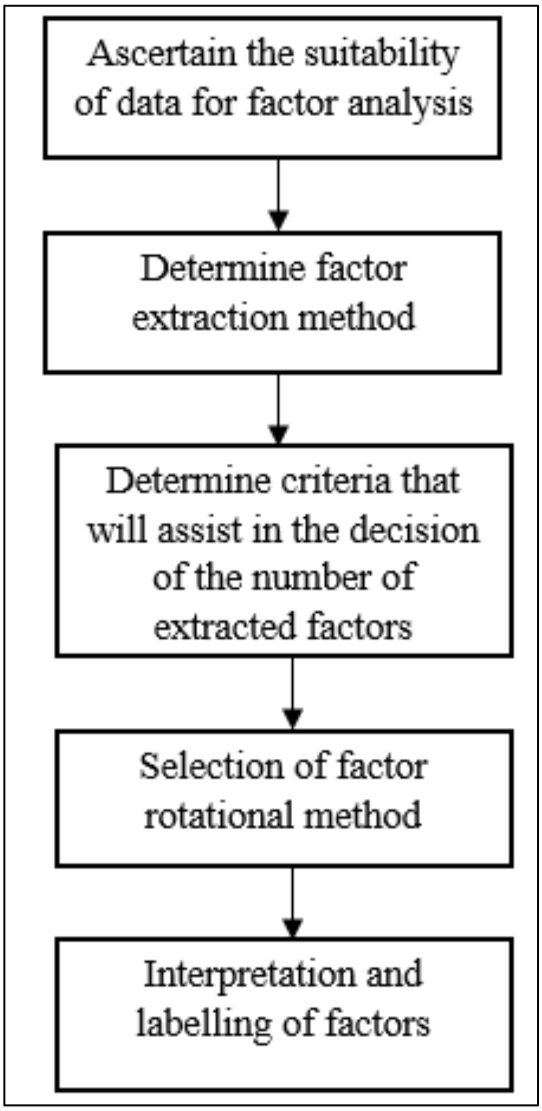


Figure 6.1 The 5-step Exploratory Factor Analysis Protocol

Adapted from "Exploratory factor analysis: A five-step guide for novices," by B. Williams, T. Brown, & Onsman, A. 2010, *Australasian Journal of Paramedicine*, 8(3), p. 4. Adapted with permission.

The first step of the EFA protocol relates to the suitability of the data for factor analytic techniques. As mentioned previously, there are varying opinions regarding the ideal sample size, with differing guiding rules of thumb cited in the literature. Tabachnick and Fidell (2007) recommend a sample as large as 300, whereas Hair, Anderson, Tatham, and Black (1995) propose a less stringent criterion of 100 cases as adequate for factor analysis. Moreover, the guidelines proposed by Comrey and Lee (1992) are also often cited: 100 as poor, 200 as fair, 300 as good, 500 as very good, and more than 1000 as excellent. Another set of guidelines pertaining to the sample (number of participants) to variable (number of variables) ratio (N:p) can also serve as a reference point for researchers; however, there exists equally disparate recommendations as to the ideal ratio, ranging anywhere from 3:1 to 20:1 (Gorusch, 1983; Hair et al., 1995; Pett, Lackey, & Sullivan, 2003; Tabachnick & Fidell, 2007). It can thus be seen that solely depending on the sample size as a guide for factor analysis suitability can lead to erroneous decisions. Thus, a visual inspection of the correlation matrix of the variables, where the majority of coefficients should be more than 0.30, was also referred to, along with two other statistical measures, to assess the suitability of the respondent data for factor analysis. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy (Kaiser, 1974) and Bartlett's Test of Sphericity (Bartlett, 1950) tests were conducted, with the KMO value being above 0.50 and a significant Bartlett's Test of sphericity (p <.05) used as criterion.

Step two and three of the protocol relates to the factor extraction method. The goal of both extraction methods of principal components analysis (PCA) and principal axis factoring (PAF) are to maximise the variance extracted by orthogonal factors (Tabachnick & Fidell, 2007). However, some authors have speculated that the reason for the popularity of PCA is due to it being the default option in statistical software programs (Costello & Osborne, 2005), rather than for its statistical properties. Rather than being a method of factor analysis, PCA instead reduces the dimensions that reject measurement errors, which often results in an overestimation of factor loadings and the variance explained by the factors (Ferrando & Anguiano-Carrasco, 2010). In contrast, PAF aims to produce a set of factor loading estimates that is as close as possible to reproducing the common variance within a correlation matrix (de Winter & Dodou, 2012) and is usually preferred when the goal is theoretical explorations of the underlying factor structure. Thus, PAF was conducted on the correlation matrix as the theoretical foundation of this extraction method is more favourable for scale development (Worthington & Whittaker, 2006). Nevertheless, it is worth noting that robust theoretical factor structures tend to emerge regardless of the extraction method used (Tabachnick & Fidell, 2007).

Of all the decisions made in exploratory factor analysis, determining the number of factors is perhaps the most critical, because incorrect specification of extracted factors will result in obscuring the factor structure of the scale (Izquierdo, Olea, & Abad, 2014; Watkins, 2006). Multiple researchers (Costello & Osborne, 2005; Fabrigar, Wegener, MacCallum, & Strahan, 1999; Patil, Singh, Mishra, & Donavan, 2008) have shown that the Kaiser criterion (1960), which is to retain factors with eigenvalues greater than 1.0, usually recommends an excessive number of factors as compared to other procedures, even though it is often a default setting in most statistical packages. Although an acceptable alternative, the scree plot approach, is highly recommended by some (e.g., Costello & Osborne, 2005) due to its visual nature, accessibility and the straightforward process of creating the plot, in situations where multiple changes in slope or a more gradual change of slope exist, interpretation of the scree plot becomes much more challenging. Additionally, because this method is centred around each individual's perception of where the "elbow" of the plot levels out, its interpretation is inherently subjective (O'Conner, 2000).

An alternative to these criteria is the well-accepted Parallel Analysis method (Horn, 1965) that involves the generation of randomised data sets with the same number of variables and cases as the data set of interest (Hayton, Allen, & Scarpello, 2004), and this technique has shown good results in the extant literature (Fabrigar et al., 1999; Hayton et al., 2004; Humphreys & Ilgen, 1969; Humphreys & Montanelli Jr., 1975). Multiple criteria were thus used within this dissertation to determine the number of factors to extract. Using the software program *Monte Carlo* (Watkins, 2000) for parallel analysis, factors were retained if the eigenvalue from the actual data was larger than the corresponding eigenvalue extracted from the random data. The scree plot was also visually inspected, and theoretical considerations about the number of factors the ECCA was created to measure were taken into account.

The fourth step of the EFA protocol pertains to the factor rotational method, in which the researcher considers if a variable might relate to more than one factor (Williams et al., 2010). Rotation maximises high item loadings and minimises low item loadings, thereby producing a more interpretable and simplified scale structure. Following theoretical expectations that universally, specific cognitive abilities areas are all positively correlated (known as the "positive manifold"; Spearman, 1904), the data was obliquely rotated using Promax rotation so as to allow the ECCA factors to be correlated (Vogt, 1993).

The final step involves running the EFA, and the subsequent interpretation of the resulting factors. Once factor analyses are conducted, Kline (1994) recommends regarding factor loadings as high if they are greater than 0.60, and moderately high if values are above 0.30. Item loadings, cross-loadings and communalities were examined to determine which items to retain and which to delete. A more stringent criterion of factor loadings being greater than .40 for items to be retained, as suggested by (Floyd & Widaman, 1995), was used due to the smaller than recommended sample size (i.e., having less than 10 participants per item). Additionally, items found to have communality values below .30 should be deleted as recommended by Pallant (2011), as such items are considered to not fit well with the other items in its factor. Finally, as for the cross-loading criterion, a difference between the primary target loading and secondary loading of at least .30 is typically recommended (Costello & Osborne, 2005). Cross-loadings are indicative of a variable that has an inherently complex latent nature (Yong & Pearce, 2013), much like cognitive ability (e.g., solving a jigsaw puzzle involves skills subsumed under both *Gf* and *Gv* broad ability areas). Hence, cross-loadings were visually inspected but this guideline was not adhered to definitively in the current study.

6.3 Results

All analysis was conducted using SPSS (version 25). Across all ECCA scale items, there was an average of 0.92% missing data, and no item exceeded 2% missing data. Little's MCAR test was run separately for each subscale that contained nonresponses. There were no missing values for the *Gr* and *Gv* subscales, and results indicated that no significant differences were obtained between variable means before and after applying the EM imputation method, therefore it was concluded that there was no pattern for the missing data across the remaining *Gf, Gc, Gwm, Gl, Ga* and *Gs* subscales.

The data was screened for univariate outliers by converting item scores to z-scores and identifying cases which had item z-scores of an absolute magnitude greater than 3.29 ($p < .001$; Tabachnick & Fidell, 2007). No univariate outliers were identified. Multivariate outliers were then examined using Mahalanobis' distance. The maximum value of 180.57 was significantly higher than the critical value of 97.03 ($df = 58$, $p < .001$), with a total of 42 cases exceeding the critical value. An inspection of Cook's distance and Leverage scores indicated that there were no cases demonstrating significant and undue influence on the data, so ultimately, no cases were removed.

Table 6.3 presents the descriptive and distributional statistics for each of the ECCA items. Absolute skew and kurtosis values were less than 2 and 7 respectively for all items, indicating that the data was univariate normal (Curran, West, & Finch, 1996; George & Mallery, 2010). All 58 items were found to be satisfactory based on the criterion of .30 being the minimum acceptable corrected item-total correlation (Nunnally & Bernstein, 1994).

Table 6.3

Descriptive and Distributional Statistics for the Hypothesised CHC Broad Abilities in the Estimates of Children's Cognitive Abilities Measure

Item	*M* (SD)	Skew[a]	Kurtosis[b]	*r*IT
Gf subscale (Cronbach's α = .93)				
Gf.Q1 Come up with a solution when confronted with a problem	3.53 (1.19)	0.32	-0.74	.76
Gf.Q2 Figure out the number that comes next in a number series	3.53 (1.47)	0.05	-0.85	.70
Gf.Q3 Think logically to solve a problem	3.42 (0.12)	0.12	-0.80	.74
Gf.Q4 Sort and classify objects according to a given category	3.83 (1.26)	-0.33	-0.27	.76
Gf.Q5 Identify relationships between concepts	3.75 (1.29)	0.14	-0.52	.77
Gf.Q6 Use previously acquired knowledge to solve new problems	3.68 (0.12)	0.12	-0.80	.80
Gc subscale (Cronbach's α = .94)				
Gc.Q1 Provide synonyms and antonyms of a word	3.58 (1.30)	0.06	-0.59	.79
Gc.Q2 Provide oral definitions for words	3.59 (1.22)	0.21	-0.49	.80
Gc.Q3 Demonstrate good general knowledge about the world around them	3.82 (1.24)	-0.04	-0.72	.76
Gc.Q4 Correctly name pictures of familiar object	4.28 (1.03)	-0.35	0.87	.74
Gc.Q5 Engage in conversations with others (peers and adults)	4.04 (1.13)	0.03	-0.41	.62
Gc.Q6 Display general knowledge of learnt social rules	4.05 (1.13)	-0.06	-0.23	.69
Gc.Q7 Communicate effectively what he/she is trying to say	3.77 (1.14)	0.31	-0.54	.71
Gc.Q8 Identify similarities between two common objects	3.98 (1.16)	-0.14	-0.09	.77
Gwm subscale (Cronbach's α = .93)				
Gwm.Q1 Remember the numbers given in a verbally spoken math problem for long enough to note it down on paper	3.30 (1.38)	0.12	-0.81	.77
Gwm.Q2 Listen to and repeat a list of words in correct order	3.31 (1.26)	0.17	-0.48	.81
Gwm.Q3 Remember two or more instructions in the correct order	3.70 (1.17)	0.31	-0.57	.85
Gwm.Q4 Follows instructions to do two or more tasks	3.74 (1.18)	0.26	-0.56	.81
Gwm.Q5 Write down dictated sentences accurately, ignoring spelling and punctuation errors	3.34 (1.27)	0.45	-0.63	.80
Gwm.Q6 Accurately solve verbally presented math sums by using mental arithmetic	3.24 (1.40)	0.28	-0.80	.77
Gl subscale (Cronbach's α = .95)				
Gl.Q1 Listen to and retell the important parts of an orally presented story	3.67 (1.20)	0.04	-0.48	.79
Gl.Q2 Remember the order of events that happened in a movie or TV show	3.78 (1.27)	-0.37	-0.24	.75
Gl.Q3 Recall facts about what he/she has read in a book	3.72 (1.25)	0.01	-0.50	.82
Gl.Q4 Remember information presented visually	3.96 (1.14)	-0.32	0.22	.76
Gl.Q5 Retain new information after only a few learning trials	3.48 (1.28)	0.44	-0.65	.83
Gl.Q6 Learn and remember the relationship between unrelated objects	3.51 (1.22)	0.13	-042	.79
Gl.Q7 Apply effective strategies to learn new information	3.27 (1.23)	0.54	-0.35	.80

Table 6.2 continued

Gr subscale (Cronbach's α = .95)				
Gr.Q1 Quickly name objects presented on a page	3.90 (1.14)	-0.37	0.15	.79
Gr.Q2 Quickly name letters presented on a page	3.95 (1.25)	-0.38	-0.09	.80
Gr.Q3 Quickly name whole words presented on a page	3.72 (1.32)	-0.02	-0.66	.83
Gr.Q4 Rapidly give numerous examples from the same category	3.62 (1.31)	-0.01	-0.56	.79
Gr.Q5 Quickly name connected information when presented with a visual or verbal cue	3.60 (1.25)	0.07	-0.54	.81
Gr.Q6 Rapidly give examples of words that share a semantic feature	3.43 (1.32)	0.12	-0.60	.79
Gr.Q7 Quickly recall information related to a learnt topic	3.60 (1.18)	0.30	-0.62	.84
Gv subscale (Cronbach's α = .93)				
Gv.Q1 Build a model from a picture of the completed model	3.74 (1.38)	-0.53	-0.35	.51
Gv.Q2 Put parts of an image together to form the whole image	3.63 (1.36)	-0.47	-0.38	.63
Gv.Q3 Differentiate between similarly shaped symbols	3.74 (1.36)	-0.30	-0.35	.78
Gv.Q4 Pick out relevant information in the presence of distracting visual stimuli	3.51 (1.30)	-0.25	-0.25	.70
Gv.Q5 Extract visual information from graphs and handouts	3.50 (1.27)	-0.05	-0.50	.76
Gv.Q6 Interpret information presented in a visual format	3.56 (1.25)	0.03	-0.52	.77
Gv.Q7 Accurately judge distances between objects	3.60 (1.30)	-0.27	-0.43	.68
Gv.Q8 Find specific information on a printed page or computer screen	3.49 (1.25)	0.11	-0.65	.84
Ga subscale (Cronbach's α = .97)				
Ga.Q1 Sound out unfamiliar words	3.58 (1.31)	0.16	-0.72	.79
Ga.Q2 Know which sounds go with which letters	3.82 (1.27)	-0.13	-0.46	.78
Ga.Q3 Understand what is being said when background noise is present	3.66 (1.17)	0.10	-0.32	.76
Ga.Q4 Differentiate between speech sounds	3.78 (1.24)	-0.11	-0.38	.82
Ga.Q5 Hear the difference between similar sounding words	3.79 (1.20)	0.09	-0.25	.83
Ga.Q6 Blend letter sounds fluently to form words	3.70 (1.29)	-0.01	-0.58	.81
Ga.Q7 Filter out background noise to listen to my voice in the classroom	3.58 (1.21)	0.16	-0.38	.76
Ga.Q8 Identify the beginning, middle, and ending sounds in words	3.72 (1.31)	-0.07	-0.54	.82
Ga.Q9 Manipulate sounds within words to form new words	3.54 (1.36)	0.07	-0.72	.84
Gs subscale (Cronbach's α = .94)				
Gs.Q1 Quickly complete a series of simple maths sums presented on a worksheet	3.49 (1.52)	0.01	-0.99	.71
Gs.Q2 Quickly and accurately copy information from the board	3.32 (1.28)	0.55	-0.56	.78
Gs.Q3 Respond to simple questions quickly	3.96 (1.11)	0.06	-0.38	.72
Gs.Q4 Quickly identify differences and similarities between two objects	3.85 (1.18)	0.05	-0.26	.80
Gs.Q5 Work quickly and efficiently on routine tasks	3.57 (1.26)	0.37	-0.65	.76
Gs.Q6 Scan and quickly determine important information on a page	3.18 (1.33)	0.75	-0.42	.84
Gs.Q7 Completes a task in the same amount of time as their peers	3.25 (1.39)	0.70	-0.76	.83

Note. *N* = 309. *r*IT is the corrected item-total correlation for each item with the remaining items of the subscale. Gf = fluid reasoning, Gc = comprehension-knowledge; Gwm = short-term working memory; Gl = learning efficiency; Gr = retrieval fluency; Gv = visual processing; Ga = auditory processing; Gs = processing speed

[a]Skew *SE* = .14

[b]Kurtosis *SE* = .28

Inspection of the correlation matrix of all ECCA items (Table 6.5) revealed the presence of many coefficients of .30 and above. The Kaiser-Meyer-Olkin measure of sampling adequacy index was .98, exceeding the recommended value of .60 (Kaiser, 1974), and Bartlett's (1950) test of sphericity was significant, supporting the factorability of the correlation matrix. Table 6.4 details the results of a parallel analysis with five hundred generated datasets for the ECCA data using the *Monte Carlo* software program. Parallel analysis revealed that the eigenvalues for the first three factors in the actual data (35.47, 3.26 and 2.14 respectively) were clearly larger than those for the random data (1.97, 1.88, and 1.81 respectively). The eigenvalues for the remaining five factors was lower than that observed for the simulated data. These results thus indicate that a three-factor solution was most optimal. The pattern of eigenvalues in the scree plot (Figure 6.2) also indicated a three-factor solution, despite the measure being constructed to measure eight latent constructs. Based on these considerations, it was decided that the analysis should be rerun with the specification of three obliquely rotated factors to be extracted.

Table 6.4

Parallel Analysis of the Estimates of Children's Cognitive Abilities EFA Sample

Factor	Adjusted eigenvalue	Unadjusted eigenvalue
1	1.9707	35.467
2	1.8810	3.263
3	1.8115	2.144
4	1.7520	1.393
5	1.7012	1.127
6	1.6499	.978
7	1.6063	.955
8	1.5652	.822

Note. Five hundred random datasets were generated. ECCA = Estimates of Children's Cognitive Abilities; EFA = exploratory factor analysis

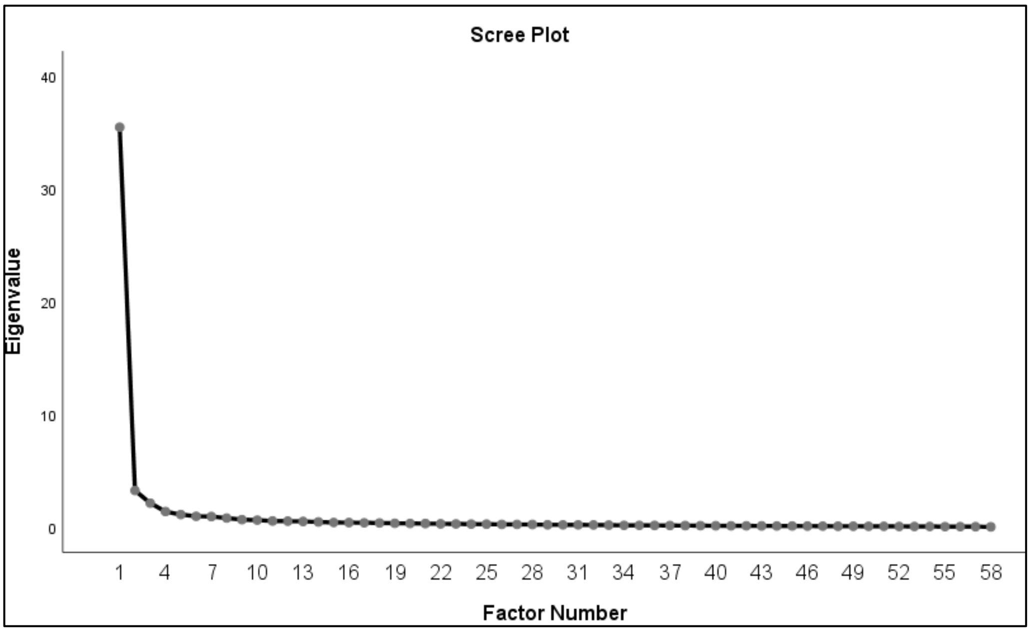


Figure 6.2 Scree plot of eigenvalues from principal axis factoring of 59-item Estimates of Children's Cognitive Abilities measure

Table 6.5

Bivariate Correlation Coefficients of Estimates of Children's Cognitive Abilities Scale Items in EFA Sample (N = 309)

	Gf.1	Gf.2	Gf.3	Gf.4	Gf.5	Gf.6	Gc.1	Gc.2	Gc.3	Gc.4	Gc.5	Gc.6	Gc.7	Gc.8	Gwm.1	Gwm.2
Gf.1	-															
Gf.2	.55	-														
Gf.3	.62	.78	-													
Gf.4	.62	.75	.75	-												
Gf.5	.66	.61	.70	.75	-											
Gf.6	.78	.62	.67	.70	.77	-										
Gc.1	.62	.47	.53	.51	.63	.65	-									
Gc.2	.64	.46	.55	.56	.59	.65	.83	-								
Gc.3	.62	.45	.57	.52	.58	.66	.69	.76	-							
Gc.4	.56	.46	.50	.52	.57	.61	.68	.69	.72	-						
Gc.5	.57	.28	.35	.43	.48	.53	.55	.58	.60	.55	-					
Gc.6	.62	.36	.44	.47	.54	.60	.58	.64	.65	.59	.75	-				
Gc.7	.63	.37	.42	.47	.53	.56	.59	.64	.62	.58	.75	.72	-			
Gc.8	.65	.47	.50	.53	.60	.62	.69	.73	.65	.68	.61	.73	.72	-		
Gwm.1	.58	.76	.70	.68	.61	.62	.56	.51	.47	.49	.34	.40	.42	.51	-	
Gwm.2	.55	.62	.64	.61	.58	.55	.64	.64	.60	.59	.42	.50	.53	.63	.76	-
Gwm.3	.68	.55	.59	.64	.63	.70	.65	.68	.66	.62	.61	.65	.71	.68	.66	.72
Gwm.4	.64	.49	.54	.59	.63	.66	.63	.64	.65	.63	.57	.70	.66	.67	.59	.69
Gwm.5	.61	.46	.51	.53	.54	.61	.66	.69	.60	.58	.52	.61	.65	.64	.59	.69
Gwm.6	.61	.85	.77	.73	.52	.67	.53	.52	.49	.47	.35	.42	.42	.49	.83	.65
Gl.1	.63	.44	.54	.56	.62	.61	.71	.74	.67	.62	.567	.56	.63	.67	.54	.68
Gl.2	.50	.52	.56	.54	.56	.56	.62	.64	.57	.56	.46	.47	.50	.55	.63	.69
Gl.3	.59	.49	.54	.56	.61	.64	.74	.74	.68	.64	.54	.55	.59	.65	.57	.69
Gl.4	.53	.46	.53	.57	.52	.54	.62	.66	.64	.60	.52	.54	.57	.60	.54	.68
Gl.5	.67	.54	.58	.54	.61	.69	.68	.71	.70	.64	.55	.60	.65	.65	.60	.69
Gl.6	.63	.46	.49	.54	.61	.67	.68	.68	.63	.60	.50	.57	.58	.63	.57	.62
Gl.7	.67	.46	.54	.52	.60	.64	.65	.69	.63	.56	.56	.63	.64	.62	.57	.61
Gr.1	.59	.63	.61	.66	.63	.63	.62	.61	.57	.69	.46	.48	.50	.57	.68	.67
Gr.2	.52	.64	.58	.65	.61	.60	.64	.61	.54	.60	.42	.46	.51	.56	.69	.70
Gr.3	.57	.60	.58	.61	.63	.64	.68	.65	.61	.59	.44	.49	.54	.60	.67	.73
Gr.4	.60	.60	.63	.61	.64	.65	.62	.61	.61	.61	.44	.50	.51	.58	.65	.63
Gr.5	.61	.55	.59	.56	.64	.65	.66	.67	.69	.66	.52	.58	.61	.63	.59	.66
Gr.6	.60	.52	.54	.54	.58	.64	.65	.64	.62	.59	.48	.51	.52	.60	.58	.67

Gwm.3	Gwm.4	Gwm.5	Gwm.6	Gl.1	Gl.2	Gl.3	Gl.4	Gl.5	Gl.6	Gl.7	Gr.1	Gl.2	Gr.3	Gr.4	Gr.5	Gr.6
-																
.87	-															
.77	.75	-														
.65	.58	.57	-													
.68	.69	.71	.52	-												
.61	.59	.62	.60	.74	-											
.71	.70	.69	.57	.82	.77	-										
.65	.65	.64	.52	.74	.77	.79	-									
.75	.74	.73	.61	.70	.66	.76	.70	-								
.69	.68	.73	.54	.69	.66	.69	.65	.74	-							
.71	.68	.78	.56	.69	.64	.73	.69	.77	.76	-						
.65	.61	.56	.66	.63	.58	.63	.59	.59	.63	.55	-					
.62	.60	.61	.64	.59	.59	.67	.61	.63	.62	.55	.85	-				
.68	.64	.65	.62	.63	.63	.72	.67	.71	.69	.64	.82	.88	-			
.61	.56	.57	.64	.62	.60	.65	.56	.61	.58	.62	.75	.71	.72	-		
.66	.65	.64	.58	.66	.63	.67	.64	.70	.62	.65	.69	.68	.72	.81	-	
.64	.60	.67	.57	.70	.63	.67	.61	.62	.67	.70	.67	.64	.65	.75	.78	-

Table 6.5 continued

	Gr.6	Gr.7	Gv.1	Gv.2	Gv.3	Gv.4	Gv.5	Gv.6	Gv.7	Gv.8	Ga.1	Ga.2	Ga.3	Ga.4	Ga.5
Gr.6	-														
Gr.7	.72	-													
Gv.1	.42	.39	-												
Gv.2	.49	.48	.79	-											
Gv.3	.58	.60	.56	.68	-										
Gv.4	.58	.54	.62	.66	.71	-									
Gv.5	.58	.63	.53	.62	.69	.64	-								
Gv.6	.59	.64	.46	.55	.64	.62	.86	-							
Gv.7	.54	.52	.55	.61	.68	.60	.65	.61	-						
Gv.8	.70	.71	.41	.52	.68	.61	.65	.68	.63	-					
Ga.1	.63	.68	.35	.41	.55	.45	.53	.54	.45	.72	-				
Ga.2	.63	.64	.30	.43	.56	.48	.52	.51	.47	.70	.87	-			
Ga.3	.62	.62	.33	.39	.53	.48	.52	.52	.49	.60	.64	.65	-		
Ga.4	.68	.64	.36	.44	.62	.55	.58	.61	.51	.74	.80	.84	.72	-	
Ga.5	.66	.67	.38	.46	.63	.54	.46	.57	.54	.72	.81	.84	.73	.90	-
Ga.6	.66	.67	.32	.43	.58	.46	.56	.57	.49	.73	.85	.86	.70	.85	.88
Ga.7	.58	.66	.33	.40	.59	.52	.53	.54	.51	.63	.59	.59	.80	.67	.70
Ga.8	.65	.64	.35	.45	.65	.54	.52	.53	.50	.71	.79	.81	.70	.86	.85
Ga.9	.70	.67	.32	.45	.63	.56	.56	.58	.50	.73	.81	.83	.69	.85	.86
Gs.1	.52	.56	.45	.55	.73	.61	.65	.63	.59	.61	.51	.51	.43	.54	.54
Gs.2	.59	.69	.30	.43	.60	.46	.60	.60	.48	.66	.60	.60	.64	.62	.61
Gs.3	.59	.66	.35	.44	.52	.47	.54	.56	.49	.59	.49	.49	.59	.54	.55
Gs.4	.60	.71	.44	.54	.65	.55	.50	.61	.59	.65	.57	.56	.57	.62	.63
Gs.5	.56	.12	.35	.43	.58	.48	.58	.61	.51	.62	.53	.49	.61	.53	.57
Gs.6	.69	.76	.35	.44	.60	.56	.64	.66	.50	.77	.69	.66	.64	.68	.67
Gs.7	.60	.76	.34	.45	.60	.51	.63	.65	.51	.73	.65	.63	.67	.64	.64

Ga.6	Ga.7	Ga.8	Ga.9	Gs.1	Gs.2	Gs.3	Gs.4	Gs.5	Gs.6	Gs.7
-										
.67	-									
.85	.66	-								
.84	.65	.92	-							
.51	.51	.58	.58	-						
.64	.68	.60	.65	.58	-					
.55	.63	.53	.56	.49	.67	-				
.61	.66	.65	.67	.57	.67	.76	-			
.57	.70	.54	.59	.52	.77	.75	.73	-		
.69	.65	.66	.72	.57	.75	.64	.69	.73	-	
.67	.69	.63	.68	.59	.81	.67	.70	.81	.87	-

The resulting three extracted factors accounted for 70.47% of the shared variance, with Factor 1 explaining 61.15% of the variance, Factor 2 explaining 5.63%, and Factor 3 explaining 3.70%. Although the three-factor solution as revealed by the pattern matrix (Table 6.5) was not in accordance with a *priori* theoretical expectations, it could be loosely described as collections of abilities that are in line with Carroll's (1993) factor-analytic survey of human cognitive abilities. Factor 1 was composed of 21 items, Factor 2 was consisted of 21 items, and Factor 3 contained 16 items. The highest loading items on Factor 1 were all items that contained cognitive constructs (i.e., *Ga*, *Gl*, and *Gr*) that were not measured in Wechsler intelligence batteries, and so was labelled MISC. The highest loading items of Factor 2 were all items that were related to verbal ability, and so was labelled Gc. The highest loading items of Factor 3 were all items that were relating to the *Gf/Gv* hybrid factor as reported by Carroll (1993), and was thus labelled Gf/Gv.

In terms of item removal, Gr.Q5 *quickly name connected information when presented with a visual or verbal cue* was eliminated as it did not contribute to a simple factor structure and failed to meet the criterion of having a primary factor loading of .4 or above. Visual inspection of cross-loading also revealed that multiple items had a difference of less than .3 between factors. Traditionally, items with cross-loadings of less than .3 difference between factors are recommended to be removed, since cross-loading indicates poor discriminant validity for the item (Tabachnick & Fidell, 2007). However, this suggestion was not observed in the production of the final version of the scale, and reasons for this decision is further explained in the Discussion section. Instead, a content analysis and in-depth exploration as to why some of these items may not have fit cleanly onto just one factor is detailed in the subsequent section of this chapter.

Correlations between the three factors were strong, indicating that the decision for an oblique rotation was appropriate for the types of constructs under consideration. Specifically, correlations between the MISC factor and Gf/Gv and Gc factors were r = .79

and $r = .73$ respectively, and the correlation between the Gf/Gv and Gc factors was $r = .73$. A general rule of thumb for a scale to demonstrate adequate internal consistency is that Cronbach's alpha is greater than .70, with alphas greater than .80 being considered as very good. Therefore, the level of internal consistency present in each of the ECCA subscales was highly acceptable.

ECCA item	Factor Loading			h^2
	MISC	Gc	Gf/Gv	
Ga.Q2 know which sounds go with which letters	**1.08**	-.14	-.09	.83
Ga.Q8 identify the beginning, middle and ending sounds in words	**.97**	-.13	.05	.82
Ga.Q1 sound out unfamiliar words	**.96**	-.01	-.09	.78
Ga.Q6 blend letter sounds fluently to form words	**.95**	.04	-.11	.82
Ga.Q9 manipulate sounds within words to form new words	**.93**	-.03	.01	.84
Ga.Q4 differentiate between speech sounds	**.91**	-.03	.01	.81
Ga.Q5 hear the difference between similar sounding words	**.90**	.02	-.01	.82
Gr.Q3 quickly name whole words presented on a page	**.69**	-.08	.31	.76
Gl.Q3 recall facts about what he/she has read in a book	**.62**	.26	.02	.72
Gr.Q2 quickly name letters presented on a page	**.62**	-.21	.49	.74
Gc.Q1 provide synonyms and antonyms of a word	**.61**	.30	-.05	.68
Gl.Q6 learn and remember the relationship between unrelated objects	**.52**	.34	.01	.67
Gc.Q2 provide oral definitions for words	**.51**	.42	-.07	.70
Gl.Q2 remember the order of events that happened in a movie or TV show	**.51**	.09	.23	.59
Gc.Q4 correctly name pictures of familiar objects	**.49**	.31	.01	.59
Gv.Q8 find specific information on a printed page	**.46**	.17	.30	.72
Gr.Q6 rapidly give examples of words that share a semantic feature	**.45**	.21	.20	.63
Ga.Q3 understand what is being said when background noise is present	**.45**	.45	-.07	.63
Gl.Q4 remember information presented visually	**.42**	.29	.13	.59
Gwm.Q2 listen to and repeat a list of words in correct order	**.42**	.10	.38	.67
Gr.Q5 quickly name connected information when presented with a cue	.37	.33	.19	.66
Gc.Q5 engage in conversations with others	-.11	**.99**	-.21	.60
Gc.Q7 communicate effectively what he/she is trying to say	-.01	**.98**	-.20	.70
Gc.Q6 display general knowledge of learnt social rules	-.01	**.92**	-.16	.66
Gs.Q5 work quickly and efficiently on routine tasks	-.20	**.85**	.19	.71
Gs.Q3 respond to simple questions quickly	-.11	**.79**	.11	.62
Gf.Q1 come up with a solution when confronted with a problem	-.10	**.72**	.23	.67
Gwm.Q3 remember two or more instructions in the correct order	.10	**.70**	.14	.78
Gwm.Q4 follows instructions to do two or more tasks	.19	**.68**	.02	.73
Gs.Q7 completes a task in the same amount of time as their peers	.08	**.66**	.17	.75
Gs.Q2 quickly and accurately copy information from the board	.06	**.65**	.15	.66

Table 6.5 continued

Gc.Q8 identify similarities between two common objects (e.g., how are dogs and cats alike?)	.28	**.62**	-.06	.66
Ga.Q7 filter out background noise to listen to teacher's voice	.22	**.56**	.05	.62
Gc.Q3 demonstrate good general knowledge about the world around them	.29	**.55**	-.01	.62
Gs.Q6 scan and quickly determine important information on a page	.25	**.53**	.14	.74
Gr.Q7 quickly recall information related to a learnt topic	.21	**.52**	.20	.74
Gl.Q7 apply effective strategies to learn new information	.42	**.51**	-.06	.70
Gs.Q4 quickly identify differences and similarities between two objects	.05	**.49**	.35	.67
Gwm.Q5 write down dictated sentences accurately, ignoring spelling and punctuation errors	.42	**.49**	-.03	.69
Gf.Q6 use previously acquired knowledge to solve new problems	.04	**.48**	.37	.68
Gl.Q5 retain new information after only a few learning trials	.44	**.48**	-.01	.74
Gl.Q1 listen to and retell the important parts of an orally presented story	.44	**.44**	-.02	.67
Gf.Q2 figure out the number that comes next in a number series	-.06	-.12	**.99**	.74
Gwm.Q6 accurately solve verbally presented maths sums by using mental arithmetic	-.05	.01	**.92**	.79
Gs.Q1 quickly complete a series of simple math sums presented on a worksheet	.05	-.15	**.91**	.71
Gf.Q3 think logically to solve a problem (e.g., could select the right piece in the puzzle below)	-.11	.12	**.83**	.70
Gv.Q2 put parts of an image together to form the whole image	.01	-.11	**.82**	.55
Gf.Q4 sort and classify objects according to a given category	-.11	.16	**.80**	.71
Gv.Q3 differentiate between similarly shaped symbols	.13	-.06	**.80**	.74
Gwm.Q1 remember the numbers given in a verbally spoken math problem for long enough to note it down on paper	.13	-.05	**.79**	.72
Gv.Q1 build a model (e.g., with Lego or blocks) from a picture of a completed model	-.07	-.10	**.75**	.40
Gv.Q4 pick out relevant information in the presence of distracting visual stimuli	.11	-.07	**.74**	.60
Gv.Q5 extract visual information from graphs and handouts	-.10	.24	**.72**	.68
Gv.Q7 accurately judge distances between objects	.01	.06	**.69**	.56
Gv.Q6 interpret information presented in a visual format	-.07	.35	**.58**	.66
Gr.Q1 quickly name objects presented on a page	.35	-.01	**.52**	.68
Gf.Q5 identify relationships between concepts	.05	.33	**.47**	.62
Gr.Q4 rapidly give numerous examples from the same category	.32	.10	**.46**	.59

6.4 Discussion

The aim of this chapter was to report findings of the initial exploration of the underlying structure of the ECCA scale, in the interest of answering the research question of whether the scale would display adequate internal validity. Specifically, it was proposed that an interpretable eight-factor solution would be found based on the eight cognitive ability constructs contained within CHC theory, that being *Gf, Gc, Gwm, Gl, Gr, Gv, Ga,* and *Gs*.

Descriptive statistics of the theorised eight subscales prior to factor extraction revealed highly acceptable levels of internal consistency, with Cronbach alpha values greater than .80 for all subscales. However, it is important to acknowledge a caveat of Cronbach's alpha - although its value is widely reported, it is now regarded as obsolete as a guide for construct homogeneity (the usual connotation of internal consistency), or reliability (Sitjsma, 2009). This indicates that alpha does not convey much information about on the internal structure of a scale, and should not be used as a sole measure of its internal consistency. Future studies could consider reporting Cronbach's alpha in combination with the greatest lower bound (glb; e.g., Woodhouse & Jackson, 1977) to the reliability, which Sitjsma (2009) purports to be an ideal solution to a better reliability estimation practice.

When exploratory factor analysis techniques were applied to the data, a three-factor model was found and confirmed via multiple criteria including a visual inspection of the scree plot, parallel analysis, and interpretability of the solutions. Although all correlations between the three factors were less than .85, providing evidence of the discriminant validity of the subscales, this three-factor model contained items that were originally intended to load on different factors, leading to a completely different model than was hypothesised. Hence, even though the initial dimensionality of the ECCA does display satisfactory internal validity, the hypothesis that an eight-factor solution would be found as theorised *a priori* was not supported by the three-factor solution that was deemed to be most optimal by the data set.

The three factors found in the initial solution provided by EFA could be loosely described as collections of cognitive abilities that relate to previous conceptualisations of intelligence, most notably by Carroll (1993) in his seminal work of a factor-analytic survey of human cognitive abilities. The highest loading items on Factor 2 comprised mostly of items representing verbal ability, and this language-specific factor was thus labelled Gc. As for Factor 3, Carroll (1993) reported that tests measuring *Gv* and *Gf* frequently form a hybrid factor, and the fact that factor analysis does not always successfully differentiate the indicators of these two broad abilities is well-known (McGrew, 2016). Hence, that the current study found Factor 3 (Gf/Gv) to comprise of a combination of mostly items relating to the behavioural operationalisations of *Gf* and *Gv* is not unexpected. Additionally, many tests of fluid reasoning are typically visual in nature (Baron, 2004), highlighting the inter-relatedness of fluid reasoning and visual-spatial processing. For example, the Ravens Progressive Matrices (RPM) is a commonly used measure of relational reasoning, requiring participants to identify patterns on the basis of the spatial organisation of a series of objects (Ferrer, O'Hare, & Bunge, 2009). Finally, the majority of Factor 1 (MISC) comprised of items relating to behavioural operationalisations of several abilities, and thus was a more challenging factor to interpret cleanly as a single broad ability. The primary loading variables on the MISC factor were all related to *Ga*, and many of the other items involved words, or the retrieval of semantic information. Hence, MISC could be interpreted as an auditory-linguistic content facet factor, since some narrow *Ga* abilities such as Phonetic coding, Speech sound discrimination, and Resistance to auditory stimulus distortion are classified as auditory-speech abilities, as they are language-related (Schneider & McGrew, in press).

An item-culling process that is based on pre-determined criteria, such as cross-loading items and items with low communality, is typically undertaken after the initial EFA is carried out in order to improve the structure of the measure. After such undesirable items are eliminated, subsequent EFAs are then run to ascertain simple structure of the scale. However, due to the exploratory nature of this phase of the study, this recommendation was not adhered to. If the item-culling process were to have taken place, it would have resulted in a total of 15 items that would have been removed from the scale, with the number of items representing each broad ability construct being reduced greatly (e.g., the *Gl* factor would have only one remaining item). CHC theory posits the existence of numerous narrow abilities that is contained within each broad ability area, so a reduction in items would have likely resulted in construct underrepresentation, thus compromising on the content validity of the measure. Additionally, the conceptual significance of an item should be examined before its removal from the set, since theoretical knowledge is more relevant than statistical relevance (Beavers et al., 2013). Hence, a content analysis of the cross-loading scale items is detailed in the following section in an attempt to justify plausible theoretical explanations for these occurrences.

The largest proportion of cross-loading items were those that loaded on the MISC and Gc factors. Although the wording of the *Gl* items were adequate in describing skills and behaviours related to the relevant broad ability construct, the included examples meant to aid in item clarity could have introduced connotations of a verbal-related ability, thus resulting in cross-loading. Item Gl.Q1: *listen to and retell the important parts of an orally presented story* reflects the narrow ability of Meaningful Memory (the ability to remember narratives and other forms of semantically related information); however, the ability to articulate a cohesive recall of a story naturally requires some form of verbal ability. Since stories require the listener to understand certain conventions of language and culture, many story memory tests tend to also have a strong loading on *Gc* (Schneider & McGrew, 2012). Item Gl.Q6: *learn and remember the relationship between unrelated objects (e.g., associating the visual symbol 'a' with its corresponding sound /ah/)* primarily reflects the narrow ability of Associative Memory (the ability to remember previously unrelated information as having been paired), but the accompanying example is a behaviour that is symbolic of basic reading skills. Items

Gl.Q5 and Gl.Q7 both have associated behavioural examples that relate to the verbal-related ability of spelling. Item Ga.Q3: *understand what is being said when background noise is present*, could have problematic wording as it was not made clear how the teacher would know the child understood what was being said. Lastly, item Gwm.Q5: *write down dictated sentences accurately, ignoring spelling and punctuation errors* could also have been interpreted by respondents as having implications of a reliance on verbal-related abilities.

In contrast, it is less clear why the *Gc* items cross-loaded. For example, the cognitive tasks as reflected in Gc.Q2: *provide oral definitions for words* and Gc.Q4: *correctly name pictures of familiar objects* are similar to that of the *Comprehension* subtest tasks contained within the WISC (for children 6 years and above), and the WPPSI (for children 6 years and below) scales respectively. These two tasks have been identified to be measures of General Verbal Information, a narrow ability of *Gc* (Flanagan, Ortiz, et al., 2013), and should have had primary loadings on the Gc factor. That Gc.Q4 cross-loaded may be explained by a common finding in intelligence research, wherein fewer factors (simple structure) are often found with younger children versus older children. This may reflect actual differences in the structure of intelligence for younger children, or that it may be more difficult to assess some abilities with younger children (Keith & Reynolds, 2012), as is reflected in the student age group being skewed toward primary-aged children.

The recent segregation of Long-term Storage and Retrieval (*Glr*) into two separate abilities of Learning Efficiency and Retrieval Fluency within CHC theory (Schneider & McGrew, in press) highlights the difference between efficiently storing information into long-term memory (*Gl*), and being able to retrieve that information with ease (*Gr*). The items Gr.Q1, Gr.Q2 and Gr.Q4 all relate to the fluent recall of words, the narrow ability of Naming Facility (the ability to rapidly call objects by their names). For a child to be able to quickly retrieve stored information from long-term memory, some reasonable level of processing speed ability is also required, and research has shown that students with deficits in naming facility are likely to also have weaknesses in processing speed (Catts, Gillispie, Leonard, Kail, & Miller, 2002; Kail & Hall, 1994). Hence, it makes sense for these items to cross-load on both the MISC and Gf/Gv factors. However, it is less clear why item Gf.Q5: *identify relationships between concepts*, a performance-based item based on the narrow ability of Induction (the ability to observe a phenomenon and discover the underlying principles that determine its behaviour) that should have loaded primarily on the PIQ factor, cross-loaded.

Finally, three items loaded on both Gc and Gf/Gv factors. The wording of item Gs.Q4: *quickly identify differences and similarities between two objects*, closely resembles item Gc.Q8: *identify similarities between two common objects (e.g., how are dogs and cats alike),* which had primary loadings on the Gc factor. Since it was not clear what the word "objects" was referring to in Gs.Q4, and this item was presented to respondents after Gc.8, it is possible that respondents also interpreted it as a verbal-related skill. The cross-loading of Item Gf.Q6, *use previously acquired knowledge to solve new problems*, may have been due to ambiguity from the lack of specificity of the word "problem", which could have differing meanings from one context to the next. Lastly, the meaning of item Gv.Q6: *interpret information presented in a visual format (e.g., reading graphs, charts, maps),* could have been misconstrued by respondents to mean the ability to use visual information in a verbal context.

The observation of multiple cross-loading items could also be attributed to the method effect of presenting operationalisations of cognitive tasks in a Likert-type response format in written language. The hierarchical concept of intelligence with *g* at the apex and broad abilities underneath that is modelled in CHC theory (Fiorello et al., 2002) means that a certain degree of commonality is to be expected between the broad ability constructs. However, correlations between variables that are measured with identical methods are known to be distorted (typically inflated), as compared to the actual relationships amongst the constructs themselves, due to the action of common method variance (Lance, Dawson, Birkelbac, & Hoffman, 2010; Spector, 2006). Cognitive abilities measured in the traditional performance-based method are presented in varying formats and methods, largely based on whether they are verbal or nonverbal. For example, the test-taker is required to put blocks together to obtain a subtest score that contributes to the measurement of the non-verbal ability area of *Gv*, and to verbally articulate responses to questions to obtain a score that contributes to the measurement of the verbal ability area of *Gc.* This means that in traditional performance-based measures, nonverbal abilities are typically assessed with methods that do not rely on verbal abilities, and the same can be said for the assessment of verbal abilities. However, the nature of informant-report scales is that all items (verbal or nonverbal) are presented in a written format – this common method of measuring cognitive constructs in the ECCA instrument has thus likely increased commonality more than what would typically be found in performance-based measures, and resulted in some occurrences of cross-loading items.

At first glance, it may seem as though the finding of numerous cross-loading ECCA items indicates inherent problems with the structure of the measure. However, it has been proposed within the personality research domain that instead of being explicit markers of a single dimension, most personality traits reflect a combination of two or more dimensions (Goldberg, 1993). In the same vein, human intelligence comprises of a number of dimensions, all of which are correlated with one another (Conway & Kovacs, 2013). Using short-term and long-term memory as examples of correlated abilities, the differences between the two are distinct enough to be considered as two separate constructs; however, these two memory systems are interdependent, and it is almost impossible to measure (with a single test) one without involving the other (Schneider & McGrew, 2012). Further, Carroll's (1993) extensive review and re-analysis of a wide body of data in the cognitive ability literature revealed a hierarchical arrangement of cognitive abilities, with each level of the hierarchical model typically displaying nonzero positive inter-correlations (Gustafsson & Undheim, 1996). This indicates that the various abilities do not represent completely independent or orthogonal traits – however, they can be reliably distinguished from one another, and therefore represent unique, albeit related, abilities (Keith & Reynolds, 2012). Hence, it is not at all surprising that numerous ECCA items showed loadings on more than one factor, and this exemplifies the construct relatedness of the cognitive ability constructs.

The current study has thus provided some evidence for a factor structure that is consistent with earlier literature, and the degree of cross-loading and lack of discrimination speaks to the argument of *g* being the most relevant and important measure of ability. As evidenced by the cross-loading items and the theoretical complexities associated with the broad range of human cognitive abilities, EFA may not have been the most optimal choice of factor analytical method to determine the structure of a CHC-based informant-report measure. In the past, EFA may have been a powerful and effective method for determining the structural validity of psychological constructs (Bucik, Boben & Kranjc, 1997, as cited in Trninić et al., 2013), yet it has also been criticised for the lack of a universally-accepted basis for choosing among solutions with varying numbers of factors. For example, a five-factor solution depends to a certain extent, on the subjective interpretation of the researcher, which may lead to disagreements about the “true” number of factors (Pervin, Cervone, & John, 2005). Despite the popularity of EFA as a multivariate analytical technique that is used in both theoretical and applied psychological research, theoretical structures of factors that have been identified and established are frequently not supported by empirical research (Trninić et al., 2013), as was the case here. Moreover, EFA methods are susceptible to the magnitude of associations across variables (Wickens, 1994), and this particular element of sample-dependence can give rise to factor analytic solutions that do not align with the actual dimensions of the variables combined (Tractenberg, 2009).

A caveat of using EFA methodologies within the current study is that an initial EFA is not so compelling, in view of the strong theoretical rationale for the hypothesised latent structure of the ECCA sale (that is, its close alignment to CHC theory). Future studies could hence consider utilising exploratory structural equation modelling (eSEM; Marsh et al., 2010), a preferred method for exploratory factor analysis of item-level data. This technique retains the exploratory features of traditional EFA, and the explicit, model-fit information of CFA, thus avoiding the arbitrary and problematic features of EFA as described above.

Thus, it can be seen that the use of exploratory techniques to evaluate the psychometric properties of a newly developed scale can, at times, obtain different types or number of factors than those found in previous research, or that are contained within a known and well-established theory. When the researcher is well-versed in the theory but considers these inconsistent results as definitive, discarding what is known from theory may result in "the destruction of theory" (Bido, Mantovani, & Cohen, 2018). The limitations imposed by using exploratory techniques in scientific research can hence lead to deleterious outcomes, specifically: being unable to compare current results with previous findings, which is crucial for extending current knowledge in the field, and a subsequent disregard for existing theory, so much so that the conceptual meaning of the constructs can be lost in some cases (Bido et al., 2018).

Instead of using EFA, a theoretically-driven measurement model derived from confirmatory factor analysis (CFA) techniques could effectively integrate the information yielded from across the observed variables in the ECCA scale that represent each of the eight cognitive ability areas and, combined with theory, create a single structural model that would not be subject to the mathematically-driven constraints of an a-theoretical model that would be obtained from EFA. Both theoretical and statistical insight should thus be applied to build a measurement model with CFA to combine the variables in hypothesis-driven ways that are aligned with what is already known in theory. Hence, the subsequent CFA as reported in Chapter 7 of this dissertation, provides the best guide to the latent structure of the ECCA scale.

6.5 Conclusion

The findings from this initial exploration of the underlying structure of the ECCA measure have shown that although the expected eight factors corresponding to CHC theory was not found, the three-factor dimensionality of the scale demonstrates a model that is reflective of how human intelligence and intelligence testing was conceptualised 40 to 50 years ago, and is consistent with the statistical methodologies used at that time. There is hence some potential for such an informant-report measure to provide valid estimations about a child's level of cognitive functioning, and it could be used by practitioners to make hypotheses about an individual's profile of cognitive strengths and weaknesses. However, further evidence in support of the structural validity in terms of its alignment with CHC theory is needed. Although EFA is a necessary first step in the process of validating a newly developed scale, its methodological limitations in attempting to replicate a theorised eight-factor model were brought to light in the current study. Hence, the next phase of data analysis was conducted to investigate the factorial structure of the ECCA in accordance with the eight ability areas contained within CHC theory using CFA techniques, the methodology and results of which are presented in the following chapter (Chapter 7).

Chapter 7

Scale Validation with CFA

Following the results found with using exploratory factor analysis (EFA) to investigate the initial validity of the ECCA, this chapter details the second stage of data analysis to address the third research question of this dissertation: the structural validity of the ECCA. The same dataset outlined in the previous chapter was used for this analysis – as highlighted in the previous chapter, although it is not considered best practice to undertake exploratory and confirmatory analyses on the same dataset, the decision was best given the limitations brought about by the current study's sample size. The following section presents the rationale for using confirmatory factor analysis (CFA), and the statistical plan followed for analysing the data. Results are then presented and discussed.

7.1 Confirmatory Factor Analysis

An alternative to the data-driven technique of EFA is CFA, a theory-driven method. That is, the planning of the analysis is determined by the theorised relationships among observed (the items on the instrument) and unobserved (the latent constructs that the instrument purports to measure) variables (Schreiber, Amaury, Stage, Barlow, & King, 2006). Structural equation modelling (SEM), a technique nested within CFA, uses a combination of factor analysis and multiple regression analyses to estimate relationships among hypothesised latent constructs to allow for the evaluation of theoretical propositions of links between constructs.

SEM is typically used to consider two aspects of an instrument: measurement and structure (Costello, 2016). The measurement model is used to test the reliability of the observed variables, and the extent of the interrelationships and covariation (or lack thereof) between items within a single construct; hence, the results from this analysis indicate the measurement properties (i.e., the reliability and validity) of the observed (the scale items) and latent (the hypothesised constructs) variables. At this stage, the factor loadings, unique variables, and modification indexes are estimated to derive the best fit for the measurement models (Schreiber et al., 2006). Subsequently, the structural model is fitted to explore the relationship between the latent variables and across higher order constructs or factors. This final model displays the interrelations among the latent constructs and observable variables as a series of structural equations, similar to running several regression equations.

Applying these principles to the current study, single-factor congeneric modelling at the lower-order factor level allows for an investigation of each CHC broad ability area as measures, and within each broad ability area, the convergent validity of items that make up the construct can be explored. This is a measure of the direct measurement relationship between observed items and latent construct. The most important criterion is that factor loadings should be significantly different from zero, and a threshold of .70 is desirable as it corresponds to approximately 50% of the variance being explained (Costello, 2016; Kline, 2011). For the single-factor congeneric model to be accepted, the indicator variables that contribute toward the overall measurement of the latent variable must be a good reflection of the construct. That is, they must all represent the same typical true score, or be valid measures of the one latent trait. The construct validity of each single-factor congeneric measurement model can be confirmed by inspection of the goodness-of-fit measures and modification indexes.

Structural equation modelling techniques can also be used to test the degree to which the constructs in a model are different (i.e., discriminant validity). Following single-factor congeneric modelling at the lower-order factor level, a structural model can be built to simultaneously specify and estimate the relationships among the latent independent (exogenous; ECCA items) and dependent (endogenous; CHC broad ability constructs and general ability *g*) variables. Since the factor loadings and measurement error variances were fixed in the measurement component of the model, the parameters to be estimated in the structural part of the model are the regression coefficients (Webster & Fisher, 2001). The bidirectional covariances indicate the magnitude (weak or strong) and direction (positive or negative) of the relationships among the latent variables. The overall model fit in this instance is an indication of how robust the structure of the model is, rather than its functionality as a measurement tool. Hence, the fit indices and modification indices will be inspected; however, the former is of secondary importance at this level of the analysis, since the purpose of the structural model is less about fit and more about discriminant validity between the measurement composites, of which modification indices are an indicator.

The aim of the current study is thus to apply SEM techniques to the ECCA scale to evaluate and improve its psychometric properties. The aim is to establish the measurement properties of the instrument at the construct level, and to consider convergent and discriminant validity of the ECCA items. The measurement models to be conducted are the latent variables of the eight broad cognitive ability areas contained within CHC theory, and their indicators (the corresponding ECCA items). The cognitive ability areas under investigation are: Fluid Reasoning (*Gf*), Comprehension-Knowledge (*Gc*), Short-Term Working Memory (*Gwm*), Learning Efficiency (*Gl*), Retrieval Fluency (*Gr*), Visual-Spatial Processing (*Gv*), Auditory Processing (*Ga*), and Processing Speed (*Gs*). The structural model depicting the eight latent constructs, their interrelationships and covariances, will subsequently be specified using derived scores.

7.2 Method

7.2.1 Participants

The sample used in the current study is the same that was used in the previous chapter, comprising of 309 primary and secondary school teachers (refer to section 6.2.1. for a detailed description of the participant sample).

7.2.2 Statistical analysis plan

When there are large numbers of latent variables resulting in an even larger number of observed variables, the number of parameters to be estimated is also large. This limitation means that such models are not robust due to the confounding of measurement and structural parameter estimation problems, which implies that fitting large, complex models is difficult and problematic. This study thus uses the Holmes-Smith and Rowe (1994) approach to fitting complex models, which is a modification of the general two-step method as proposed by Anderson and Gerbing (1988), and incorporates the following steps:

1. Conduct a series of single-factor congeneric measurement models to ensure each is fitting well.
2. Use the factor score regression weights obtained from the measurement models to create a single weighted composite measure for each case on each construct.
3. Calculate the reliability of each composite measure using the Hancock and Mueller (2001) coefficient *H* formula.
4. Calculate the factor loadings in the regressions of each construct on its respective composite measure together with its associated variance.
5. Specify and run full structural equation model using each single composite measure as a single reflective indicator of its associated construct.

7.2.3 Fit indices

A review of research literature was undertaken to identify the most common and appropriate goodness-of-fit and modification indices that would be used to confirm and improve the hypothesised factor models. There are numerous model fit indices available in structural equation modelling, and a comprehensive discussion of all of them is beyond the scope of this dissertation. Given the plethora of fit indices available for researchers to choose from, it can be tempting for researchers to simply choose those that indicate best fit. This should be avoided at all costs, since it is essentially sweeping important information under the carpet (Hooper, Coughlan, & Mullen, 2008). Additionally, going by what is most frequently used is not necessarily good practice, as some statistics are often relied on purely for historical reasons, rather than for their statistical sophistication (Hooper et al., 2008). As per the recommendations made by Hu and Bentler (1999), a combination of absolute and incremental fit indexes will be used in this study to minimise Type I and Type II errors under various conditions.

The chi-square (χ^2) value is the traditional measure for evaluating if the hypothesised model is an exact fit to the data by "assessing the magnitude of discrepancy between the sample and fitted covariance matrices" (Hu & Bentler, 1999, p. 2). As it is a test of the null hypothesis that the hypothesised model is generalisable to the population, a good model fit thus evidenced by a statistically non-significant result at a .05 level (Barrett, 2007). The current study uses Bollen-Stine bootstrapped significance values due to improved accuracy compared to the conventional significance value (Bollen & Stine, 1992). Although the χ^2 statistic is a popular fit statistic, it is sensitive to sample size, which means it is almost always statistically significant for large sample sizes (Bentler & Bonett, 1980), and may not adequately discriminate between good fitting and poor fitting models for small samples (Kenny & McCoach, 2009).

Despite the numerous problems associated with the χ^2, the consensus is that this statistic, along with its degrees of freedom and associated *p*-value, should at all times be reported (Hayduk, Cummings, Boadu, Pazderka-Robinson, & Boulianne, 2007; Hooper et al., 2008; Kline, 2005). However, given its limitation of high sensitivity to sample size, other fit indices were also considered to determine the best fit of the hypothesised factor models, and these are outlined subsequently.

The Root Mean-Square Error of Approximation (RMSEA) is less sensitive to sample size, and takes into account the error of approximation in the population, thus relaxing the stringent requirement on χ^2 that the model holds exactly in the population. The RMSEA has been considered as one of the most informative fit indices (Diamantopoulos & Sigauw, 2000). Interpretation of the RMSEA value was as follows: 0 = exact fit, < .05 = close fit, .05 to .08 = fair fit, .08 to .10 = mediocre fit, and >.10 = poor fit (Moreno, 2003).

The Goodness of Fit index (GFI) is another popular measure of model fit, and shows how closely the hypothesised model comes to replicating the observed covariance matrix (Diamantopoulos & Sigauw, 2000). Its interpretation is similar to that of a correlation coefficient, ranging from 0 (poor fit) to 1 (perfect fit), and is considered satisfactory when the value is more than .90 (Tabachnick & Fidell, 2007). However, the GFI has been shown to be overestimated when there is a large ratio of the degrees of freedom to sample size (Sharma, Mukherjee, Kumar, & Dillon, 2005). Hence, given the often detrimental effect of sample size on the GFI, it is not usually relied upon as a standalone index.

The Standardised Root Mean Square Residual (SRMR) is the square root of the difference between the residuals of the sample covariance matrix and the hypothesised covariance model (Hooper et al., 2008). Values range from 0 to 1.0, and models are considered to display good fit if the SRMR value obtained is than 0.05 (Byrne, 1998; Diamantopoulos & Sigauw, 2000) – as the SRMR value approaches zero, the clearer the indication of ideal fit between the hypothesised and sample models.

The Comparative Fit Index (CFI) is an indication of the amount of improvement in the overall fit of the hypothesised factor models, relative to the most restricted factor model (1-factor; P. Kline, 1998). It is an incremental measure that performs well even when sample size is small (Tabachnick & Fidell, 2007), making it one of the most popularly reported fit index (Fan, Thompson, & Wang, 1999).Values for this statistic range between 0 (poor fit) and 1.0 (perfect fit), with values closer to 1.0 indicating ideal fit. For example, if the CFI value is .80, the relative overall fit of the hypothesised factor model is considered to be 80% better than that of the 1-factor (null model) model estimated with the sample data. A cut-off of more than 0.95 is presently recognised as indicative of good fit (Hu & Bentler, 1999).

Lastly, the Tucker-Lewis Index (TLI) is an incremental fit index that compares the lack of fit between the tested model and the null model. A cut-off of .95 or greater indicates a good fit for the model (Hu & Bentler, 1999). An issue that may arise with the TLI is that due to being non-normed, values may be larger than 1 or slightly below 0.

It can thus be seen that each fit index when used as a standalone has their own limitations; therefore, reporting a combination of fit statistics is usually necessary and considered best practice because different fit indices reflect a different aspect of model fit (Crowley & Fan, 1997; Hooper et al., 2008).

7.2.4 Modification indices and squared multiple correlation

When the fit indices as mentioned above suggest an inadequate fit, the model may be modified or respecified, and subsequently retested (MacCallum, Roznowski, & Necowitz, 1992). Modification indices can be used to detect the exact areas of misfit in a CFA solution, and they reflect an approximation of the reduction in the overall model χ^2 value if the fixed parameter is freely estimated (Whittaker, 2012). Since they are also sensitive to sample size, software programs often provide the values of the expected parameter change for each modification index. These are an estimate of how much the parameter is expected to change in magnitude (a positive or negative direction) if it were to be freely estimated in a subsequent analysis. Although modification indices provide information for how the model fit can be improved in particular areas, any amendments should only be undertaken if they are justifiable on empirical or conceptual grounds (MacCallum et al., 1992). Further model misspecification and introducing redundant parameter estimates due to chance associations in the sample data (i.e., over-fitting) can often result from atheoretical specification searches.

The squared multiple correlation (SMC), indicated by the R^2 value, is the communality estimate for an indicator variable, and measures the proportion of variance in a given indicator variable (scale item) that is explained by its latent variable (factor) – it may be interpreted as the reliability of the indicator variable (Arbuckle, 2013; McCoach & Newton, 2016). Using guidelines as recommended by Moore, Notz, and Flinger (2013), inspection of the R^2 values were carried out to determine if the communality estimate of each scale item was low ($R^2 < .30$), moderate ($.50 > R^2 < .70$), or strong ($R^2 > .70$). If a scale item has low theoretical importance and a low SMC value, it may be targeted for removal in model modification (Hox & Bechger, 1998).

7.2.5 Item keywords

To aid interpretation of item content in the model diagrams, a descriptive keyword was derived for each item instead of using numeric codes. A summary of the items and their respective keywords is provided in table 7.1.

Table 7.1

Keywords of Items on the Estimates of Children's Cognitive Abilities Measure

Item	Keyword/Code	Item	Keyword/Code	Item	Keyword/Code
Gf.Q1	Solution	Gl.Q1	Story	Gv.Q7	Judge
Gf.Q2	Number series	Gl.Q2	TV/movie	Gv.Q8	Find info
Gf.Q3	Logic puzzle	Gl.Q3	Book	Ga.Q1	Decode
Gf.Q4	Classify	Gl.Q4	Visual info	Ga.Q2	Letter-sound
Gf.Q5	Relationships	Gl.Q5	Learning trial	Ga.Q3	Background noise
Gf.Q6	Knowledge	Gl.Q6	Unrelated	Ga.Q4	Speech sounds
Gc.Q1	Syn. and ant.	Gl.Q7	Strategies	Ga.Q5	Rhyming
Gc.Q2	Definition	Gr.Q1	Objects	Ga.Q6	Blend
Gc.Q3	Gen knowledge	Gr.Q2	Letters	Ga.Q7	Filter
Gc.Q4	Familiar objects	Gr.Q3	Words	Ga.Q8	B M E sounds
Gc.Q5	Conversations	Gr.Q4	Category	Ga.Q9	Manipulate
Gc.Q6	Social rules	Gr.Q5	Cue	Gs.Q1	Simple math
Gc.Q7	Communicate	Gr.Q6	Semantic	Gs.Q2	Copy
Gc.Q8	Identify similarities	Gr.Q7	Learnt topic	Gs.Q3	Respond
Gwm.Q1	Remember numbers	Gv.Q1	Model	Gs.Q4	Identify
Gwm.Q2	Repeat words	Gv.Q2	Jigsaw	Gs.Q5	Routine tasks
Gwm.Q3	Remember instructions	Gv.Q3	Symbols	Gs.Q6	Scan
Gwm.Q4	Tasks	Gv.Q4	Distracting	Gs.Q7	Complete tasks
Gwm.Q5	Dictated sentences	Gv.Q5	Extract		
Gwm.Q6	Mental math	Gv.Q6	Interpret		

7.3 Results

7.3.1 Single-factor congeneric modelling

7.3.1.1 Fluid Reasoning (*Gf*)

The initial measurement model for the Fluid Reasoning (Gf) factor contained six items, and demonstrated a less than ideal fit. Inspection of the modification indices suggested significant improvement could be attained by co-varying error terms for items Gf.Q1: *come up with solution when confronted with a problem* and Gf.Q6*: use previous knowledge to solve new problems*, and for items Gf.Q2*: figure out the number that comes next in a number series (e.g., 1, 2, 4, 7, 11, …?)* and Gf.Q3: *think logically to solve a problem (e.g., could select the right piece in the puzzle below*, suggesting the presence of multidimensionality within the construct. While these modifications resulted in an improvement of fit, item Gf.Q1 was removed from the measurement model due to low squared multiple correlations, and the error terms for item Gf.Q5: *identify relationships between concepts* and Gf.Q6 were covaried. These modifications resulted in good fit for the construct of *Gf*. Table 7.2 describes the model fit, and the final model is shown in figure 7 1. The parameter estimates of all items reached values above .70 (ranging from .77 to .91), indicating excellent convergent validity of the retained *Gf* items.

Table 7.2

Fit Indices for Fluid Reasoning

Description	X^2	*df*	Bollen *p*	GFI	RMSEA	SRMR	CFI	TLI
Measurement								
Initial	148.24	9	.005	.851	.224	.051	.907	.845
2	83.64	8	.005	.914	.175	.039	.949	.905
3	42.59	7	.005	.955	.128	.026	.976	.949
4	40.00	4	.005	.949	.171	.029	.970	.925
Final	**14.21**	**3**	**.040**	**.982**	**.110**	**.015**	**.991**	**.969**

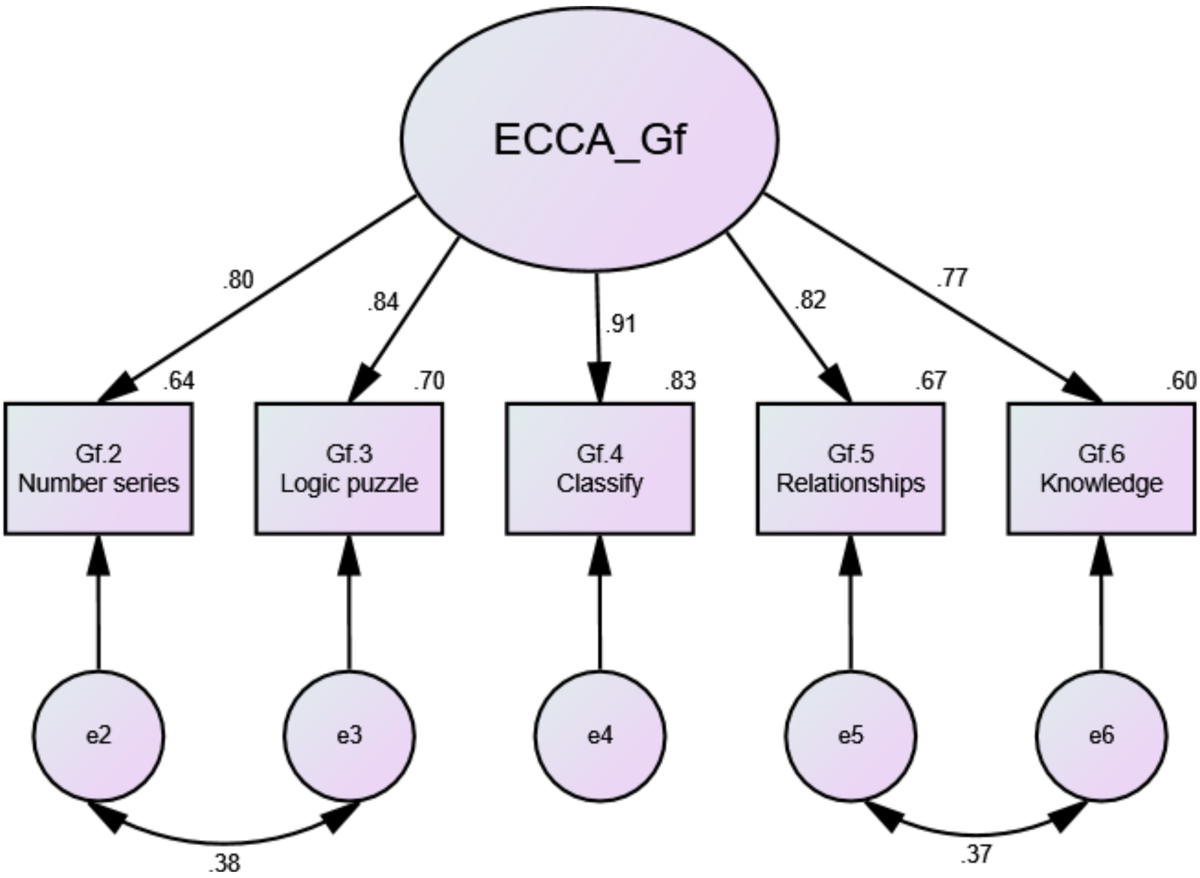


Figure 7.1 Structural diagram with standardised regression pathways for *Fluid Reasoning*

7.3.1.2 Comprehension-Knowledge (*Gc*)

The initial measurement model contained eight items for the *Gc* factor, and did not show an ideal fit. Inspection of the modification indices suggested that significant improvement could be achieved by covarying the error terms for items Gc.Q1: *provide synonyms and antonyms of a word* and Gc.Q2: *provide oral definitions for words*, and for items Gc.5: *engage in conversations with others (peers and adults)* and Gc.Q7: *communicate effectively what he/she is trying to say*. To further improve model fit, items Gc.Q5 and Gc.Q7 was removed from the model due to low squared multiple correlations, and the error terms of Gc.6: *display general knowledge of learnt social rules (e.g., able to answer "what should you do if you find a mobile phone on the ground?")* and item Gc.Q8: *identify similarities between two common objects* were covaried. These modifications resulted in good fit for the construct of *Gc*. Table 7.3 describes the steps taken and their corresponding fit statistics, with figure 7.2 illustrating the final measurement model of the construct of *Gc*. The parameter estimates of all items reached values above .70 (ranging from .73 to .88), indicating excellent convergent validity of the retained *Gc* items.

Table 7.3

Fit Indices for Comprehension-Knowledge

Description	X2	df	Bollen p	GFI	RMSEA	SRMR	CFI	TLI
Measurement								
Initial	256.35	20	.005	.806	.196	.056	.886	.840
2	175.34	19	.005	.863	.163	.050	.924	.888
3	131.10	18	.005	.901	.143	.043	.945	.915
4	86.18	13	.005	.920	.135	.037	.932	.958
5	61.59	12	.005	.948	.116	.030	.971	.950
6	51.58	8	.005	.951	.133	.028	.970	.944
Final	25.59	7	.060	.974	.093	.017	.987	.973

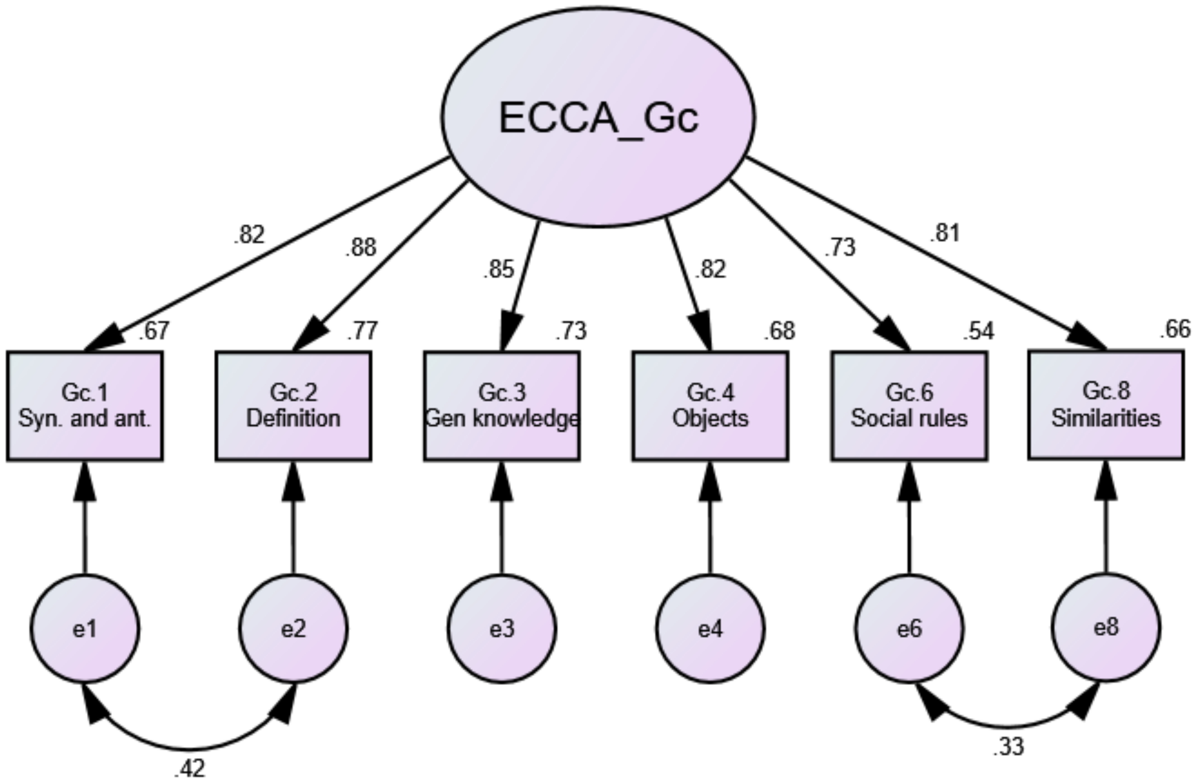


Figure 7.2 Structural diagram with standardised regression pathways for *Comprehension-Knowledge*

7.3.1.3 Short-Term Working Memory (Gwm)

The initial measurement model for the Gwm factor contained six items, and resulted in a less than satisfactory fit. Inspection of the modification indices suggested significant improvement could be achieved by covarying the error terms for items Gwm.Q1*: remember the numbers given in a verbally spoken math problem for long enough to note it down on paper* and Gwm.Q6: *accurately solve verbally presented math sums by using mental math (i.e., no pencil or paper)*, and for items Gwm.Q1 and Gwm.Q2: *listen to and repeat a list of words in the correct order*, suggesting the presence of multidimensionality within the construct. While these modifications resulted in an improvement of fit, item Gwm.Q6 was ultimately removed from the measurement model due to low squared multiple correlations (SMC; R^2 = .47) to attain good model fit. While item Gwm.Q1 also displayed less than ideal SMC (R^2 = .51), its removal did not make a significant improvement in model fit, thus it was retained in the final *Gwm* model. Table 7.4 describes the steps taken for the model fit, and figure 7.3 illustrates the final measurement model of *Gwm*. The parameter estimates of all items ranged from .68 to .95, indicating acceptable convergent validity of the retained *Gwm* items.

Table 7.4

Fit Indices for Short-Term Working Memory

Description	X2	df	Bollen p	GFI	RMSEA	SRMR	CFI	TLI
Measurement								
Initial	270.75	9	.005	.771	.307	.076	.842	.737
2	41.43	7	.010	.957	.126	.035	.979	.955
Final	16.19	4	.129	.980	.099	.020	.990	.976

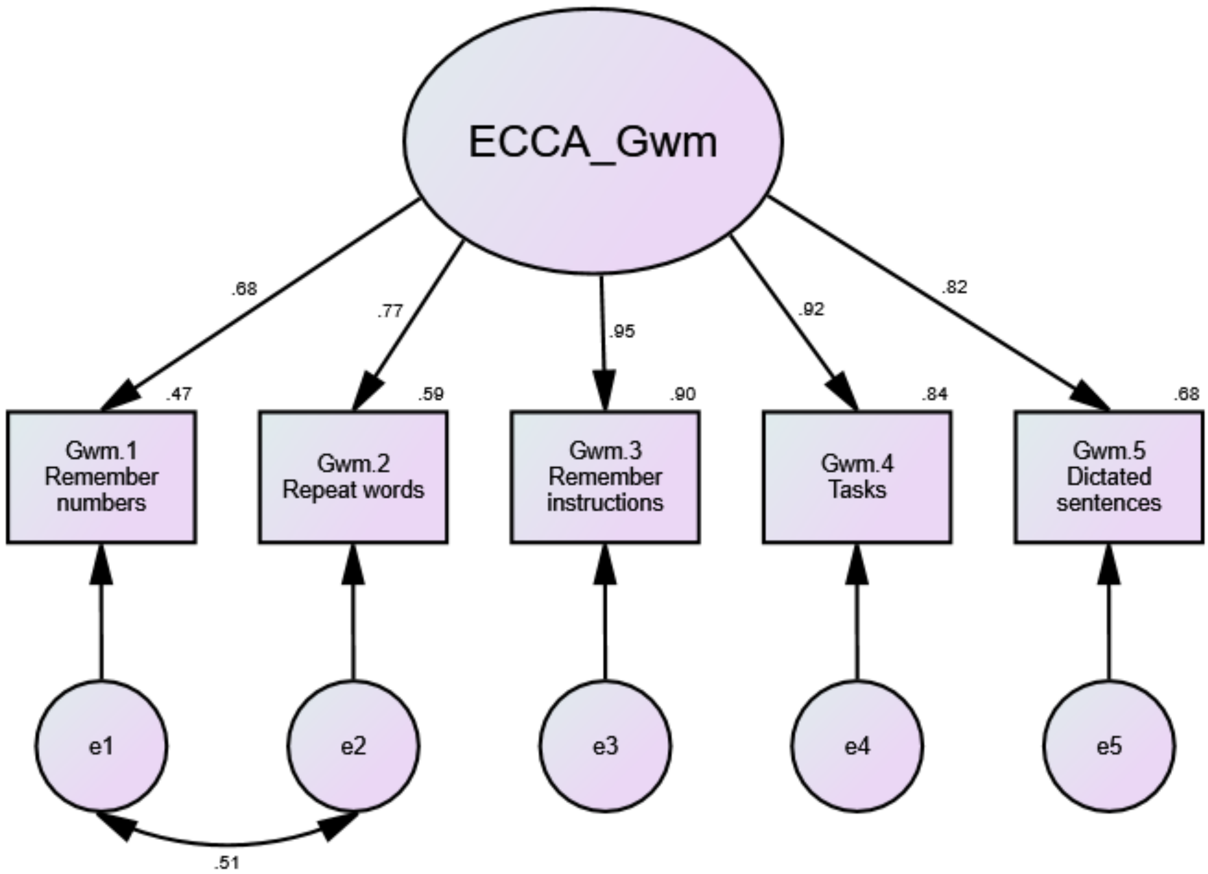


Figure 7.3 Structural diagram with standardised regression pathways for *Short-Term Working Memory*

7.3.1.4 Learning Efficiency

The *Gl* factor comprised of seven items for its initial measurement model, and did not demonstrate an ideal model fit. Examination of the modification indices indicated that significant improvement could be achieved by covarying the error terms for items Gl.Q5: *retain information after only a few learning trials (e.g., learning a spelling list)*, Gl.Q6: *learn and remember the relationship between unrelated objects* and Gl.Q7: *apply effective strategies to learn new information*, signifying the presence of multidimensionality within the construct. Although this led to an improvement in model fit, Gl.Q6 was removed from the model due to low squared multiple correlations (R^2 = .581). These modifications resulted in good fit for the *Gl* construct. Table 7.5 describes the steps taken for the model fit, and figure 7.4 depicts the final measurement model of *Gl*. The parameter estimates of all items reached values above .70 (ranging from .78 to .92), indicating excellent convergent validity of the retained *Gl* items.

Table 7.5

Fit Indices for Learning Efficiency

Description	X2	df	Bollen p	GFI	RMSEA	SRMR	CFI	TLI
Measurement								
Initial	105.2	14	.005	.899	.145	.035	.954	.930
2	71.2	13	.020	.938	.121	.031	.970	.952
3	51.2	12	.070	.957	.103	.025	.980	.965
4	23.1	11	.557	.980	.060	.013	.994	.988
Final	15.0	8	.587	.984	.053	.013	.996	.992

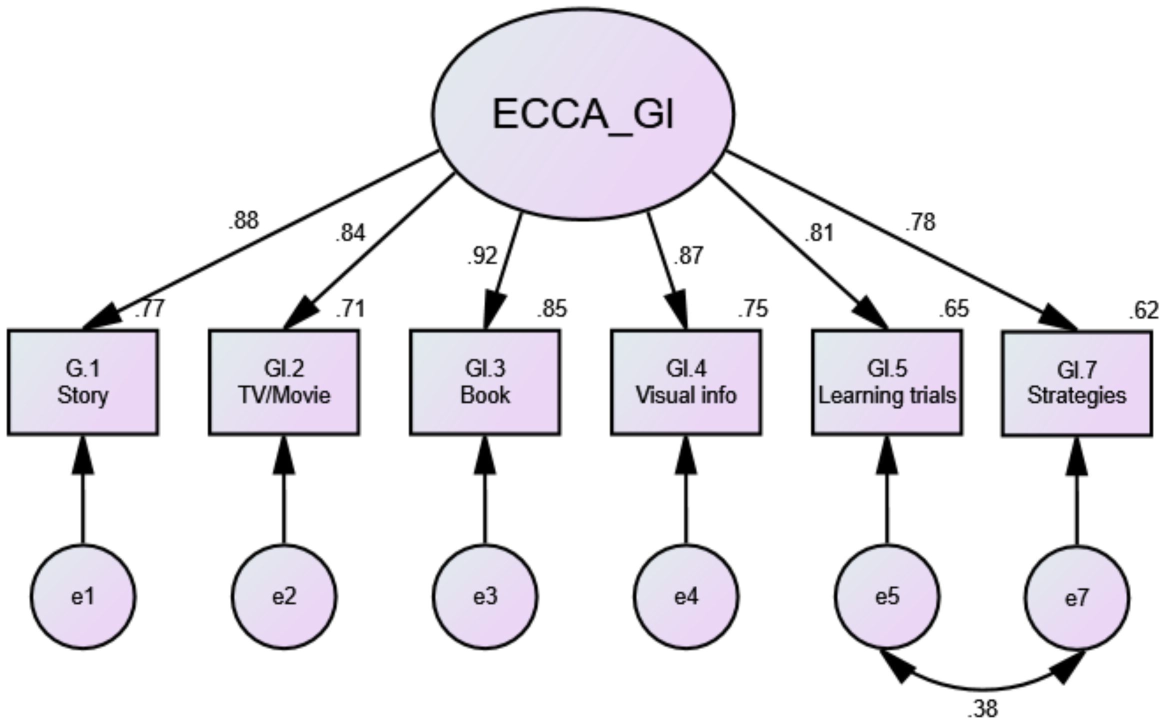


Figure 7.4 Structural diagram with standardised regression pathways for *Learning Efficiency*

7.3.1.5 Retrieval Fluency

The *Gr* factor did not demonstrate an ideal fit with its initial measurement model containing seven items. Inspection of the modification indices suggested that significant improvement could be achieved by covarying the error terms for items Gr.Q2: *quickly name letters presented on a page* and Gr.Q3: *quickly name whole words presented on a page*. Further review of the modification indices for subsequent steps in the model fit process indicated that the error terms for item Gr.Q1: *quickly name objects presented on a page* should also be covaried with Gr.Q2 and Gr.Q3 to achieve better fit. Such covariation within error terms suggests multidimensionality within the *Gr* construct. To achieve the best fit for the *Gr* construct, item Gr.Q2 was deleted from the model due to its low squared multiple correlations, and error terms for items GrQ3 and GrQ7: *quickly recall information related to a learnt topic* were covaried. Table 7.6 details the steps taken to ensure good model fit, and the final measurement model is depicted in figure 7.5. The parameter estimates of all items reached values above .70 (ranging from .79 to .91), indicating excellent convergent validity of the retained *Gr* items.

Table 7.6

Fit Indices for Retrieval Fluency

Description	X2	df	Bollen p	GFI	RMSEA	SRMR	CFI	TLI
Measurement								
Initial	259.3	14	.005	.762	.239	.053	.887	.830
2	159.2	13	.005	.868	.191	.044	.933	.891
3	110.1	12	.005	.907	.163	.036	.955	.921
4	39.8	11	.124	.966	.092	.019	.987	.975
5	34.2	8	.124	.966	.103	.020	.984	.970
Final	23.09	7	.294	.977	.086	.016	.990	.979

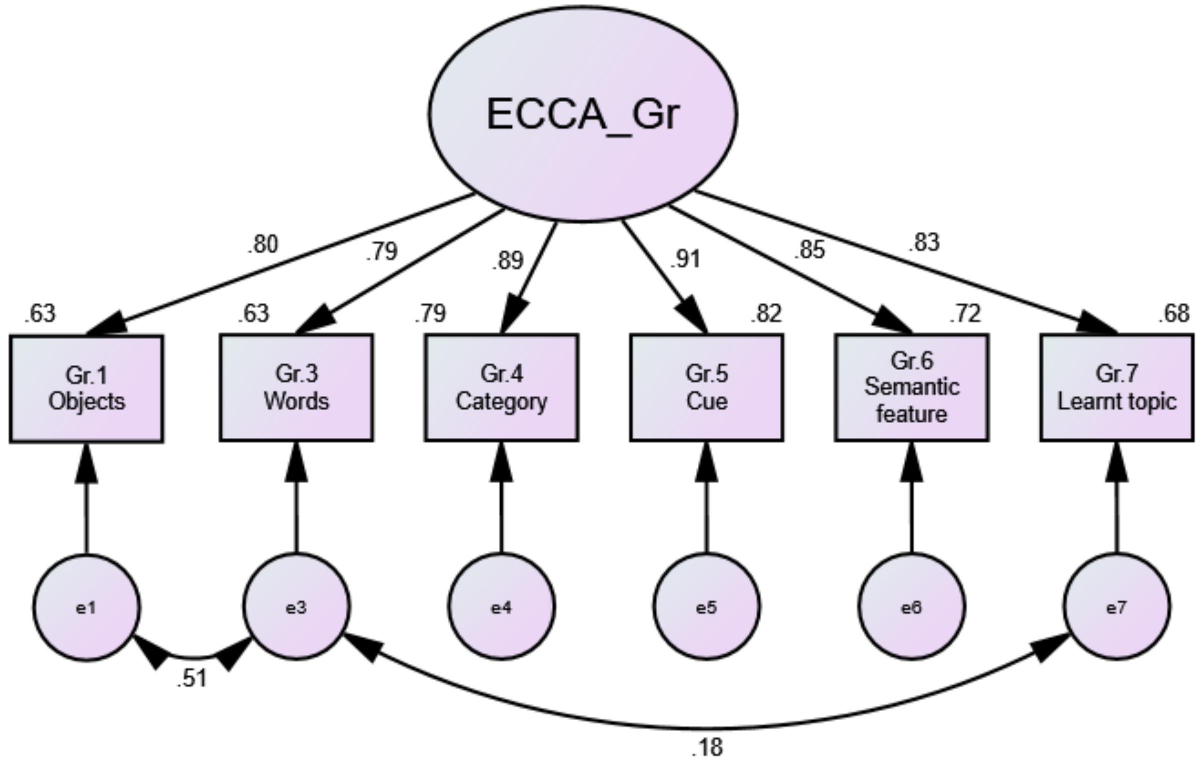


Figure 7.5 Structural diagram with standardised regression pathways for *Retrieval Fluency*

7.3.1.6 Visual-Spatial Processing (Gv)

The initial measurement model for the *Gv* factor contained eight items, and did not demonstrate an ideal fit. Examination of the modification indices indicated significant improvement of model fit would be achieved if the error terms for items Gv.Q1: *build a model from a picture of the completed model* and Gv.Q2: *put parts of an image together to form the whole image*, and for items Gv.Q5: *extract visual information from graphs and handouts* and Gv.Q6: *interpret information presented in a visual format (e.g. reading graphs, charts, maps, etc.)* were covaried, indicating the presence of multidimensionality within the *Gv* construct. Although these steps led to some improvement in model fit, deleting item Gv.Q1 from the model due to low squared multiple correlations and covarying the error terms for Gv.Q6 and Gv.Q8 *find specific information on a printed page or computer screen* resulted in the best fit for the construct of *Gv*. Even though covarying error terms with more than one item is usually not recommended practice, the deletion of item Gv.Q6 resulted in a poorer model fit (as seen in Step 5 in table 7.3.6). Hence, it was retained in the final model. Table 7.7 details the model fit with the final measurement model displayed in figure 7.6. The parameter estimates of all items reached values above .70 (ranging from .76 to .87), indicating excellent convergent validity of the retained *Gv* items.

Table 7.7

Fit Indices for Visual-Spatial Processing

Description	X2	df	Bollen p	GFI	RMSEA	SRMR	CFI	TLI
Measurement								
Initial	312.3	20	.005	.793	.218	.066	.848	.788
2	61.88	18	.055	.952	.089	.034	.977	.965
3	42.63	13	.070	.961	.086	.027	.982	.970
5	24.49	9	.189	.974	.075	.024	.987	.978
Final (4)	27.44	12	.303	.976	.065	.022	.990	.983

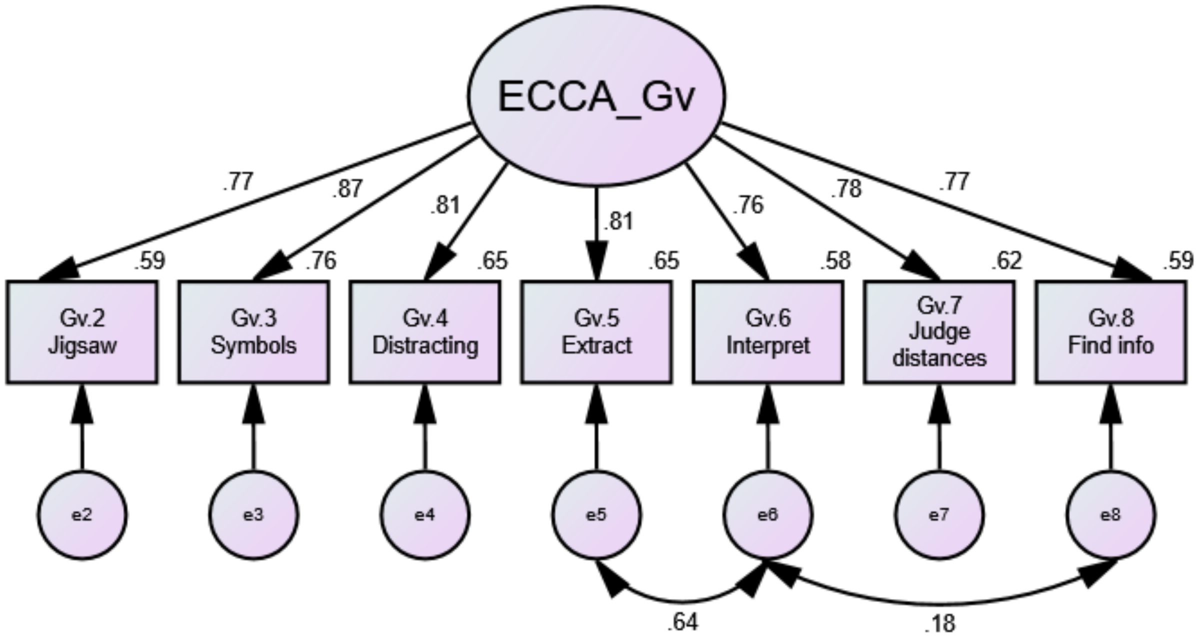


Figure 7.6 Structural diagram with standardised regression pathways for *Visual-Spatial Processing*

7.3.1.7 Auditory Processing (Ga)

The initial measurement model for the construct of *Ga* included 9 items, but demonstrated a less than ideal fit. Items Ga.Q3: *understand what is being said when background noise is present* and Ga.Q7: *filter out background noise to listen to my voice in the classroom* were deleted due to low squared multiple correlations (R^2 = .59 and .52 respectively) and inspection of the modification indices suggested that significant improvement in model fit would be achieved if the error terms for items Ga.Q8: *identify the beginning, middle, and ending sounds in words* and Ga.Q9: *manipulate sounds within words to form new words*, and Ga.Q1: *sound out unfamiliar words* and Ga.Q2: *know which sounds go with which letters* were covaried. The steps taken and corresponding fit statistics are displayed in table 7.8 with the final measurement model illustrated in figure 7.7. The parameter estimates of all items reached values above .70 (ranging from .88 to .94), indicating excellent convergent validity of the retained *Ga* items.

Table 7.8

Fit Indices for Auditory Processing

Description	X2	df	Bollen p	GFI	RMSEA	SRMR	CFI	TLI
Measurement								
Initial	294.97	27	.005	.827	.180	.044	.926	.902
2	164.6	20	.005	.879	.153	.021	.956	.939
3	83.3	19	.020	.929	.105	.017	.981	.971
4	45.9	18	.343	.964	.071	.012	.992	.987
Final	39.49	12	.289	.966	.086	.011	.991	.984

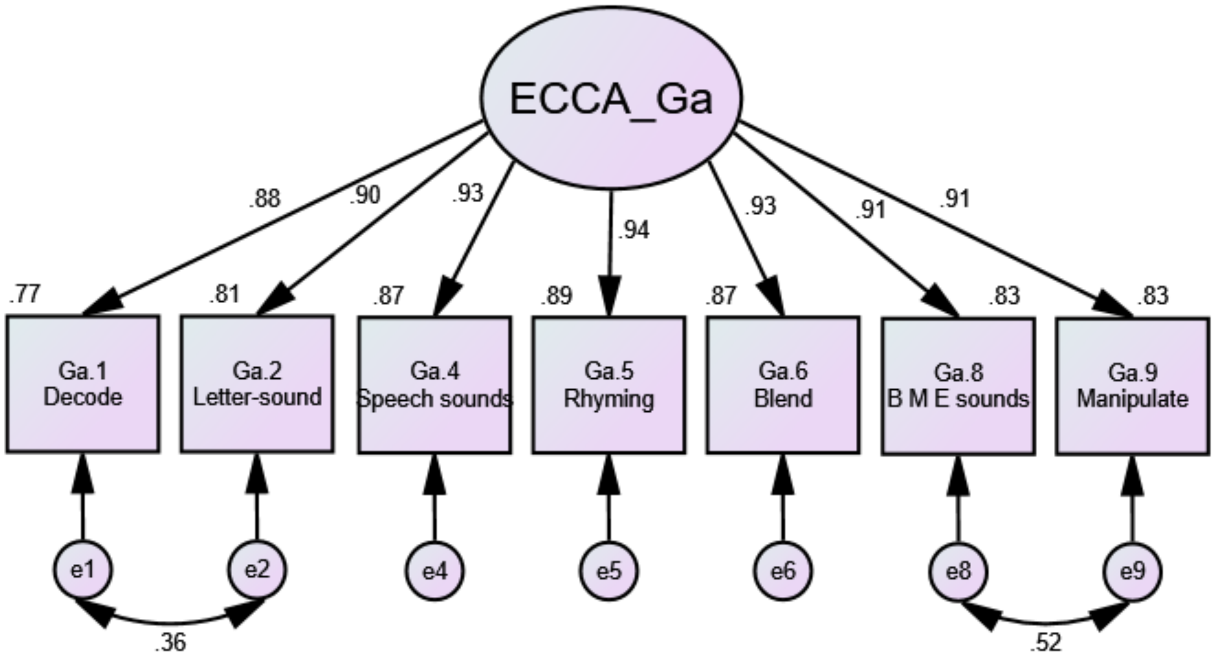


Figure 7.7 Structural diagram with standardised regression pathways for *Auditory Processing*

7.3.1.8 Processing Speed (Gs)

The initial measurement model for the *Gs* factor contained seven items, and demonstrated a less than ideal fit. A low squared multiple correlation (R^2 = .41) indicated that Gs.Q1: *quickly complete a series of simple math sums presented on a worksheet* did not fit well with the model and was thus deleted, and inspection of the modification indices suggested that significant improvement of model fit would occur if the error terms of item Gs.Q3: *respond to simple questions quickly*) were covaried with error terms of items Gs.Q4: *quickly identify differences and similarities between two objects* and Gs.Q5: *work quickly and efficiently on routine tasks*. In the next steps, Gs.Q3 was removed from the model due to low squared multiple correlations, and the error terms of Gs.Q4 and Gs.Q5 were covaried, which led to good model fit for the construct of *Gs*. Table 7.9 describes the steps taken for model fit, with the final measurement model displayed in figure 7.8. The parameter estimates of all items reached values above .70 (ranging from .75 to .96), indicating excellent convergent validity of the retained *Gs* items.

Table 7.9

Fit Indices for Processing Speed

Description	X2	df	Bollen p	GFI	RMSEA	SRMR	CFI	TLI
Measurement								
Initial	132.07	14	.005	.883	.165	.038	.938	.907
v2	68.3	8	.005	.928	.156	.032	.965	.935
v3	45.05	7	.005	.953	.133	.025	.978	.953
v4	40.9	5	.005	.949	.153	.025	.975	.949
Final	22.3	4	.020	.972	.122	.018	.987	.968

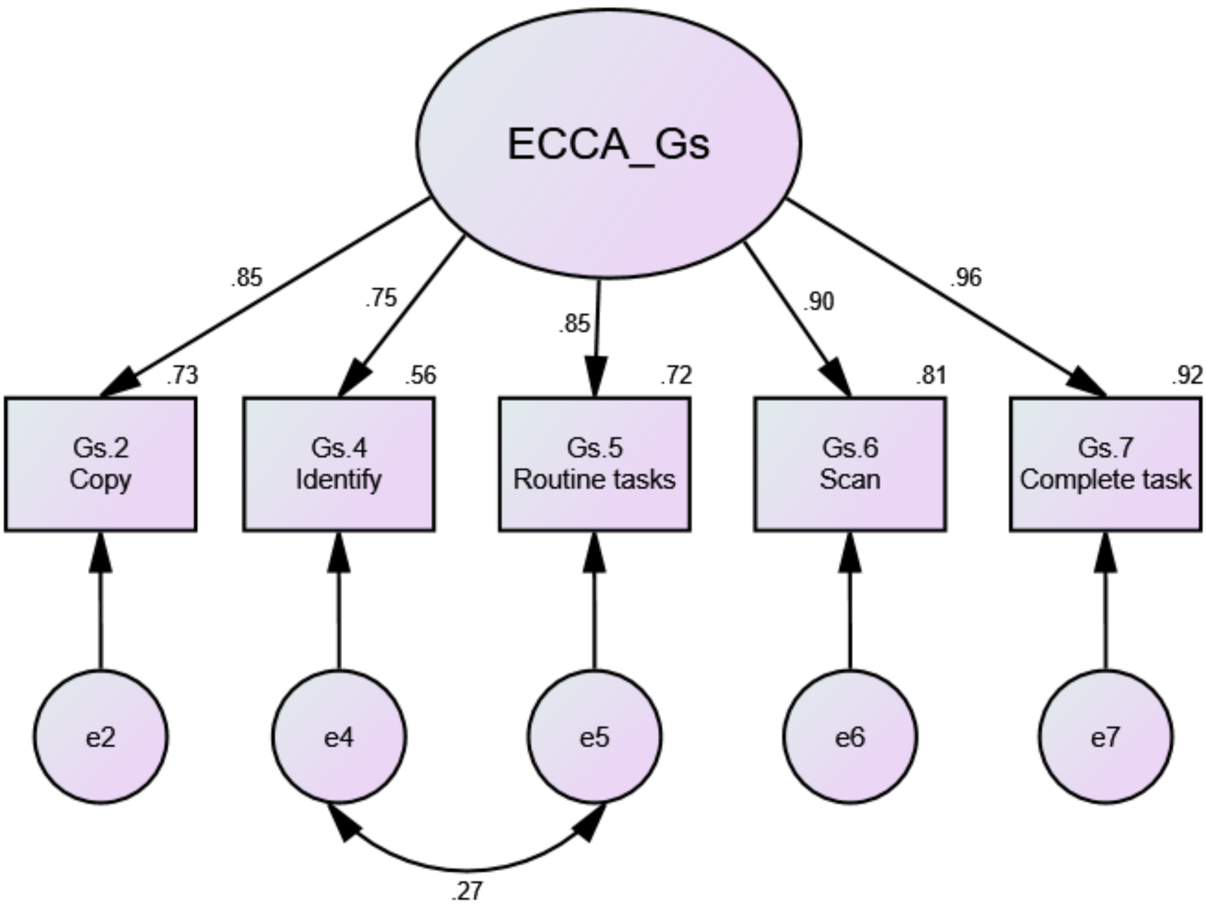


Figure 7.8 Structural diagram with standardised regression pathways for *Processing Speed*

7.3.2 Derived factor model for ECCA measure

Composite values for each ECCA construct were derived using factor score weights. Following the guidelines outlined by Holmes-Smith and Rowe (1994), the standard deviation and reliability of each facet was then used to estimate the factor loading and error variance for each derived facet. Instead of using Cronbach's alpha for reliabilities, which assumes tau equivalence and is an underestimate of the reliability for congeneric measures (Graham, 2006), coefficient *H* reliability was used as it is a better estimate of the composite reliability if a congeneric measurement model has been fitted (Hancock & Mueller, 2001). By fixing the regression coefficients and measurement error variances as obtained in the construction of the measurement models, a significant reduction in inconsequential error was achieved (Munck, 1979). Subsequently, the complete CHC model was specified, and the structural diagram of the ECCA instrument is illustrated in figure 7.3.2.1. This final version of the ECCA contained a total of 47 items that were retained in the full structural equation model.

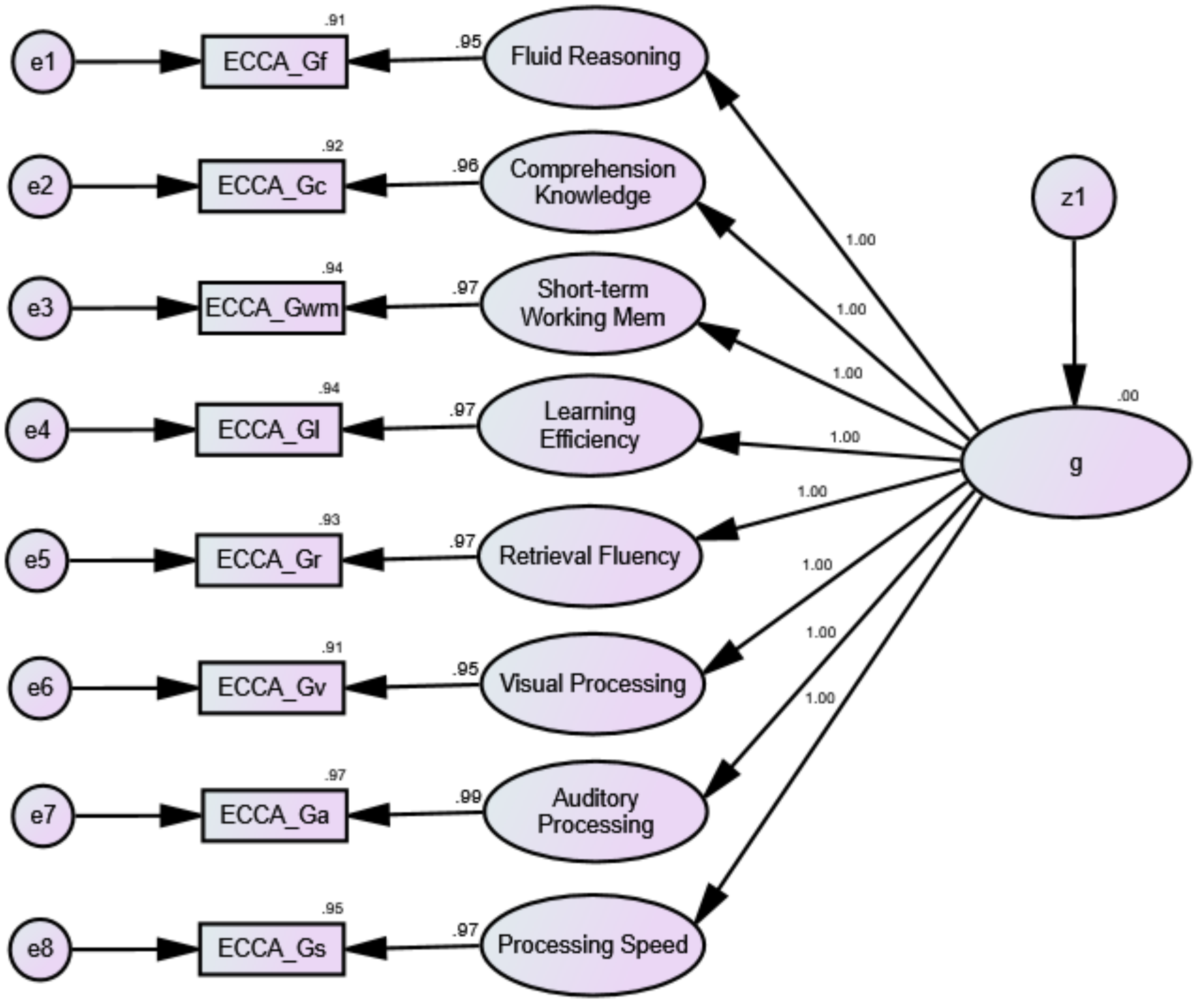


Figure 7.9 Structural diagram for *Estimates of Children's Cognitive Abilities* measure using derived facet scores

The purpose of this exercise was to examine evidence of discriminant validity across the broad ability factors, given that convergent validity was already demonstrated within factors previously as part of conducting the measurement models. A sample correlation matrix was run in SPSS based on the derived factor scores, with inter-correlations between the constructs ranging from r = .67 to .84. Four out of the 31 additional relationships possible between the factors and latent constructs demonstrated modification index scores higher than 10, warranting further investigation. Although the overall structural model fit is of secondary importance, when the recommended modifications were implemented (i.e., the covarying of error terms between the measurement models), the resultant structural model (χ^2 = 4565.17, *df* = 20, *p* < .001, SRMR = .15, RMSEA = .86) was not significantly improved on the initial structural model (χ^2 = 4740.82, *df* = 28, *p* < .001, SRMR = .15, RMSEA = .74). The initial structural model was thus retained without modifications, but a further discussion regarding their meaning is provided in the subsequent section of this chapter. The derived factor correlation matrix is provided in table 7.10, and a summary of the suggested modification indices is provided in table 7.11. Lastly, the descriptive statistics of the ECCA instrument after structural equation modelling is provided in table 7.12.

Table 7.10

Derived Factor Correlation Matrix

	ECCA_Gf	ECCA_Gc	ECCA_Gwm	ECCA_Gl	ECCA_Gr	ECCA_Gv	ECCA_Ga	ECCA_Gs
ECCA_Gf	1.00							
ECCA_Gc	.73	1.00						
ECCA_Gwm	.78	.79	1.00					
ECCA_Gl	.71	.83	.81	1.00				
ECCA_Gr	.78	.80	.79	.81	1.00			
ECCA_Gv	.81	.70	.78	.72	.79	1.00		
ECCA_Ga	.67	.81	.75	.80	.78	.71	1.00	
ECCA_Gs	.75	.75	.84	.77	.78	.72	.73	1.00

Table 7.11

Modification Indices for Estimates of Children's Cognitive Abilities Structural Model

Factor	Modification Index	Resulting SR change
Gwm → Gs	15.54	0.06
Gf → Gv	55.47	0.26
Gs → Gwm	13.24	0.05
Gv → Gf	54.19	0.07

Table 7.12

Descriptive Statistics and Reliability for Estimates of Children's Cognitive Abilities Measure

Factor/Item	Mean	SD	Skew[a]	Kurtosis[b]	Reliability *H*
Fluid Reasoning					
ECCA_Gf	1.79	0.58	.249	-.715	.926
Gf.Q2	3.53	1.46	.047	-.847	
Gf.Q3	3.42	1.42	.117	-.803	
Gf.Q4	3.83	1.26	-.330	-.268	
Gf.Q5	3.75	1.29	-.137	-.519	
Gf.Q6	3.68	1.23	.124	-.795	
Comprehension-Knowledge					
ECCA_Gc	4.03	1.06	.222	-.412	.930
Gc.Q1	3.58	1.90	.061	-.592	
Gc.Q2	3.59	1.22	.206	-.486	
Gc.Q3	3.82	1.24	-.037	-.719	
Gc.Q4	4.28	1.03	-.345	.871	
Gc.Q6	4.05	1.13	-.062	-.232	
Gc.Q8	3.98	1.16	-.144	-.093	
Short-term Working Memory					
ECCA_Gwm	3.62	1.11	.453	-.551	.948
Gwm.Q1	3.30	1.38	.117	-.810	
Gwm.Q2	3.31	1.26	.166	-.482	
Gwm.Q3	3.70	1.17	.306	-.570	
Gwm.Q4	3.74	1.18	.260	-.556	
Gwm.Q5	3.34	1.27	.449	-.631	
Learning Efficiency					
ECCA_Gl	3.71	1.10	.162	-.312	.948
Gl.Q1	3.67	1.20	.040	-.481	
Gl.Q2	3.78	1.27	-.367	-.236	
Gl.Q3	3.72	1.25	.006	-.502	
Gl.Q4	3.96	1.14	-.317	.220	
Gl.Q5	3.48	1.28	.439	-.648	
Gl.Q7	3.27	1.23	.540	-.345	

Table 7.12 continued

Retrieval Fluency					
ECCA_Gr	3.62	1.12	.368	-.607	.943
Gr.Q1	3.90	1.14	-.367	.147	
Gr.Q3	3.72	1.32	-.024	-.664	
Gr.Q4	3.62	1.31	-.005	-.564	
Gr.Q5	3.60	1.25	.071	-.535	
Gr.Q6	3.43	1.32	.115	-.602	
Gr.Q7	3.60	1.18	.301	-.617	
Visual-Spatial Processing					
ECCA_Gv	3.59	1.11	.090	-.331	.927
Gv.Q2	3.63	1.36	-.470	-.381	
Gv.Q3	3.74	1.36	-.302	-.353	
Gv.Q4	3.51	1.30	-.249	-.248	
Gv.Q5	3.50	1.27	-.049	-.497	
Gv.Q6	3.56	1.25	.026	-.516	
Gv.Q7	3.60	1.30	-.273	-.428	
Gv.Q8	3.49	1.25	.106	-.654	
Auditory Processing					
ECCA_Ga	3.73	1.18	.102	-.383	.975
Ga.Q1	3.58	1.31	.159	-.719	
Ga.Q2	3.82	1.27	-.133	-.458	
Ga.Q4	3.78	1.24	-.108	-.380	
Ga.Q5	3.79	1.20	-.086	-.254	
Ga.Q6	3.70	1.29	-.007	-.581	
Ga.Q8	3.72	1.31	-.065	-.543	
Ga.Q9	3.54	1.36	.073	-.723	
Processing Speed					
ECCA_Gs	3.32	1.24	.766	-.573	.957
Gs.Q2	3.32	1.28	.554	-.563	
Gs.Q4	3.85	1.18	.045	-.260	
Gs.Q5	3.57	1.26	.368	-.654	
Gs.Q6	3.18	1.33	.753	-.487	
Gs.Q7	3.25	1.39	.697	-.760	

7.4 Discussion

The aim of the current study was to evaluate the psychometric properties of the ECCA scale, specifically, its structural and construct validity in terms of the convergent and discriminant validity of the latent constructs. This was performed using a two-step structural equation modelling technique as recommended by Holmes-Smith and Rowe (1994) in dealing with complex models with a large number of variables. First, the fundamental measurement model at the scale item level was specified and fitted to determine the convergent validity of the items. Secondly, the complete instrument was structurally modelled using derived composite scores to determine the discriminant validity of the broad ability constructs.

The current study utilised 58 indicator variables to measure eight latent variables. Fitting single-factor congeneric measurement models is advantageous when working with many variables for numerous reasons. Aside from the obvious benefit of reducing large data sets to a more manageable level, the technique also accounts for differences in the extent to which each individual measure contributes to the overall composite scale, thus providing a more accurate representation of the data (Holmes-Smith, 2014). Additionally, testing the robustness of the measurement model before conducting the structural model makes sense – if the chosen indicators for the construct do not adequately measure that construct, modifications should be made before the structural relationships between constructs are tested (Teo, Tsai, & Yang, 2013). In this way, the researcher can identify sources of poor fit of a full structural model and decide if this poor fit is due to the individual measurement models or the overall structural model (Webster & Fisher, 2001).

The results of the single-factor congeneric modelling of the constructs were very favourable in indicating the convergent validity of items within each construct of the newly developed ECCA instrument, with the fit statistics of each measurement model providing supporting evidence of the construct validity of the indicator variables used to measure their respective latent construct. Following the systematic removal of undesirable items with low squared multiple correlations, or covarying the error terms of items where indicated by the modification indices in accordance with theoretically sensible conditions, ideal model fit was displayed for all factors, indicating the presence of construct validity for each broad cognitive ability area. Given that all factors met the requirements for good model fit, no further discussion was warranted about measurement model fit. Moreover, excellent convergent validity of each broad ability area was evident as indicated by almost all ECCA items achieving factor loadings of above .70 onto their respective factors. The exception to this was item Gwm.1: *remember the numbers given in a verbally spoken math problem for long enough to note it down on paper,* which achieved a parameter estimate of .68, nonetheless indicating very good convergent validity.

There is a need to elaborate on the choice to covary error terms of items rather than simply deleting these "problematic" items with low SMC values. There is much debate surrounding the practice of correlating measurement errors in order to improve model fit, with many authors cautioning against it (e.g., Brannick, 1995; Cortina, 2002; Gerbing & Anderson, 1984; Hermida, 2015; Kaplan, 1990). This exercise is often considered suspect as it enables researchers to achieve ideal model fit in spite of excluding relevant variables from their models based on post hoc modifications (Cortina, 2002), which may result in an improvement of model fit, but not of our understanding of the phenomenon in question. Common advice is to confine correlating residuals to only those that are theoretically defensible (MacCallum, 1986; Silvia & MacCallum, 1988); however, Cole, Ciesla, and Steiger (2007) demonstrated that this practice often leads to the under-inclusion of covaried error terms that are completely justified on the basis of measurement theory and research design. The basis for allowing for a correlation between measurement errors can therefore be data-driven (e.g., as suggested by modification indices) or theoretically-motivated. Covarying error terms of variables can be empirically and conceptually meaningful in certain circumstances, such as compelling evidence of underlying multidimensionality within the construct, or the existence of local dependence between items (one "causes" another in some way; Costello, 2016). The failure to include such covaried residuals can thus alter the meaning of the extracted latent variable, and generate potentially meaningless results (Cole et al., 2007).

In the current study, all eight factors utilised covaried error terms in some form in order to achieve improved model fit, with local dependence between items and underlying multidimensionality attributed to be the main cause for doing so. An example of local dependence between items is reflected in items Gr.1 and Gr.3: *quickly name objects/whole words presented on a page*, given that the natural trajectory of development for children's learning indicates an acquisition in the skill of object-naming preceding that of word-naming (Landau, Smith, & Jones, 1998). This suggests that if the child is unable to successfully perform the task described in Gr.1 (name objects), he or she is likely unable to perform the task described in Gr.3 (name whole words).

Additionally, the conceptualisation of human cognitive ability as a multidimensional construct is well-established (Schneider & Newman, 2015); therefore, it is not unexpected for this phenomenon to also exist in operationalisations through scale items that attempt to measure it. For example, the *Gc* subscale demonstrated good model fit using six out of the initial eight items; however, the domain of Comprehension-Knowledge includes several specific narrow abilities such as General Verbal Information (K0), Language Development (LD) and Lexical Knowledge (VL), highlighting the multifaceted nature of the broad ability *Gc*. However, the presence of multidimensionality within each narrow ability construct itself is also evident, such that the inspection of modification indices suggesting that the error terms of Gc.1: *provide synonyms and antonyms of a word* and Gc.2: *provide oral definitions for words* should be covaried to improve model fit is not unexpected. Research has shown that the development of one's vocabulary knowledge is a complex process, with various types of knowledge providing related but distinct contributions to the richness of a semantic and conceptual system of lexical knowledge (Neugebauer, Kieffer, & Howard, 2015; Scott, 2015). Hence, rather than risking construct underrepresentation, covaried error terms were retained. Finally, although relying solely on undesirable SMC values for item removal would have further improved the model fit, this may have been too excessive, resulting in a limited number of items representing each ability construct. Hence, a conservative approach to item removal was undertaken for the current study, and resulted in a total of 47 items that were retained on the ECCA scale for construction of the structural model.

The final model was explored using derived composite scores that were built on the parameters established during the single-factor congeneric modelling analyses. Overall, the ECCA instrument largely demonstrated adequate discriminant validity across factors, with correlations between derived factor scores ranging from r = .67 to .84. Given what is known about the positive inter-correlations between cognitive abilities as measured traditional performance-based methods attributable to the "positive manifold" (Spearman, 1904) phenomenon, that relationships of this magnitude were found between the ability constructs as measured by the ECCA instrument was not surprising. Moreover, a mean correlation of r = .77 with variance explained of 58.74% suggests that within this model, overall *g* would account for approximately 60% of the variance in the broad abilities. Although this is reflective of convenience sampling characteristics and is likely to be an inflation of normative ones, it is consistent with what previous research has shown about *g* typically accounting for 40-50% of the variance explained in performance-based IQ test batteries (Mackintosh, 2011), as well as what CHC theory posits about the hierarchical structure of human abilities, wherein the broad abilities at Stratum II are subsumed by *g* at Stratum III (Flanagan, Ortiz, et al., 2013; Schneider & McGrew, 2012).

Of the 31 additional relationships that could have been specified between the factors and latent constructs, only four demonstrated a modification index score of a significant magnitude to be taken into consideration. Further inspection of these suggested regression coefficients seem to make some semantic and theoretical sense. That the analysis revealed links between the *Gf* and *Gv* constructs is not surprising, given that this is also seen in performance-based batteries of cognitive abilities. For example, in the Perceptual Reasoning Index (PRI) of previous versions of the Wechsler scales, the Block Design and Picture Arrangement subtests are examples of problem-solving tasks that also involve visual-spatial ability (Horn & Hofer, 1992; Kaufman & Lichtenberger, 2002), hence reflecting the interplay of *Gf* and *Gv* abilities in the successful completion of these types of activities. Similar relations were also found for the *Gwm* and *Gs* constructs in the current study, and is consistent with the depiction of the interaction of these two abilities in the Cognitive Proficiency Index (CPI) of the Wechsler scales, and the Cognitive Efficiency cluster in the Woodcock Johnson Tests of Cognitive Ability. The CPI and Cognitive Efficiency cluster represents a set of functions relating to the proficiency with which an individual can process cognitive information automatically (McGrew, LaForte, & Schrank, 2014), highlighting the interdependence of *Gwm* and *Gs* abilities for the capacity to process information efficiently. These suggested modifications are thus consistent with what has been found about the inter-relatedness of some abilities, and is reflective of the nuances of human cognitive ability.

Model fit could have been improved by specifying these relationships, which in the current study would have resulted in a difference in χ^2 values of 175.65 between the original and resultant model, no change in SRMR values, and a reduction in RMSEA values of .12. This indicates that making changes to the model would not create a significantly more parsimonious model, and would achieve little to further our understanding of what is currently known about human cognitive abilities. It is commonly agreed amongst authors within the field that modification of a structural equation model changes the confirmatory nature of the analyses into a rather more exploratory one (Whittaker, 2012); hence, making modifications to the structural model simply for modification's sake is likely to be an unfruitful endeavour.

The current study thus provides initial support of a fine-grained operationalisation of human cognition into eight correlated but distinguishable dimensions that could be reliably used to collect information from teachers about a child's profile of cognitive ability and learning behaviours. However, although a CFA study based on strong theoretical support is an important part of the scale validation process, more evidence is generally needed before one can say with confidence that a newly developed scale is truly valid (DiStefano & Hess, 2005). The lion's share of the literature with regard to scale development has traditionally recommended obtaining a large sample size that can be reliably split into two to conduct EFA first, then CFA on the separate samples, to evaluate the EFA-informed *a priori* theory about the instrument's factor structure and psychometric properties (Cabrera-Nguyen, 2010). Additionally, existing restrictions of the sample size precluded the study's ability to cross-validate findings across balanced samples, a limitation of the current study given that cross-validation is highly recommended to ensure the predictive validity of modified models (MacCallum et al., 1992).

The stages of research at this point have focused primarily on the internal and structural validity of this newly developed scale. According to Strauss and Smith (2009), convergent validity (evidenced by strong associations with independent measurement tools of similar constructs) and discriminant validity (demonstrated by substantially weak correlations with measurement tools of dissimilar constructs) are critical to the validation of any psychological test. However, due to constraints often associated with ensuring the timely completion of a PhD dissertation, these important aspects of construct validity were not investigated in the current study and thus have yet to be ascertained for the ECCA scale.

Being able to compare scores obtained on the ECCA scale (i.e., teacher-report of child's cognitive ability) with scores obtained on performance-based measures (i.e., child's actual cognitive ability levels) would not only establish the predictive validity of the instrument, but could also inform the accuracy of teacher judgments. Utilising modification indices to ascertain discriminant validity across factors is also a limitation, since modification indices are derived from data and are hence susceptible to capitalisation on chance variation (Mungas et al., 2014). Although its use in the current study was restricted, this nevertheless raises questions about the replicability of findings related to discriminant validity, and it will be important for future research to extend the current study's findings with different samples.

It has been argued that using factor analysis to test an instrument's construct validity could produce differing factor structures when applied in normative versus clinical populations (Delis, Jacobson, Bondi, Hamilton, & Salmon, 2003). Nevertheless, construct validity is a cumulative scientific process, and this study is the first step toward establishing initial evidence for the construct validity of the ECCA measure. Given that the current study utilised a sample of typically developing children, future research could test the extent to which the factor structure of the ECCA scale is generalisable to clinical populations, such as in children with specific learning disabilities (SLD). Findings from such studies would extend and provide evidence of the potential utility of the scale in its use as a screening tool for teachers to identify children who may benefit from a comprehensive psychoeducational assessment to ascertain their cognitive strengths and weaknesses.

Chapter 8

Final Conclusions

This dissertation underwent a series of progressive stages in order to find answers to the overarching questions framing the rationale of the research, and its final chapter aims to synthesise the body of literature reviewed and the findings of the resulting studies that were conducted within this research endeavour. There is currently a need for a practical tool that could be used in conjunction with cognitive assessment techniques such as the cross-battery approach (XBA) to assist practitioners in deciding which of the wide array of broad and narrow cognitive abilities should form the focus of a cognitive assessment, so that the practical and financial efficacy of this traditionally lengthy and costly exercise can be ensured. Hence, this thesis builds upon the work of Jacobs (2012) in the development of a self-report measure of three cognitive ability areas, and extends the previous study's findings to the development of an informant-report measure of eight cognitive ability areas. The aim was to determine if using a psychometrically and theoretically defensible model of cognitive abilities, such as Cattell-Horn-Carroll (CHC) theory, would allow for the development of a parent- and teacher-report measure that would elicit a valid index of children's cognitive ability via informant-report methods.

Prior to developing an informant-report measure, the capacity of parents and teachers to provide reliable and valid information about children's cognitive abilities first needed to be determined. This was explored by conducting meta-analytic reviews of the extant literature that have compared the relationship between parent- and teacher- reports of children's cognitive and academic abilities, and children's actual performance on standardised tests. The human and method factors that have been known to influence informant-reports were concurrently investigated as potential moderators. Informed by insights gained from the first phase of the research endeavour pertaining to the appropriate methodological design of informant-report items, the process of scale development and validation sought to devise informant-report items of CHC broad ability constructs that facilitate valid estimations from parents and teachers of children's cognitive abilities. By extension, the current research also examined the proposition that a lack of theory guiding the development of measures used to obtain parent and teacher estimates of children's abilities has been more limiting than has been acknowledged. An integrative summary of chapters and overview of the major findings and conclusions of the respective studies is provided, and the findings are also discussed in terms of its original contribution to current knowledge in the field. Subsequently, the implications for furthering knowledge of cognitive ability assessment and scale development, and practical uses for the newly created valid informant-report measure within the educational context are highlighted. Finally, the chapter will conclude with a discussion of the limitations of the study, and provide suggestions for how future research can address these issues.

8.1 Overview of Chapters, Major Findings and Conclusions

The introductory chapter of this thesis discussed the context of the research, that being the need for practical and economic solutions to the time-consuming and expensive traditional performance-based method of obtaining an index of cognitive functioning. Since parents and teachers are already frequently utilised as valuable sources of information for children's functioning in other psychological constructs, it was proposed that an informant-report measure of children's cognitive abilities, used in combination with XBA techniques, could help practitioners hone in on the abilities which are most relevant to the referral concern, which could streamline the assessment process. However, efforts to develop informant-reports based on theoretically robust psychometric principles have thus far been lacking. The failure to take into consideration basic psychometric principles in developing informant-report methods means that inadequate measurement conditions of existing informant-report measures could hinder the capacity of parents and teachers to provide valid estimations of children's cognitive functioning. Thus, this dissertation set out to develop an informant-report instrument of eight CHC broad abilities that have been linked to the successful acquisition of literacy and numeracy skills, which could be used to inform and support the results of a traditional cognitive assessment. Subsequently, the next chapter detailed a review of the extant literature pertaining to the measurement of intelligence, parents and teachers as informants of children's cognitive abilities, and the human (child gender, child age, parent gender, professional experience of teacher) and methodological (criterion test match, number of items per construct, type of comparison, item reference) characteristics that have been known to influence how accurate these informants are in their estimations.

The first two research questions involved the examination of the factors that influence the validity of parent-reports and teacher-reports of children's cognitive abilities, and this first study of the dissertation is reported in Chapters 3 and 4. This was explored by conducting meta-analytic reviews of previous research that investigated the relationship between parent- and teacher-reports of, and performance-based measures of, children's cognitive and academic ability. Inspection of the summary correlation coefficients revealed that overall, parents and teachers can accurately estimate children's abilities to a moderate degree, and can therefore be considered as valid informants of children's cognitive functioning. Moderator analyses of certain human and measurement characteristics in their potential effect on informant accuracy provided some preliminary indications of how a psychometrically and theoretically defensible informant-report measure of children's abilities could be developed, and these insights guided the development of the ECCA that was undertaken in the second study of this dissertation. The individual differences that naturally occur amongst humans, namely the child's gender and age, parent gender, and how experienced the teacher is, were found to have nonsignificant moderating effects on how accurate parents and teachers are when reporting on a child's cognitive and academic functioning. Instead, the methodological aspects of measures used to obtain parent- and teacher-estimates were found to play a more significant role in informant accuracy. Overall, the findings highlight the importance of selecting appropriate instruments to use when assessing parent and teacher estimations. Specifically, the measurement conditions that were found to facilitate valid estimations from informants included the use of inter-individual comparisons in the scale design, and using theoretically analogous criterion measures when ascertaining the validity of informant estimations.

The second study conducted aimed to address the third research question of this thesis, which was to determine an appropriate method for devising scale items that facilitate valid estimations from informants regarding children's cognitive abilities, and to investigate the content validity of these developed scale items. This research endeavour was reported in Chapter 5. Following scale development guidelines as outlined by the Standards for Educational and Psychological Testing (APA, AERA, & NCME, 2014), and as recommended by DeVellis (2017), the multi-item, inter-individual *Estimate of Children's Cognitive Abilities* (ECCA) instrument was created with behavioural operationalisations of the following CHC broad abilities: Fluid Reasoning (*Gf*), Comprehension-Knowledge (*Gc*), Short-Term Working Memory (*Gwm*), Learning Efficiency (*Gl*), Retrieval Fluency (*Gr*), Visual-Spatial Processing (*Gv*), Auditory Processing (*Ga*), and Processing Speed (*Gs*). These abilities were chosen as they have consistently been shown to be important for the successful acquisition of literacy and numeracy skills (Flanagan, Ortiz, et al., 2013). The ECCA scale utilised a multi-item format to ensure adequate construct representation of each broad ability area as reflected in the numerous narrow abilities subsumed by broad abilities within the framework of CHC theory. Additionally, multi-item formats allow for aggregation to occur, which decreases error variance present in a measure and leads to increased validity (Ackerman & Wolman, 2007; Epstein, 1983; Paulhus, Lysy, & Yik, 1998). Finally, inter-individual comparisons were utilised in the response format of the ECCA due to findings from the current meta-analysis and from previous research showing that scale items that require respondents to use a frame of social reference results in greater validity in estimations (Mabe & West, 1982) – simply put, the evidence from the current thesis and in the extant literature suggests that it is more meaningful to benchmark performance compared to peers. Results of the expert panel review step of the scale development process as part of the second study indicated that item quality, and therefore content validity of the ECCA, could be further improved in terms of construct relevance, and item clarity and conciseness. Therefore, after writing revisions of the problematic items as identified by the expert panel, the additional step of pre-testing items via cognitive interviewing with psychologists and educational specialists was implemented to provide further evidence of content validity. The inclusion of cognitive interviews as a formal step in the scale development process proved to be a worthwhile endeavour, as it improved the content validity of the ECCA scale items. This indicates a progression towards the successful translation of psychological jargon into layperson language, an important issue that has been given limited attention in the development of similar parent- and/or teacher-report measures of children's abilities.

Subsequently, the overarching aim of the fourth research question of this dissertation was to investigate the structural validity of the ECCA scale via administration to an initial validation sample, and findings from this first phase of the fourth study were reported in Chapter 6. Due to the small sample size obtained, a two-step process was utilised to avoid the effects of sampling error that is associated with factor analysis. Using exploratory factor analysis techniques, the results indicated that a three-factor solution was most optimal for the dataset, which was not consistent with *a priori* expectations of an eight-factor solution that was in accordance with the eight broad cognitive abilities contained within CHC theory. Instead, the factors found were considered to be collections of abilities that resembled what was found in Carroll's (1993) seminal work in his factor-analytic survey of human cognitive abilities. A content analysis of "problematic" items revealed some justification as to why these items had loadings on more than one factor, mainly to do with the theoretical similarities and inherent system interdependence that some cognitive abilities are known to share. The initial exploration of the ECCA scale's underlying structure therefore showed some evidence for one that is consistent with earlier cognitive ability literature, hence highlighting the potential for such an informant-report measure to provide valid estimations about a child's cognitive ability. However, further evidence in support of the ECCA's structural validity in terms of its alignment was needed, giving rise to the second phase of the third study as reported in Chapter 7, which used the confirmatory factor analytic technique of structural equation modelling.

Following guidelines proposed by Holmes-Smith and Rowe (1994), a series of single-factor congeneric measurement models were conducted to first assess the convergent validity of items for each of the eight cognitive ability constructs. After the culling of items with low squared multiple correlations or covarying of error terms of items in accordance with modification indices, inspection of the various fit indices used indicated ideal fit for all measurement models, thus displaying adequate convergent validity for each of the eight cognitive ability areas under investigation. The structural model of the ECCA instrument was then built with derived composite scores built on the parameters established in the construction of the measurement models, and overall, displayed highly acceptable discriminant validity across all factors. The results of this phase thus provided initial evidence of a fine-grained operationalisation of human cognition into eight correlated but differentiable dimensions that can be reliably used to collect information from teachers about a child's profile of cognitive strengths and weaknesses.

The results of this project led to the overall conclusion that the development of an informant-report measure of cognitive abilities such as the ECCA can be considered as a promising first step in addressing several issues. First, the use of such a measure could alleviate some of the concerns regarding the traditional psycho-educational assessment process, and could be a practical tool that parents, teachers, and students can benefit from. Additionally, the development of the ECCA instrument represents a progression in knowledge of current methodologies of scale development, and also extends the assessment of the ecological validity of CHC theory in educational contexts. The following section discusses the wider implications of this research, followed by suggestions for future research aimed at extending and addressing the limitations of the current dissertation.

8.2 Implications

8.2.1 A solution to lengthy and expensive assessment procedures

At the core of being an educational psychologist is developing competencies in the administration, scoring, interpretation, and reporting the results of measures that assess an individual's level of intellectual abilities and academic skills, and other attributes relating to educational performance (Bowles et al., 2016). Psychoeducational assessments are important as they pave the way to facilitating diagnostic decisions, and can inform the development of educational interventions that capitalise on the individual's cognitive strengths to compensate for their weaknesses (Fiorello, Hale, & Wycoff, 2012). However, other activities, such as individual and/or group counselling, research, and partaking in professional development courses, are also important aspects of a psychologist's professional responsibility. Research that has investigated how educational psychologists allocate their working hours to the various psychological activities has reported that 50-70% of time is spent engaged in the intellectual assessment process (Bramlett, Murphy, Johnson, Wallingsford, & Hall, 2002; Falotico, 2015; Harris & Joy, 2010; Nastasi, Varjas, Bernstein, & Pluymert, 1998). Additionally, the discrepancies reported by school psychologists in the ideal and actual amount of time spent on psychoeducational assessments (Reschly & Wilson, 1995) imply that there could be some dissatisfaction in the disproportionate amount of time spent on testing and evaluation.

According to Gothard (2013), the length of time taken to conduct a truly comprehensive evaluation can vary from one child to the next, but the process can take up to five or six hours, sometimes conducted over a span of several days. Research that has investigated psychologists' allocation of time in psychoeducational evaluations has found that length of time taken for an assessment (including administration, scoring, and interpretation of cognitive ability test batteries) for the presence of learning disability and intellectual disability takes on average, 4.63 hours and 4.75 hours respectively (Taub & Valentine, 2014). Their associated standard deviations (1.30 and 1.33 hours respectively) also suggests that the total time spent on psychoeducational assessment was often reported as over an hour more than was indicated in the mean values. These findings are consistent with Jacobs' (2012) interpretation of the results from similar studies conducted previously (Camara, Nathan, & Puente, 2000; Sweet, Peck, Abramowitz, & Eitzweiler, 2002; as cited in Jacobs, 2012), indicating that these figures and their implications on cognitive ability assessments have remain largely unchanged for almost two decades.

It is thus apparent that the time-consuming aspect of comprehensive cognitive assessments is a long-standing issue from the practitioner's perspective, since there is much less time available for psychologists to commit to counselling and consultation related activities that are equally important to the profession (Falotico, 2015; Harris & Joy, 2010). On the client's part, however, the problems associated with this lengthy exercise could extend to monetary issues as well. To ensure appropriate professional use of test batteries and that test results are reliable and valid, it is often the case that only specialists (e.g., psychologists) who have been appropriately trained in the administration and interpretation of standardised tests can be contracted for psychoeducational assessment services (Zucker, 2004). However, as was described in Chapter 2 (section 2.1.1), the current recommended hourly rate for a psychologist to conduct assessments is $251 in Australia (APS, 2018). Combined with the lengthy nature of assessment processes as delineated above, it can therefore be inferred that the accessibility of this important service to all segments of society, in particular those of lower socioeconomic statuses who are likely most in need of it, is doubtful. Although the option of accessing free psychoeducational assessment services via the public sector is available for clients in need of financial aid, such as through psychologists operating in government schools (Jacobs, 2012), long waiting lists and having to meet strict eligibility requirements (e.g., only referral concerns pertaining to the possible diagnoses of Intellectual Disability or Severe Language Disorder are accepted for educational assessment referrals) indicate that the likelihood of obtaining a timely intellectual assessment in this way is low.

There is thus a real need to find economical solutions, in terms of time and cost, to the performance-based assessment process, to improve practitioner's time allocation to this activity and to better meet and serve clients' needs. Recent technological innovations in the cognitive ability testing field has led to the development of the Q-interactive system, a digital tablet-based platform that allows for the administration, scoring, and interpretation of the Wechsler intelligence scales and other individually-administered tests (Clark, Gulin, Heller, & Vrana, 2017). Although this represents a significant advancement for intelligence testing in terms of the improved convenience and flexibility, and is certainly less cumbersome than the traditional paper-and-pencil version, it does not come without its limitations. The Q-interactive is the first attempt by any testing company at developing a computer-supported administration of an individualised test (Noland, 2017). Hence, some current practitioners who are already well-versed with the paper-and-pencil method may feel that the move to a digitised version is wholly unnecessary, and thus could express reluctance to undergo further training for the Q-interactive testing system. Additionally, at the time of writing, the developers of the Woodcock Johnson have yet to publish a digital version that can be used with the Q-interactive – given that it is the only battery of cognitive abilities that provides adequate measures of all eight CHC broad ability areas that are known to be important for academic achievement, practitioners who choose to follow the "digital only" route may be short-changed in their capacity to conduct cognitive ability assessments that are truly comprehensive.

An informant-report tool such as the ECCA scale could therefore be used as an adjunct to alleviate some of the financial and practical problems associated with cognitive assessment. Using the ECCA scale in conjunction with techniques such as cross-battery assessment (XBA) can help to streamline the traditionally lengthy process of cognitive ability assessment by assisting practitioners in focusing on the cognitive and academic domains that are most relevant to the referral concerns and subsequent planning of educational interventions (Groth-Marnat, 1999). The CHC taxonomy of human cognitive abilities purports that individuals should be assessed for the total range of abilities specified by the theory (Carroll, 1997). However, Carroll (1997) also indicated that "research is needed to spell out how the assessor can select what abilities need to be tested in particular cases" (p. 129). Hence, extensive research into the investigation of how deficits in specific broad ability domains are related to and manifested in specific academic domains (Cormier, Bulut, McGrew, & Frison, 2016; Evans, Floyd, McGrew, & Leforgee, 2002; Floyd, Evans, & McGrew, 2003; Proctor, 2012; Taub et al., 2008) has provided a strong evidence base arguing for the selective, flexible, and referral-based intelligence testing, of which narrow abilities should form the primary focus (McGrew & Wendling, 2010). Further, significant variability within people's cognitive ability profiles is especially evident in clinical populations such as learning disability (Fletcher, 1994), and within the gifted populations as well (Lohman, Gambrell, & Lakin, 2008). This means that any two individuals who present with the same referral concern are likely to exhibit varying patterns of peaks and troughs in cognitive ability areas (Jacobs, 2012). A cognitive ability assessment that places emphasis on testing the child's performance narrow cognitive abilities is thus more likely to pick up on these individual differences in learning profiles.

Extensive advances in the theory and research regarding the XBA approach (Flanagan & McGrew, 1997; Flanagan et al., 2000; Flanagan & Ortiz, 2001; McGrew & Flanagan, 1998) for the past twenty years has enabled practitioners to make reliable, systematic and theory-based interpretations of any test battery, with the option of supplementing it with additional subtests from other cognitive, academic, and neuropsychological batteries. The XBA approach thus allows for the formation of a psychometrically defensible and holistic view of an individual's cognitive strengths and weaknesses (Flanagan, Ortiz, et al., 2013). However, choosing which ability areas, from the wide array of cognitive abilities available, should form the main focus of the assessment could still be overwhelming to many practitioners. Using a tool such as the ECCA could facilitate this process, as practitioners can reliably and validly obtain accurate perceptions from parents and teachers, who are significant stakeholders in a child's life, about where the child's cognitive strengths and weaknesses lie, and ensure that emphasis is placed on these ability areas during assessment.

Comprehensive psychoeducational assessments are often referred to as an iterative process, due to the necessity of having to properly integrate data obtained from the cognitive ability assessment, as well as from other sources such as background information and behavioural observations. This implies that psychologists may have already engaged in what they thought was a comprehensive psychoeducational evaluation, only to find at a later stage that further data is required to assist with informing results of the initial assessment (Dombrowski, 2015). With the use of an instrument such as the ECCA, clinicians can readily make informed decisions about which abilities are most pertinent to the referral problem and therefore, should form the focus of the assessment. This may reduce the need for additional cognitive assessments, and thereby minimise the time and costs that are typically associated with the process. Subsequently, this may lead to increased accessibility of such services to wider segments of the population.

Finally, aside from monetary benefits to the client, the reduction of time spent on cognitive ability assessments may also allow for a measurement process that seeks to reduce error that is outside of the test administrator's control. Some degree of measurement error is inevitable in the testing process, which can be minimised by test administrators (e.g., ensuring the test is conducted in a conducive environment with little visual and auditory distractions) . However, other sources of error that are known to impact the validity of test results can arise from non-cognitive factors that are intrinsic to test-takers (Jacobs, 2012). These include test anxiety, reduced motivation and concentration, as well as test fatigue, an unfortunate but often inevitable result of lengthy and extensive cognitive testing (Jacobs, 2012; Johnson, 1953; Reeve & Bonaccio, 2008; Sievertsen, Gino, & Piovesan, 2016). Hence, if the clinician is able to plan and carry out cognitive assessments that are individually tailored to adequately answer each client's individual referral concern and inform subsequent interventions, this would reduce the time and effort on the client's part to engage in the assessment process. This potentially would lead to a minimisation of measurement error, and a likely increase in the validity of the test results obtained.

8.2.2 Enabling laypersons' understanding of CHC theory

The development of CHC theory was in part inspired by the need to establish a common nomenclature of the vast range of human cognitive abilities, much like what the periodic table is to chemistry (Flanagan, Ortiz, et al., 2013; Horn, 1998; Schneider & McGrew, 2012). The adoption of CHC theory as a common set of terms and definitions in the domain of human intelligence was argued for by McGrew (2009) in its utility for facilitating communication, guarding against misinterpretation, and as a theoretical basis for testing *a prioi* hypotheses regarding aspects of human cognitive abilities (Flanagan, Ortiz, et al., 2013). The results of the current dissertation, an extension of Jacob's (2012) development of a self-report measure of the CHC ability areas of *Gf, Gc,* and *Gv*, has therefore successfully replicated and extended its findings in the development of an informant-report measure of these three ability areas and the other CHC ability areas of *Gwm, Gl, Gr, Ga* and *Gs*. It also provides further evidence for the potential of the nomenclature of CHC theory to be extended beyond the domain of scientific research to practical contexts for use by members of the general public, as findings indicate that laypeople are able to understand the various CHC ability constructs and the distinctions between them, if the technicalities that are often characteristic of scientific theories are suitably translated into jargon-free language (Jacobs, 2012).

Psychological reports are the primary means for an educational psychologist to communicate the results of cognitive assessment to the client and other stakeholders such as parents and teachers (Mastoras, Climie, McCrimmon, & Schwean, 2015). As such, to be of value to those working with the child, it is important for the report to be understandable and meaningful for its readers. The use of technical jargon has been found to be a significant factor in negatively affecting the readability of psychological reports and the reader's ability to comprehend the results of the assessment (Brenner, 2003; Groth-Marnat & Horvath, 2005; Harvey, 1997, 2006; Mastoras et al., 2015). If a psychoeducational report contains technical terms that hinders its readability, it negatively impacts on parents' and teachers' understanding of what it means for the child. The likelihood of its results being used to inform the planning of interventions to help the child is then likely reduced, thereby rendering the psychological report and all the pertinent information it contains useless. Hence, the introduction of an informant-report tool that translates the technicalities of CHC ability constructs into a language that is readily understood by parents and teachers can assist in practitioners ensuring that the psychoeducational report produced at the end of the assessment process is similarly free of technical jargon.

The creation of a valid informant-report scale of children's cognitive abilities that has been undertaken in the current study thus represents a significant step toward bridging the gap in laypersons' knowledge of CHC theory. This gap is apparent with psychologists' higher level of familiarity with the model due to years of extensive training in psychometric theory (Fiorello et al., 2009). Although understanding the intricacies and mechanics of the CHC ability constructs is important for psychologists, parents and teachers are likely to be more interested in the practical implications of different CHC abilities in classroom activities (Fiorello et al., 2009). Hence, by completing an informant-report measure that contains operationalisations of the CHC cognitive abilities as day-to-day behaviours and cognitive tasks that are often seen in the classroom, these significant adults in a child's life could begin to gain some insight as to how certain cognitive deficits are manifested in academic tasks. The use of CHC theory as common terminology that can be understood by professionals and laypeople alike would also assist in conducive communication and collaboration between the psychologist and all parties who are involved (i.e., parent, teacher and child) to effectively carry out the resulting recommendations from a cognitive assessment (Jacobs, 2012). Once parents and teachers achieve an appropriate level of understanding of the practical implications of CHC theory, they would be much more likely to appreciate the significance of why particular educational strategies were recommended based on the results found from the cognitive assessment about the child's CHC profile. Subsequently, this then increases the willingness of these important stakeholders in accepting and implementing the CHC-based interventions and strategies to support the child's learning and academic life.

A shared language for CHC cognitive abilities as introduced by an informant-report measure can also have the potential to affect teaching and learning directly (Fiorello et al. 2009; Jacobs, 2012). If teachers are able to wholly understand how the different CHC broad abilities affect learning and achievement outcomes, they would be more likely to meaningfully integrate them into their instructional planning and implementation.

Additionally, it is important that teachers are able to identify the underlying cognitive abilities that may be affecting the student's performance (Petrucelli, Fiorello, & Thurman, 2010), so that they are able to accurately refer the student for the appropriate support and implement interventions that will be successful in helping the student. The use of the ECCA instrument can also be extended to older-aged students (e.g., those in upper primary and high school), who could complete such a measure to develop greater self-insight in beginning to understand and gain awareness of where their strengths and weaknesses lie in their cognitive profile. Interpreting the unique constellation of cognitive and academic abilities can help them understand what compensatory strategies and accommodations might be most effective for their learning (Flanagan, Fiorello, & Ortiz, 2010). Empowering students with the knowledge of CHC abilities can thus give them an increased capacity to recognise their own cognitive strengths and limits. Moreover, being able to accurately estimate one's cognitive capabilities has been determined to be important for successful life outcomes. Erroneous self-perceptions of intellectual capacity have been shown to negatively impact self-confidence and psychological health (Stone, Dodrill, & Johnson, 2001); hence, it is important for young people to develop self-insight about their own cognitive capabilities sooner rather than later.

8.2.3 A typical performance measure of CHC abilities

As was described in Chapter 2 (Section 2.1), the measurement of cognitive ability has historically been assessed with tests of maximal performance, where the individual's performance on cognitive tasks is assessed within a controlled setting that is devised to facilitate the test-taker's best performance (Dewey, Crawford, & Kaplan, 2003). Tests of typical performance, in contrast, attempt to estimate how the individual usually behaves in everyday situations, and are traditionally used with constructs such as personality, where the administrator is more concerned with how the person normally feels instead of what the person is capable of feeling (Fiske & Butler, 1963). Although the clinical and diagnostic utility of maximal performance measures of cognitive ability is widely accepted, scores obtained on such measures do not always guarantee that a similar performance will be exhibited under everyday circumstances (Cronbach, 1949), and does not enable a comprehensive assessment of an individual's strengths and weaknesses (Ackerman, 1994; Livingston, Jennings, Reynolds, & Gray, 2003). In addition, it can be argued that a child's school performance is much more appropriately considered to be his or her "typical behaviour", as the criterion is best conceptualised as what the individual achieves over a period of time, rather than in a single occasion (Ackerman & Kanfer, 2004). Thus, a maximal performance measure of intelligence (e.g., Wechsler scales) that is used in conjunction with a typical performance measure of cognitive skills (e.g., ECCA instrument) may provide a more holistic insight into an individual's cognitive profile of strengths and weaknesses than using either measure alone (Lamb, 2008). Additionally, the prediction of children's academic achievement could also be improved beyond what is possible with using maximal performance measures alone (Yen, Konold, & McDermott, 2004).

Typical performance measures of abilities for use with children are few and far between. The Learning Disabilities Diagnostic Inventory (LDDI) is the one of only two measures that are commercially available for use with 8 to 17 year old children in the identification of skill patterns in specific academic ability areas (e.g., reading or writing) that are characteristic of individuals known to have a learning disability in that area (e.g., dyslexia or dysgraphia; Hammill & Bryant, 1998). The LDDI is designed be completed by a professional (e.g., teacher, speech pathologist) who is well-versed in the child's capacity in key academic areas such as reading, listening, writing, and mathematics. Another tool is Dehn's (2012) Children's Psychological Processing Scale, a teacher-report measure of children's cognitive abilities and other neuropsychological processes such as executive functioning and working memory. Other ventures into the development of informant-report measures of typical academic and cognitive skills include the Parent Rating Scale of Everyday Cognitive and Academic Abilities (PRECAA; Williams et al., 1991) and the

Ratings of Everyday Academics and Cognitive Skills (REACS; Lamb, 2008). The psychometric evidence of these informant-report measures of high internal consistency and predictive validity of children's academic achievement highlights that parents and teachers are indeed capable of accurately estimating children's typical level of cognitive abilities. The creation of the informant-report ECCA scale in the current study thus builds upon the success of previous attempts at developing typical performance measures of cognitive abilities, and further extends past research in its item development being guided by CHC theory, which has not been done before. The potential of the ECCA in providing a timely and easy to administer assessment of a child's typical functioning in various CHC cognitive domains via everyday behaviours could therefore be said to provide further evidence of the ecological validity of CHC theory in a practical context.

8.2.4 Extending knowledge in scale development processes

The results of the current research project indicate that parents and teachers are indeed valid and reliable reporters of children's cognitive functioning, and are more so when information is obtained from them using instruments utilising specific measurement conditions that facilitate the occurrence of valid estimations. The methodological decisions made by scale developers can thus determine how accurate parents' and teachers' estimations can be, and is an important finding that highlights the significance of undertaking theoretically robust scale development processes in future designs of informant-report measures.

The process of developing a scale often involves sophisticated and systematic processes requiring theoretical and methodological rigour (L. A. Clark & Watson, 1995; Nunnally, 1967), and it has been said that scale development is critical to expanding the knowledge base in human and social sciences (Morgado, Meireles, Neves, Amaral, & Ferreira, 2017). The development of measurement scales without adequate and rigorous theoretical justification is known to be a major cause of a rather incomplete state of knowledge in certain fields of research such as information systems (Recker & Rosemann, 2007). Likewise in the area of performance-based measures of cognitive ability, studies of human intelligence have been carried out for over a century; however, it is only in the recent two decades that revisions of intelligence batteries, such as in the Stanford-Binet Intelligence Scales – Fifth Edition (Roid, 2003), Kaufman Assessment Battery for Children – Second Edition (A. S. Kaufman & Kaufman, 2004), and Differential Ability Scales – Second Edition (Elliott, 2007), that adherence to psychometrically robust theory is explicitly stated in their respective technical manuals.

In terms of existing measures pertaining to parent- and teacher-reports of children's cognitive functioning, the well-established concept of intelligence as a multidimensional construct (Carroll, 1993; Gustafsson & Undheim, 1996; Schneider & Newman, 2015) implies that any scale that is designed to measure it should comprise of multiple items crafted to reflect an adequate theoretical understanding of each dimension (Carpenter, 2018). However, the lack of theoretical foundation used as a guiding framework in the development of questionnaires to obtain parent- and/or teacher-reports of children's cognitive and/or academic skills is the norm rather than an exception within the current literature. The example of "*How well do you think your child is doing at reading?*", a questionnaire item contained within a measure used by Pesu et al. (2016) to obtain informant's expectations about their child's reading ability, could be considered as overly simplistic. The broad ability of reading is multifaceted as it is comprised of several narrow abilities, that being reading fluency, reading decoding, and reading comprehension (McGrew et al., 2014). That the expectation of such items, which are inherently inadequate theoretical representations of the multidimensionality of abilities as constructs, is to produce valid estimates from parents and teachers about their children therefore seems to be theoretically and psychometrically unfounded.

The current author acknowledges the existence of informant-report measures of children's intelligence that have been underpinned by Gardner's (1983) Multiple Intelligences (MI) theory (e.g., Furnham, 2000; Furnham & Petrides, 2004; Hernández-Torrano et al., 2014). However, as mentioned in Chapter 2 (Section 2.6.1.1), the educational appeal of MI theory is offset by its lack of psychometric evidence (Allix, 2000; Sternberg & Kaufman, 1998; Waterhouse, 2006). Since MI theory does not possess measurable psychological subcomponents for each of the seven intelligences (Allix, 2000), there exists no standardised performance-based tool to measure the constructs in MI theory. Given the importance of using criterion measures of ability that are theoretically parallel to the informant-report measure of ability in facilitating accurate estimations that was found in the current dissertation, it can be inferred that valid conclusions about the accuracy of informant-reports of the intelligence constructs contained within MI theory cannot be drawn. Therefore, future research should take into consideration the caveats of certain intelligence theories while making the choice in relation to the theoretical framework that should be used as a guide in the development of informant-report measures.

The recent evolution in cognitive theory and testing brought about by the introduction of CHC theory means that the use of inappropriately developed measures of informant-reports of cognitive abilities either without any theoretical foundation, or with the use of non-validated theories, is now inexcusable. At the time of writing, however, there is a very limited application of CHC theory in the development of parent- and/or teacher-report measures of children's cognitive abilities. Only one parent-report study (Waschbusch et al., 2000), published almost twenty years ago, used the predecessor of CHC theory (*Gf-Gc*) as a theoretical basis for scale item development. Although there are studies that have developed measures intended for teacher responses (Fiorello et al., 2012; Singh, 2016) using scale items based on CHC theory, the focus of this research was surrounding the importance of CHC abilities in the classroom – the teacher participants were not directly estimating children's level of competencies in the ability areas, and so these measures cannot be considered as teacher-report measures of children's cognitive ability levels. The development of the ECCA in this dissertation is therefore a worthwhile endeavour in an attempt to facilitate valid estimations of children's intelligence from parents and teachers, and contributes to supporting evidence of the ecological validity of CHC theory.

The current dissertation also augmented the traditional method of scale development processes by integrating cognitive interviewing as a way to qualitatively pre-test potentially problematic items before releasing the scale to the general public. As part of the item development process, items are typically reviewed by a panel of experts for content validity in terms of clarity, relevance to the construct, and construct-irrelevant content (C. H. Peterson, 2012). Ratings by individuals who are considered to be knowledgeable in the field of interest can yield useful and important information, yet such expert ratings do not include an in-depth analysis of the response processes that respondents would go through when answering the developed scale, such as understanding the intended meaning of the scale item, retrieving relevant information from memory, and correctly selecting a response that reflects the respondents' actual judgment (C. H. Peterson, Peterson, & Powell, 2017; Tourangeau, 1984). Since any of these cognitive operations are a potential source of error, and experts may fail to recognise problems from the respondent's perspective, the scale developer may not detect if respondent interpretation of an item is different to the developer's intention without the use of cognitive interviewing (Hughes, 2004; Presser et al., 2004). Since respondent interpretation of scale items is the cornerstone of all findings and inferences that are made from the measure, the validity of the measure is inevitably negatively affected by respondent misinterpretation (AERA, APA, & NCME, 2014).

Despite the many benefits afforded by cognitive interviewing, its application as a pre-testing technique within the scale development literature has occurred only recently (Castillo-Díaz & Padilla, 2013; Dietrich & Ehrlenspiel, 2010; Woolley, Bowen, & Bowen, 2006), and is even more limited in the research domain of informant-reports of cognitive abilities. Although the extra step of cognitive interviewing that was undertaken in the current dissertation translated into more time and effort spent on developing the ECCA scale, it proved to be a meaningful endeavour as it not only provided more evidence of the measure's content validity, but also revealed areas and reasons for misalignment between respondent interpretation and the planned meaning for several items. Additionally, since the interviews were completed early in the scale development process, it allowed for enough time to identify and correct identified issues in item comprehension and content coverage before subjecting the scale to more extensive validity studies. Hence, future research is encouraged to adopt cognitive interviewing in the development of new scales, as the adoption of best practices pertaining to scale development can yield stronger theoretical and practical concepts, and subsequently, a more stable foundation of knowledge (Carpenter, 2018).

8.3 Limitations and Directions for Future Research

Although the empirical investigation conducted within the current research project has extended previous research and current knowledge by systematically developing a theoretically and psychometrically defensible informant-report measure of children's cognitive abilities, there are a number of limitations that need to be highlighted as potential issues for future studies to address.

As was acknowledged in Chapter 7 (section 7.4), the current study provides preliminary evidence of the validity of the ECCA scale. However, further research is required to confirm and extend the findings of the current study. Within the current dissertation, the convergent validity of ECCA items with their respective broad abilities was ascertained via a series of single-factor congeneric measurement models, and the discriminant validity of each broad ability area was established with the overall structural model. Although these results are promising, other important aspects of validity such as predictive and concurrent validity should also be considered before the ECCA scale can be described as having adequate construct validity. Given its potential uses, such as for practitioners to narrow down the choice of which abilities should form the focus of the cognitive assessment, or as a screening tool for teachers to identify students with learning difficulties and/or possible learning disabilities, the implication for future research is to investigate the ability of the ECCA instrument to predict children's performance on traditional intelligence tests. This will also provide evidence of the accuracy of parents and/or teachers when reporting on children's CHC cognitive abilities.

The original intention of the current thesis was to develop an informant-report tool that could be used to obtain information about a child's level of cognitive functioning in the various CHC ability areas from both parents and teachers. This would have involved recruiting parent and teacher participants to allow for cross-validation to compare validity results across both types of informants. However, the time constraints that are often associated with the timely completion of a PhD dissertation precluded the current study's ability to recruit parent participants for the studies investigating the validity of the scale. The resulting consequence of this restriction is two-fold. Firstly, cross-validation using two separate samples is often recommended to allow for replication of the scale's factor structure by conducting EFA on one sample, and CFA on the other (Brown, 2006). However, this could not be achieved due to small ratios of participants to variables (5:1), with reviews of factor analytic literature showing that only 10% of studies with small ratios (2:1) producing correct solutions (Costello & Osborne, 2005). Even though the effects of sampling error were successfully avoided by using a two-step process as was explained in Chapter 6, future research could attempt to replicate and extend results of the current study by recruiting a new set of teacher participants and parent participants for cross-validation purposes. Further, it would be of interest to ascertain how the language used in the measure is understood by parents, who understandably may have less experience with and knowledge of how cognitive deficits may manifest in tasks and behaviours.

The second caveat of not being able to recruit parents as participants is that the ECCA instrument cannot yet be considered a multi-informant measure. The utilisation and interpretation of reports from multiple informants are critical aspects of best practice in evidence-based assessment (Dirks, De Los Reyes, Briggs-Gowan, Cella, & Wakschlag, 2012). An instrument that collects the subjective opinion of behaviours from multiple informants can give the practitioner a better picture of the child's behaviours in different contexts (e.g., home and school). However, previous research has found that discrepancies are common among multiple informants' reports (Achenbach, 2006), even when parallel or identical measures are completed (De Los Reyes, 2011). Therefore, the implication is for future research to determine if parents and teachers are able to provide concordant ratings of the child that is being referred, as it is paramount to the inter-rater reliability and clinical utility of the ECCA as a multi-informant measure.

Finally, although not critical, the sample size obtained could be larger and more varied in terms of participant characteristics to ensure generalisability of results. An inspection of the demographical information of the participants suggested an over-representation of female teachers (84.1%) reporting on male students (66.7%) in government (63.4%) primary schools (64.5%). This is not unexpected given that the teaching profession has historically been dominated by females (Drudy, 2008; McGrath & Van Bergen, 2017), and that women are more inclined to participate in surveys than men (Cull, O'Connor, Sharp, & Tang, 2005; Curtin, Presser, & Singer, 2000; Singer, Van Hoewyk, & Maher, 2000). To ensure a more even representation of boys and girls being rated, future research could amend the instructions of the scale to ask for ratings of either gender using a randomiser, which is available in online survey platforms such as Qualtrics. Lastly, the validation work that was conducted in the current study focused on evaluating constructs and removing items that performed poorly with regards to the internal validity of the scale. Hence, there is opportunity for future research to revise items to improve their quality, or generate new items for the cognitive ability constructs if necessary, although it should be noted that model fit across each of the eight cognitive ability areas was highly successful.

8.4 Concluding Remarks

This thesis set out to establish an appropriate methodology to develop behavioural operationalisations of cognitive skills that would facilitate accurate parent- and teacher-estimations of children's cognitive functioning in accordance with eight broad ability areas in CHC theory that are known to be important for the acquisition of literacy and numeracy skills. In overall conclusion, the results of this dissertation provide a promising starting point for establishing the validity of a newly created informant-report measure of children's CHC cognitive abilities, with further validation studies required to extend these findings. Meta-analyses of previous research within the parent- and teacher-report domain were conducted to guide the development of the Estimates of Children's Cognitive Abilities scale, which was designed as a multi-item, inter-individual measure containing items that referred to cognitive abilities as skills, tasks and behaviours that children would typically display in their day-to-day lives at home and in the classroom. The use of expert ratings and the incorporation of cognitive interviewing as an additional qualitative step to improve scale development methods helped to establish the ECCA's content validity. Exploratory and confirmatory factor analytic methods were then employed to uncover the scale's dimensionality using data collected from a sample of primary and secondary school teachers, and subsequently, adequate convergent and discriminant validity amongst the theorised eight CHC cognitive ability constructs was found.

The conclusion of this project does not in any way suggest that the ECCA instrument is complete, finalised and ready for use with the general public – the work to be done is greater than could be achieved in a single dissertation. Despite the fact that the exploration into the ECCA instrument's validity is far from finished, for a measurement scale in its infancy stage, a high degree of psychometric rigour has thus far been demonstrated. The initial exploration of the structural validity of this newly developed scale has shown some optimistic results, yet requires more evaluation and investigation before it can be deemed to be truly valid. From a "big picture" perspective, the aim of this project was to determine the capacity of parents and teachers to accurately estimate a child's level of functioning in the broad cognitive ability areas contained within CHC theory. It can be said with a reasonable level of certainty that these important adults are capable of providing valid reports of their children's cognitive abilities, and that teachers adequately demonstrated the reliability and validity of the newly developed ECCA scale to report on children's CHC cognitive abilities. Additionally, in developing a psychometrically and theoretically robust informant-report tool that attempted to translate psychological jargon to layperson language by operationalising cognitive abilities as observable behaviours, it can be said that we are one small step closer to developing a more effective multi-informant, multi-method measurement of abilities. Additionally, in giving parents and teachers the ability to report on their children's cognitive functioning in a meaningful and valid way, the ECCA measure can also be said to help these laypeople understand how strengths and weaknesses in particular CHC ability areas can manifest in children's learning and academic achievement, a level of insight that arguably can provide more meaning than an arbitrary number on a standardised test.

References

Abrami, P. C., Cohen, P. A., & D'apollonia, S. (1988). Implementation problems in meta-analysis. *Review of Educational Research, 58*(2), 151-179. doi:10.3102/00346543058002151

Achenbach, T. M. (2006). As others see us: Clinical and research implications of cross-informant correlations for psychopathology. *Current Directions in Psychological Science, 15*(2), 94-98. doi:10.1111/j.0963-7214.2006.00414.x

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*(2), 213-232.

Ackerman, P. L. (1994). Intelligence, attention and learning: Maximal and typical performance. In D. K. Detterman (Ed.), *Current topics in human intelligence: Theories of intelligence* (Vol. 4, pp. 1-27). Norwood, NJ: Ablex.

Ackerman, P. L., & Kanfer, R. (2004). Cognitive, affective, and conative aspects of adult intellect within a typical and maximal performance framework. In D. Y. Dai & R. J. Sternberg (Eds.), *Motivation, emotion, and cognition: Integrative perspectives on intellectual functioning and development* (pp. 119-141). Mahwah, NJ: Lawrence Earlbaum Associates.

Ackerman, P. L., & Lohman, D. F. (2006). Individual differences in cognitive functions. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 139-161). Mahwah, NJ: Lawrence Erlbaum Associates.

Ackerman, P. L., & Wolman, S. D. (2007). Determinants and validity of self-estimates of abilities and self-concept measures. *Journal of Experimental Psychology: Applied, 13*(2), 57-78.

Ahmed, I., Sutton, A. J., & Riley, R. D. (2012). Assessment of publication bias, selection bias, and unavailable data in meta-analyses using individual participant data: A database survey. *British Medical Journal, 344*(7838), 16. doi:10.1136/bmj.d7762

Alfonso, V. C., Flanagan, D. P., & Radwan, S. (2005). *The impact of the Cattell-Horn-Carroll theory on test development and interpretation of cognitive and academic abilities* (2nd ed.). New York, NY: Guilford Press.

Allix, N. M. (2000). The theory of multiple intelligences: A case of missing cognitive matter. *Australian Journal of Education, 44*(3), 272-288. doi:10.1177/000494410004400306

*Alloway, T. P., Gathercole, S. E., Kirkwood, H., & Elliott, J. (2009). The working memory rating scale: A classroom-based behavioral assessment of working memory. *Learning and Individual Differences, 19*, 242-245.

Allred, R. A. (1990). Gender differences in spelling achievement in Grades 1 through 6. *The Journal of Educational Research, 83*(4), 187-193. doi:10.1080/00220671.1990.10885955

*Alvidrez, J., & Weinstein, R. S. (1999). Early teacher perceptions and later student academic achievement. *Journal of Educational Psychology, 91*(4), 731-746.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York, NY: Macmillan Publishing Company.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103*(3), 411-423.

Arbuckle, J. L. (2013). *IBM SPSS Amos 22 user's guide*. Chicago, IL: Amos Development Corporation.

Arciuli, J., Gurisik, B., & Munro, N. (2010). Parental perceptions of children's reading ability. *Asia Pacific Journal of Speech, Language and Hearing, 13*(1), 61-66. doi:10.1179/136132810805335146

Aronson, J., & Steele, C. M. (2005). Stereotypes and the fragility of academic competence, motivation and self-concept. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 436-456). New York, NY: Guilford Publications.

*Aunola, K., Nurmi, J.-E., Lerkkanen, M.-K., & Rasku-Puttonen, H. (2003). The roles of achievement-related behaviours and parental beliefs in children's mathematical performance. *Educational Psychology, 23*(4), 403-421. doi:10.1080/01443410303212

*Aunola, K., Nurmi, J.-E., Niemi, P., Lerkkanen, M.-K., & Rasku-Puttonen, H. (2002). Developmental dynamics of achievement strategies, reading performance, and parental beliefs. *Reading Research Quarterly, 37*(3), 310-327.

Australian Psychological Society. (2018). APS National Schedule of recommended fees and item numbers for psychological services. Retrieved from https://www.psychology.org.au/for-members/resource-finder/resources/Business-admin-legal/APS-National-Schedule-Recommended-Fees-and-items

Babad, E. Y. (1985). Some correlates of teachers' expectancy bias. *American Educational Research Journal, 22*(2), 175-183. doi:10.3102/00028312022002175

Bagnato, S., Neisworth, J., & Munson, S. (1989). *Linking developmental assessment and early intervention: Curriculum-based prescriptions*. Rockville, MD: Aspen.

Balsis, S., Cooper, L. D., & Oltmanns, T. F. (2015). Are informant reports of personality more internally consistent than self reports of personality. *Assessment, 22*(4), 399-404. doi:10.1177/1073191114556100

Baron, I. S. (2004). *Neuropsychological evaluation of the child.* New York, NY: Oxford University Press.

Barrett, P. (2007). Structural equation modeling: Ajudging model fit. *Personality and Individual Differences, 42*, 815-824. doi:10.1016/j.paid.2006.09.018

*Barry, J. G., Tomlin, D., Moore, D. R., & Dillon, H. (2015). Use of questionnaire-based measures in the assessment of listening difficulties in school-aged children. *Ear & Hearing, 36*(6), e300-e313. doi:10.1097/AUD.0000000000000180

Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Mathematical and Statistical Psychology, 3*(2), 77-85. doi:10.1111/j.2044-8317.1950.tb00285.x

*Bates, C., & Nettelbeck, T. (2001). Primary school teachers' judgements of reading achievement. *Educational Psychology, 21*(2), 177-187. doi:10.1080/01443410020043878

*Baudson, T. G., Fischbach, A., & Preckel, F. (2016). Teacher judgments as measures of children's cognitive ability: A multilevel analysis. *Learning and Individual Differences, 52*, 148-156. doi:10.1016/j.lindif.2014.06.001

*Baudson, T. G., & Preckel, F. (2013). Development and validation of the German test for (Highly) Intelligent Kids - T(H)INK. *European Journal of Psychological Assesement, 29*(3), 171-181.

Baumgartner, H., & Homburg, C. (1996). Applications of structural equation modelling in marketing and consumer research: A review. *International Journal of Research in Marketing, 13*(2), 139-161. doi:10.1016/0167-8116(95)00038-0

Baumgartner, H., & Steenkamp, J.-B. E. M. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research, 38*(2), 143-156. doi:10.1509/jmkr.38.2.143.18840

Beatty, P. C. (2004). The dynamics of cognitive interviewing. In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 45-66). Hoboken, NJ: John Wiley.

Beatty, P. C., & Willis, G. B. (2007). Research synthesis: The practice of cognitive interviewing. *Public Opinion Quarterly, 71*(2), 287-311. doi:10.1093/poq/nfm006

Beavers, A. S., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G. J., & Esquivel, S. L. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research & Evaluation, 18*(6), 1-13.

*Begeny, J. C., Eckert, T. L., Montarello, S. A., & Storie, M. S. (2008). Teachers' perceptions of students' reading abilities: An examination of the relationship between teachers' judgments and sutdents' performance across a continuum of rating methods. *School Psychology Quarterly, 23*(1), 43-55. doi:10.1037/1045-3830.23.1.43

*Begeny, J. C., Krouse, H. E., Brown, K. G., & Mann, C. M. (2011). Teacher judgments of students' reading abilities across a continuum of rating method and achievement measures. *School Psychology Review, 40*(1), 23-38. doi:10.1177/1534508413475644

*Benner, A. D., & Mistry, R. S. (2007). Congruence of mother and teacher educational expectations and low-income youth's academic competence. *Journal of Educational Psychology, 99*(1), 140-153. doi:10.1037/0022-0663.99.1.140

Benson, E. (2003). Intelligence across cultures. *Monitor on Psychology, 34*(2), 56.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588-606.

Bergkvist, L. I., & Rossiter, J. (2007). The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research, 44*(2), 175-184. doi:10.1509/jmkr.44.2.175

Besnoy, K. D., Dantzler, J., Besnoy, L. R., & Byrne, C. (2016). Using exploratory and confirmatory factor analysis to measure construct validity of the Traits, Aptitudes, and Behaviors Scale (TABS). *Journal for the Education of the Gifted, 39*(1), 3-22. doi:10.1177/0162353215624160

*Beswick, J. F., Willms, J. D., & Sloat, E. A. (2005). A comparative study of teacher ratings of emergent literacy skills and student performance on a standardized measure. *Education, 126*(1), 116-137.

Bido, D. S., Mantovani, D. M. N., & Cohen, E. D. (2018). Destruction of measurement scale through exploratory factor analysis in production and operations research. *Gestão & Produção, 25*(2), 384-397. doi:10.1590/0104-530x3391-16

Bird, J. E., & Berman, L. S. (1985). Differing perceptions of mothers, fathers, and children concerning children's academic performance. *The Journal of Psychology: Interdisciplinary and Applied, 119*(2), 113-124. doi:10.1080/00223980.1985.10542877

Bleeker, M. M., & Jacobs, J. E. (2004). Achievement in math and science: Do mothers' beliefs matter 12 years later? *Journal of Educational Psychology, 96*(1), 97-109. doi:10.1037/0022-0663.96.1.97

*Bohlmann, N. L., & Weinstein, R. S. (2013). Classroom context, teacher expectations, and cognitive level: Predicting children's math ability judgments. *Journal of Applied Developmental Psychology, 34*(6), 288-298.

Bollen, K. A., & Stine, R. A. (1992). Bootstrapping goodness-of-fit measures in structural equation models. *Sociological Methods & Research, 21*(2), 205-229. doi:10.1177/0049124192021002004

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Hoboken, NJ: John Wiley & Sons.

*Bouffard, S. M., & Hill, N. E. (2005). Maternal perceptions of competence and children's academic adjustment: Longitudinal relations across early elementary school. *Social Psychology of Education, 8*, 441-463. doi:10.1007/s11218-005-1327-4

Bowles, T., Scull, J., Hattie, J., Clinton, J., Larkins, G., Cicconi, V., . . . Arnup, J. L. (2016). Conducting psychological assessments in schools: Adapting for converging skills and expanding knowledge. *Issues in Educational Research, 26*(1), 10-28.

Bradstreet, L. E., Juechter, J. I., Kamphaus, R. W., Kerns, C. M., & Robins, D. L. (2017). Using the BASC-2 parent rating scales to screen for autism spectrum disorder in toddlers and preschool-aged children. *Journal of Abnormal Child Psychology, 45*(2), 359-370. doi:10.1007/s10802-016-0167-3

Bramlett, R. K., Murphy, J. J., Johnson, J., Wallingsford, L., & Hall, J. D. (2002). Contemporary practices in school psychology: A national survey of roles and referral problems. *Psychology in the Schools, 39*(3), 327-335. doi:10.1002/pits.10022

Brannick, M. T. (1995). Critical comments on applying covariance structure modeling. *Journal of Organizational Behavior, 16*(3), 201-213. doi:10.1002/job.4030160303

Brenner, E. (2003). Consumer-focused psychological assessment. *Professional Psychology: Research and Practice, 34*(3), 240-247. doi:10.1037/0735-7028.34.3.240

Brindle, S., Douglas, F., Hundley, V., & van Teijlingen, E. (2005). Midwifery research: Questionnaire survey. *RCM Midwives Journal, 8*(4), 156-158.

Brockwell, S. E., & Gordon, I. R. (2001). A comparison of statistical methods of meta-analysis. *Statistics in Medicine, 20*(6), 825-840. doi:10.1002/sim.650

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

*Brummelman, E., Thomaes, S., Nelemans, S. A., de Castro, B. O., & Bushman, B. J. (2014). My child is God's gift to humanity: Development and validation of the Parental Overestimation Scale (POS). *Journal of Personality and Social Psychology, 108*(4), 665-679. doi:10.1037/pspp0000012

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS and SIMPLIS: Basic concepts, applications and programming*. Mahwah, NJ: Lawrence Earlbaum Associates.

Cabrera-Nguyen, P. (2010). Author guidelines for reporting scale development and validation results in the Journal of the Society for Social Work and Research. *Journal of the Society for Social Work and Research, 1*(2), 99-103. doi:10.5243/jsswr.2010.8

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validity by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81-105.

Campbell, L. (1997). How teachers interpret MI theory. *Educational Leadership, 55*(1), 14-19.

Carbonaro, M., & Bainbridge, J. (2000). Design and development of a process for web-based survey research. *The Alberta Journal of Educational Research, 46*(4), 392-394.

Carpenter, S. (2018). Ten steps in scale development and reporting: A guide for researchers. *Communication Methods and Measures, 12*(1), 25-44. doi:10.1080/19312458.2017.1396583

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York, NY: Cambridge University Press.

Carroll, J. B. (1997). The three-stratum theory of cognitive abilities. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (1st ed., pp. 122-130). New York, NY: Guilford Press.

Castillo-Díaz, M., & Padilla, J.-L. (2013). How cognitive interviewing can provide validity evidence of the response processes to scale items. *Social Indicators Research, 114*(3), 963-975. doi:10.1007/s11205-012-0184-8

Cattell, R. B. (1971). *Intelligence: Its structure, growth and action*. Amsterdamn: Elsevier.

Catts, H. W., Gillispie, M., Leonard, L., Kail, R. V., & Miller, C. A. (2002). The role of speed of processing, rapid naming, and phonological awareness in reading achievement. *Journal of Learning Disabilities, 35*(6), 509-524.

*Chamorro-Premuzic, T., Arteche, A., Furnham, A., & Trickot, N. (2009). Assessing pupils' intelligence through self, parental, and teacher estimates. *Educational Psychology, 29*(1), 83-97. doi:10.1080/01443410802520662

Chamorro-Premuzic, T., & Furnham, A. (2006). Self-assessed intelligence and academic performance. *Educational Psychology, 26*(6), 769-779. doi:10.1080/01443410500390921

Chen, J.-Q. (2004). Theory of multiple intelligences: Is it a scientific theory? *Teacher's College Record, 106*(1), 13-23. doi:10.1111/j.1467-9620.2004.00313.x

Child, D. (1990). *The essentials of factor analysis* (2nd ed.). London, UK: Cassel Educational Limited.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309-319. doi:10.1037/1040-3590.7.3.309

Clark, S. W., Gulin, S. L., Heller, M. B., & Vrana, S. R. (2017). Graduate training implications of the Q-Interactive platform for administering Wechsler intelligence tests. *Training and Education in Professional Psychology, 11*(3), 148-155. doi:10.1037/tep0000155

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics, 10*, 101-129.

Cockcroft, K., Bloch, L., & Moolla, A. (2016). Assessing verbal functioning in South African school beginners from diverse socioeconomic backgrounds: A comparison between verbal working memory and vocabulary measures. *Education as Change, 20*(1), 199-215. doi:10.17159/1947-9417/2016/559

Cocodia, E. A. (2014). Cultural perceptions of human intelligence. *Journal of Intelligence, 2*, 180-196. doi:10.3390/jintelligence2040180

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37-46.

Cohn, L. D., & Becker, B. J. (2003). How meta-analysis increases statistical power. *Psychological Methods, 8*(3), 243-253. doi:10.1037/1082-989X.8.3.243

*Coladarci, T. (1986). Accuracy of teacher judgments of student responses to standardized test items. *Journal of Educational Psychology, 78*(1), 141-146.

Cole, D. A., Ciesla, J. A., & Steiger, J. H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods, 12*(4), 381-398. doi:10.1037/1082-989X.12.4.381

*Commodari, E., & Guarnera, M. (2005). Attention and reading skills. *Perceptual and Motor Skills, 100*, 375-386. doi:10.2466/pms.100.2.375-386

Comrey, A. L., & Lee, H. (1992). *A first course in factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Considine, G., & Zappalà, G. (2002). The influence of social and economic disadvantage in the academic performance of school students in Australia. *Journal of Sociology, 38*(2), 129-148. doi:10.1177/144078302128756543

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*(2), 163-183. doi:10.1016/S0160-2896(01)00096-4

Conway, A. R. A., & Kovacs, K. (2013). New and emerging models of human intelligence. *Wiley Interdisciplinary Reviews: Cognitive Science, 6*(5), 419-426. doi:10.1002/wcs.1356

Cooper, H. M., & Lemke, K. M. (1991). On the role of meta-analysis in personality and social psychology. *Personality and Social Psychology Bulletin, 17*, 245-251. doi:10.1177/0146167291173002

Cormier, D. C., Bulut, O., McGrew, K. S., & Frison, J. (2016). The role of Cattell-Horn-Carroll (CHC) cognitive abilities in predicting writing achievement during the school-age years. *Psychology in the Schools, 53*(8), 787-803. doi:10.1002/pits.21945

Cormier, D. C., Bulut, O., McGrew, K. S., & Singh, D. (2017). Exploring the relations between Cattell-Horn-Carroll (CHC) cognitive abilities and mathematics achievement. *Applied Cognitive Psychology, 31*(5), 530-538. doi:10.1002/acp.3350

Cortina, J. M. (2002). Big things have small beginnings: An assortment of 'minor' methodological misunderstandings. *Journal of Management, 28*(3), 339-362. doi:10.1177/014920630202800305

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*(7), 1-9.

Costello, S. (2016). *Person-centred psychometrics and cognitive style: Moving from variables to a person-centred approach.* (Master of Psychology (Educational & Developmental) / Doctor of Philosophy), Monash University.

Cronbach, L. J. (1949). *Essentials of psychological testing* (2nd ed.). Oxford, England: Harper.

Cronbach, L. J. (1960). *Essentials of psychological testing* (2nd ed.). New York: Harper.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281-302. doi:10.1037/h0040957

Crowley, S. L., & Fan, X. (1997). Structural equation modeling: Basic concepts and applications in personality assessment research. *Journal of Personality Assessment, 68*(3), 508-531. doi:10.1207/s15327752jpa6803_4

Cull, W. L., O'Connor, K. G., Sharp, S., & Tang, S.-F. S. (2005). Response rates and response bias for 50 surveys of pediatricians. *Health Services Research, 40*(1), 213-226. doi:10.1111/j.1475-6773.2005.00350.x

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York, NY: Taylor & Francis.

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods, 1*(1), 16-29. doi:10.1037/1082-989X.1.1.16

Curtin, R., Presser, S., & Singer, E. (2000). The effects of response rate changes on the index of consumer sentiment. *Public Opinion Quarterly, 64*(4), 413-428.

Daniel, M. H. (2007). "Scatter" and the construct validity of FSIQ: Comment on Fiorello et al. (2007). *Applied Neuropsychology, 14*(4), 291-295. doi:10.1080/09084280701719401

De Los Reyes, A. (2011). Special section: More than measurement error: Discovering meaning behind informant discrepancies in clinical assessments of children and adolescents. *Journal of Clinical Child & Adolescent Psychology, 40*(1), 1-9. doi:10.1080/15374416.2011.533405

de Winter, J. C. F., & Dodou, D. (2012). Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size. *Journal of Applied Statistics, 39*(4), 695-710. doi:10.1080/02664763.2011.610445

*Dean, R. S., & Kundert, D. K. (1981). Intelligence and teachers' ratings as predictors of abstract and concrete learning. *Journal of School Psychology, 19*(1), 78-85.

Dehn, M. J. (2012). *Children's Psychological Processes Scale.* Onalaska, WI: Schoolhouse Educational Services.

Deimann, P., & Kastner-Koller, U. (2011). Maternal evaluations of young children's developmental status: A comparison of clinic- and non-clinic-groups. *Psychological Test and Assessment Modeling, 53*(2), 214-227.

*Delgado-hachey, M., & Miller, S. A. (1993). Mothers' accuracy in predicting their children's IQs: Its relationship to antecedent variables, mothers' academic achievement demands, and children's achievement. *The Journal of Experimental Education, 62*(1), 43-59. doi:10.1080/00220973.1993.9943830

Delgado-Rico, E., Carretero-Dios, H., & Ruch, W. (2012). Content validity evidences in test development: An applied perspective. *International JOurnal of Clinical and Health Psychology Espana, 12*(3), 449-460. doi:10.5167/uzh-64551

Delis, D. C., Jacobson, M., Bondi, M. W., Hamilton, J. M., & Salmon, D. P. (2003). The myth of testing construct validity using factor analysis or correlations with normal or mixed clinical populations: Lessons from memory assessment. *Journal of the International Neuropsychological Society, 9*(6), 936-946. doi:10.1017/S1355617703960139

DeMaio, T. J., & Landreth, A. (2004). Do different cognitive interview techniques produce different results? In S. Presser, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin, & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 89-108). Hoboken, NJ: John Wiley.

*Demaray, M. K., & Elliott, S. N. (1998). Teachers' judgments of students' academic functioning: A comparison of actual and predicted performances. *School Psychology Quarterly, 13*(1), 8-24. doi:10.1037/h0088969

Desimone, L. M., & Le Floch, K. C. (2004). Are we asking the right questions? Using cognitive interviews to improve surveys in education research. *Educational Evaluation and Policy Anlaysis, 26*(1), 1-22. doi:10.3102/01623737026001001

Detterman, D. K. (2011). A challenge to Watson. *Intelligence, 39*(2-3), 77-78. doi:10.1016/j.intell.2011.02.006

DeVellis, R. F. (2003). *Scale development: Theory and application* (2nd ed.). Thousand Oaks, CA: Sage Publications.

DeVellis, R. F. (2012). *Scale development: Theory and applications* (3rd ed.). Thousand Oaks, CA: Sage Publications.

DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). Thousand Oaks, CA: SAGE Publishing.

Dewey, D., Crawford, S. G., & Kaplan, B. J. (2003). Clinical importance of parent ratings of everyday cognitive abilities in children with learning and attention problems. *Journal of Learning Disabilities, 36*(1), 87-95. doi:10.1177/00222194030360011001

DeYoung, C. G. (2009). Intelligence and personality. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence* (pp. 711-737). New York, NY: Cambridge University Press.

Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P., & Kaiser, S. (2012). Guidelines for choosing between multi-item and single-item scales for construct measurement: A predictive validity perspective. *Journal of the Academy of Marketing Science, 40*(3), 434-449. doi:10.1007/s11747-011-0300-3

Diamantopoulos, A., & Sigauw, J. A. (2000). *Introducing LISREL*. London, UK: SAGE Publications.

Diamond, K. E., & Squires, J. (1993). The role of parental report in the screening and assessment of young children. *Journal of Early Intervention, 17*(2), 107-115.

Dietrich, H., & Ehrlenspiel, F. (2010). Cognitive interviewing: A qualitative tool for improving questionnaires in sport science. *Measurement in Physical Education and Exercise Science, 14*(1), 51-60. doi:10.1080/10913670903455025

Dillman, D. A. (2007). *Mail and internet surveys: The tailored design method* (2nd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

*DiPerna, J. C., & Elliott, S. N. (1999). Development and validation of the Academic Competence Evaluation Scales. *Journal of Psychoeducational Assessment, 17*(3), 207-225. doi:10.1177/073428299901700302

Dirks, M. A., De Los Reyes, A., Briggs-Gowan, M., Cella, D., & Wakschlag, L. S. (2012). Annual research review: Embracing not erasing contextual variability in children's behavior - theory and utility in the selection and use of methods and informants in developmental psychopathology. *The Journal of Child Psychology and Psychiatry, 53*(5), 558-574. doi:10.1111/j.1469-7610.2012.02537.x

DiStefano, C., & Hess, B. (2005). Using confirmatory factor analysis for construct validation: An empirical review. *Journal of Psychoeducational Assessment, 23*(3), 225-241. doi:10.1177/073428290502300303

*Doherty, J., & Connolly, M. (1985). How accurately can primary school teachers predict the scores of their pupils in standardised tests of attainment? A study of some non-cognitive factors that influence specific judgements. *Educational Studies, 11*(1), 41-60. doi:10.1080/0305569850110105

Dolnicar, S., & Grün, B. (2014). Including don't know answer options in brand image surveys improves data quality. *International Journal of Market Research, 56*(1), 33-50. doi:10.2501/IJMR-2013-043

Dombrowski, S. C. (2015). *Psychoeducational assessment and report writing*. New York, NY: Springer.

Douglas, K. M., McGarty, C., Bliuc, A.-M., & Lala, G. (2005). Understanding cyberhate: Social competition and social creativity in online white supremacist groups. *Social Science Computer Review, 23*(1), 68-76. doi:10.1177/0894439304271538

Dreary, I. J., Leaper, S. A., Murray, A. D., Staff, R. T., & Whalley, I. J. (2003). Cerebral white matter abnormalities and lifetime cognitive change: A 67-year follow-up of the Scottish mental survey of 1932. *Psychology and Ageing, 18*(1), 140-148.

Drennan, J. (2003). Cognitive interviewing: Verbal data in the design and pretesting of questionnaires. *Journal of Advanced Nursing, 42*(1), 57-63. doi:10.1046/j.1365-2648.2003.02579.x

Drudy, S. (2008). Gender balance/gender bias: The teaching profession and the impact of feminisation. *Gender and Education, 20*(4), 309-323. doi:10.1080/09540250802190156

*DuPaul, G. J., Volpe, R. J., Jitendra, A. K., Lutz, G. J., Lorah, K. S., & Gruber, R. (2004). Elementary school students with AD/HD: Predictors of academic achievement. *Journal of School Psychology, 42*(4), 285-301.

Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development*: Psychology Press.

*Eaves, R. C., Campbell-Whatley, G., Dunn, C., Reilly, A., & Tate-Braxton, C. (1994). Comparison of the Slosson Full-Range Intelligence Test and teacher judgments as predictors of students' academic achievement. *Journal of Psychoeducational Assessment, 12*(4), 381-392. doi:10.1177/073428299401200408

*Eaves, R. C., Williams, P., Winchester, K., & Darch, C. (1994). Using teacher judgment and IQ to estimate reading and mathematics achievement in a remedial-reading program. *Psychology in the Schools, 31*(4), 261-272. doi:10.1002/1520-6807(199410)31:4<261::AID-PITS2310310403>3.0.CO;2-K

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*, 109-132. doi:10.1146/annurev.psych.53.100901.135153

*Eckert, T. L., Dunn, E. K., Codding, R. S., Begeny, J. C., & Kleinmann, A. E. (2006). Assessment of mathematics and reading performance: An examination of the correspondence between direct assessment of student performance and teacher report. *Psychology in the Schools, 43*(3), 247-265.

Elliott, C. D. (2007). *Differential Ability Scales* (2nd ed.). San Antonio, TX: Harcourt Assessmnet.

Epstein, S. (1983). Aggregation and beyond: Some basic issues on the prediction of behavior. *Journal of Personality, 51*(3), 360-392. doi:10.1111/j.1467-6494.1983.tb00338.x

Erez, A., Bloom, M. C., & Wells, M. T. (1996). Using random rather than fixed effects models in meta-analysis: Implications for situational specificity and validity generalization. *Personnel Psychology, 49*(2), 275-306. doi:10.1111/j.1744-6570.1996.tb01801.x

Evans, J. J., Floyd, R. G., McGrew, K. S., & Leforgee, M. H. (2002). The relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and reading achievement during childhood and adolescence. *School Psychology Review, 31*(2), 246-262.

Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory factor analysis: Understanding statistics*. New York, NY: Oxford University Press.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*, 272-299. doi:10.1037/1082-989X.4.3.272

Fagard, R. H., Staessen, J. A., & Thijs, L. (1996). Advantages and disadvantages of the meta-analysis approach. *Journal of Hypertension, 14*(Suppl 2), S9-S13.

Falotico, M. (2015). *School psychologists' time allocation: Striving for "lean" school psychology.* (Master of Science), Miami University.

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 56-83. doi:10.1080/10705519909540119

Fearrington, J. Y., Parker, P. D., Kidder-Ashley, P., Gagnon, S. G., McCane-Bowling, S., & Sorrell, C. A. (2014). Gender differences in written expression curriculum-based measurement in third- through eight-grade students. *Psychology in the Schools, 51*(1), 85-96. doi:10.1002/pits.21733

*Feinberg, A. B., & Shapiro, E. S. (2003). Accuracy of teacher judgment in predicting oral reading fluency. *School Psychology Quarterly, 18*(1), 52-65. doi:10.1521/scpq.18.1.52.20876

*Feinberg, A. B., & Shapiro, E. S. (2009). Teacher accuracy: An examination of teacher-based judgments of students' reading with differing achievement levels. *The Journal of Educational Research, 102*(6), 453-462. doi:10.3200/JOER.102.6.453-462

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Children Development, 59*(5), 1-173.

Ferguson, R. F. (2003). Teachers' perceptions and expectations and the Black-White test score gap. *Urban Education, 38*(4), 460-507. doi:10.1177/0042085903254970

Ferrando, P. J., & Anguiano-Carrasco, C. (2010). A structural model-based optimal person-fit procedure for identifying faking. *Educational and Psychological Measurement, 73*(2), 173-190. doi:10.1177/0013164412460049

Ferrer, E., O’Hare, E. D., & Bunge, S. A. (2009). Fluid reasoning and the developing brain. *Frontiers in Neuroscience, 3*(1), 46-51. doi:10.3389/neuro.01.003.2009

Festinger, L. (1954). A theory of social comparison processes. *Human Relations, 7*, 117-140. doi:10.1177/001872675400700202

Field, A. P. (2001). Meta-analysis of correlation coefficients: A Monte-Carlo comparison of fixed- and random-effects methods. *Psychological Methods, 6*(2), 161-180. doi:10.1037/1082-989X.6.2.161

Fiorello, C. A., Hale, J. B., Holdnack, J. A., Kavanagh, J. A., Terrell, J., & Long, L. (2007). Interpreting intelligence test results for children with disabilities: Is global intelligence relevant? *Applied Neuropsychology, 14*(1), 2-12. doi:10.1080/09084280701280338

Fiorello, C. A., Hale, J. B., McGrath, M., Ryan, K., & Quinn, S. (2002). IQ interpretation for children with flat and variable test profiles. *Learning and Individual Differences, 13*, 115-125. doi:10.1016/S1041-6080(02)00075-4

Fiorello, C. A., Hale, J. B., & Wycoff, K. L. (2012). Cognitive hypothesis testing: Linking test results to the real world. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 484-496). New York, NY: Guilford Press.

Fiorello, C. A., Thurman, S. K., Zavertnik, J., Sher, R., & Coleman, S. (2009). A comparison of teachers' and school psychologists' perceptions of the importance of CHC abilities in the classroom. *Psychology in the Schools, 46*(6), 489-500. doi:10.1002/pits.20392

Fiske, D. W., & Butler, J. M. (1963). The experimental conditions for measuring individual differences. *Educational and Psychological Measurement, 23*(2), 249-266. doi:10.1177/001316446302300203

Flanagan, D. P., Alfonso, V. C., & Reynolds, M. R. (2013). Broad and narrow CHC abilities measured and not measured by the Wechsler scales: Moving beyond within-battery factor analysis. *Journal of Psychoeducational Assessment, 31*(2), 202-223. doi:10.1177/0734282913478047

Flanagan, D. P., & Dixon, S. G. (2014). The Cattell-Horn-Carroll theory of cognitive abilities. In C. R. Reynolds, K. J. Vannest, & E. Fletcher-Janzen (Eds.), *Encyclopedia of special education: A reference for the education of children, adolescents, and adults with disabilities and other exceptional individuals* (4th ed., pp. 368–382). Hoboken, NJ: John Wiley & Sons.

Flanagan, D. P., Fiorello, C. A., & Ortiz, S. O. (2010). Enhancing practice through application of Cattell-Horn-Carroll theory and research: A "third method" approach to specific learning disability identification. *Psychology in the Schools, 47*(7), 739-760. doi:10.1002/pits.20501

Flanagan, D. P., & Kaufman, A. S. (2004). *Essentials of WISC-IV assessment*. Hoboken, NJ: Joh Wiley & Sons.

Flanagan, D. P., & McGrew, K. S. (1997). A cross-battery approach to assessing and interpreting cognitive abilities: Narrowing the gap between practice and cognitive science. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (1st ed., pp. 314-325). New York, NY: Guilford Press.

Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The Wechsler intelligence scales and Gf-Gc theory: A contemporary approach to interpretation*. Needham Height, MA: Allyn & Bacon.

Flanagan, D. P., & Ortiz, S. O. (2001). *Essentials of cross-battery assessment*. New York, NY: John Wiley & Sons.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2007). *Essentials of cross-battery assessment* (2nd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Massachusetts: John Wiley & Sons.

Flanagan, D. P., Ortiz, S. O., Alfonso, V. C., & Mascolo, J. T. (2006). *The achievement test desk reference (ADTR): A guide to learning disability identification*. Boston, MA: Allyn & Bacon.

Fletcher, J. M. (1994). Cognitive profiles of reading disability: Comparisons of discrepancy and low achievement definitions. *Journal of Educational Psychology, 86*(1), 6-23. doi:10.1037/0022-0663.86.1.6

Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment, 7*(3), 286-299. doi:10.1037/1040-3590.7.3.286

Floyd, R. G., Evans, J. J., & McGrew, K. S. (2003). Relations between measures of Cattell-Horn-Carroll cognitive abilities and mathematics achievement across the school-age years. *Psychology in the Schools, 40*(2), 155-171. doi:10.1002/pits.10083

Framingham, J. (2016). What is psychological assessment? *Psych Central.* Retrieved from https://psychcentral.com/lib/what-is-psychological-assessment/

Freund, P. A., & Kasten, N. (2012). How smart do you think you are? A meta-analysis on the validity of self-estimates of cognitive ability. *Psychological Bulletin*. doi:10.1037/a0026556

Frome, P. M., & Eccles, J. S. (1998). Parents' influence on children's achievement-related perceptions. *Journal of Personality and Social Psychology, 74*(2), 435-452. doi:10.1037/0022-3514.74.2.435

Furnham, A. (1986). Response bias, social desirability and dissimulation. *Personality and Individual Differences, 7*(3), 386-400. doi:10.1016/0191-8869(86)90014-0

Furnham, A. (2000). Parents' estimates of their own and children's multiple intelligences. *British Journal of Developmental Psychology, 18*, 583-594. doi:10.1348/026151000165869

Furnham, A., & Budhani, S. (2002). Sex differences in the estimated intelligence of school children. *European Journal of Personality, 16*(3), 201-219. doi:10.1002/per.438

*Furnham, A., & Bunclark, K. (2006). Sex differences in parents' estimations of their own and their children's intelligence. *Intelligence, 34*(1), 1-14. doi:10.1016/j.intell.2005.05.005

Furnham, A., & Fukumoto, S. (2008). Japanese parents' estimates of their own and their children's multiple intelligences: Cultural modesty and moderate differentiation. *Japanese Psychological Research, 50*(2), 63-76. doi:10.1111./j.1468-5884.2008.00362.x

Furnham, A., & Gasson, L. (1998). Sex differences in parental estimates of their children's intelligence. *Sex Roles, 38*(1/2), 151-162. doi:10.1023/A:1018772830511

Furnham, A., & Mansi, A. (2014). The self-assessment of the Cattell-Horn-Carroll broad stratum abilities. *Learning and Individual Differences, 32*, 233-237.

Furnham, A., & Petrides, K. V. (2004). Parental estimates of five types of intelligence. *Australian Journal of Psychology, 56*(1), 10-17. doi:10.1080/00049530410001688074

Furnham, A., Rakow, T., & Mak, T. (2002). The determinants of parents' beliefs about the intelligence of their children: A study from Hong Kong. *International Journal of Psychology, 37*(6), 343-352.

Furnham, A., Reeves, E., & Budhani, S. (2002). Parents think their sons are brighter than their daughters: Sex differences in parental self-estimations and estimations of their children's multiple intelligences. *The Journal of Genetic Psychology, 163*(1), 24-39. doi:10.1080/00221320209597966

Furnham, A., & Valgeirsson, H. (2007). Parents' estimations of their own intelligence and that of their children: A comparison between English and Icelandic parents. *Scandinavian Journal of Psychology, 48*, 289-298. doi:10.1111/j.1467-9450.2007.00587.x

*Gabriele, A. J., Joram, E., & Park, K. H. (2016). Elementary mathematics teachers' judgment accuracy and calibration accuracy: Do they predict students' mathematics achievement outcomes? *Learning and Instruction, 45*, 49-60. doi:10.1016/j.learninstruc.2016.06.008

Galvin, J. E., Roe, C. M., Xiong, C., & Morris, J. C. (2006). Validity and reliability of the AD8 informant interview in dementia. *Neurology, 67*, 1942-1948.

Ganley, C., & Lubienski, S. (2016). Current research on gender differences in math. *Teaching Children Mathematics.* Retrieved from https://www.nctm.org/Publications/Teaching-Children-Mathematics/Blog/Current-Research-on-Gender-Differences-in-Math/

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York, NY: Basic Book Inc.

Gardner, H., & Hatch, T. (1989). Multiple Intelligences go to school: Educational implications of the theory of multiple intelligences. *Educational Researcher, 18*(8), 4-10. doi:10.2307/1176460

Geiser, C., Mandelman, S. D., Tan, M., & Grigorenko, E. L. (2016). Multitrait-multimethod assessment of giftedness: An application of the correlated traits-correlated (methods-1) model. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(1), 76-90. doi:10.1080/10705511.2014.937792

George, D., & Mallery, P. (2010). *SPSS for Windows step by step: A simple guide and reference 17.0* (10th ed.). Boston, MA: Pearson.

Gerbing, D. W., & Anderson, J. C. (1984). On the meaning of within-factor correlated measurement errors. *Journal of Consumer Research, 11*(1), 572-580.

Gilman, L. (2001). The theory of multiple intelligences. *Human Intelligence.* Retrieved from https://www.intelltheory.com/mitheory.shtml

Gioia, G. A., Isquith, P. K., Guy, S. C., & Kenworthy, L. (2000). *Behavior Rating Inventory of Executive Function*. Odessa, FL: Psychological Assessment Resources.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*(10), 3-8.

*Goforth, K., Noltemeyer, A., Patton, J., Bush, K. R., & Bergen, D. (2014). Understanding mathematics achievement: An analysis of the effects of student and family factors. *Educational Studies, 40*(2), 196-214. doi:10.1080/03055698.2013.866890

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48*(1), 26-34. doi:10.1037/0003-066X.48.1.26

Goldin, R. L., Matson, J. L., Konst, M. J., & Adams, H. L. (2014). A comparison of children and adolescents with ASD, atypical development, and typical development on the Behavior Assessment System for Children, Second Edition (BASC-2). *Research in Autism Spectrum Disorders, 8*(8), 951-957. doi:10.1016/j.rasd.2014.04.005

Gorusch, R. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Gothard, A. (2013). 9 key things to know about psychoeducational evaluations. Retrieved from https://impactadhd.com/overcome-school-challenges/9-key-things-to-know-about-psychoeducational-evaluations/

Gottfredson, L. S. (1997). Why g matters: The complexities of everyday life. *Intelligence, 24*(1), 79-132.

Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability. *Educational and Psychological Measurement, 66*(6), 930-944. doi:10.1177/0013164406288165

Granello, D. H., & Wheaton, J. E. (2004). Online data collection: Strategies for research. *Journal of Counselling & Development, 82*(4), 387-393. doi:10.1002/j.1556-6678.2004.tb00325.x

Groth-Marnat, G. (1999). Financial efficacy of clinical assessment: Rational guidelines and issues for future research. *Journal of Clinical Psychology, 55*(7), 813-824. doi:10.1002/(SICI)1097-4679(199907)55:7<813::AID-JCLP4>3.0.CO;2-B

Groth-Marnat, G., & Horvath, L. S. (2005). The psychological report: A review of current controversies. *Journal of Clinical Psychology, 62*(1), 73-81. doi:10.1002/jclp.20201

Gustafsson, J.-E., & Undheim, J. O. (1996). Individual differences in cognitive functions. In D. C. Berliner & R. Calfee (Eds.), *Handbook of educational psychology* (pp. 186-242). New York, NY: Macmillan.

*Gut, J., Reimann, G., & Alexander, G. (2013). A contextualised view on long-term predictors of academic performance. *Journal of Educational Psychology, 105*(2), 436-443. doi:10.1037/a0031503

Haidich, A. B. (2010). Meta-analysis in medical research. *Hippokratia, 14*(Suppl 1), 29-37.

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1995). *Multivariate data analysis* (4th ed.). Upper Saddle River, NJ: Prentice-Hall, Inc.

Hall, J. A., & Rosenthal, R. (1991). Testing for moderator variables in meta-analysis: Issues and methods. *Communication Monographs, 58*(4), 437-448.

Halpern, D. F. (2000). *Sex differences in cognitive abilities*. Mahwah, NJ: Lawrence Erlbaum Associates.

Halpern, D. F., & Collaer, M. L. (2005). Sex differences in spatial abilities: More than meets the eye. In P. Shah & A. Miyake (Eds.), *The Cambridge handbook of visuospatial thinking* (pp. 170-212). New York, NY: Cambridge University Press.

Hammill, M. D., & Bryant, B. R. (1998). *Learning Disabilities Diagnostic Inventory*. Austin, TX: PRO-ED.

Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present und Future-A Festschrift in honour of Karl Joreskog*. Lincolnwood, IL: Scientific Software International.

Harris, G. E., & Joy, R., M. (2010). Educational psychologists' perspectives on their professional practice in Newfoundland and Labrador. *Canadian Journal of School Psychology, 25*(2), 205-220. doi:10.1177/0829573510366726

Hartley, B. L., & Sutton, R. M. (2013). A stereotype threat account of boys' academic underachievement. *Child Development, 84*(5), 1716-1733. doi:10.1111/cdev.12079

Harvey, V. S. (1997). Improving readability of psychological reports. *Professional Psychology: Research and Practice, 28*(3), 271-274.

Harvey, V. S. (2006). Variables affecting the clarity of psychological reports. *Journal of Clinical Psychology, 62*(1), 5-18. doi:10.1002/jclp.20196

*Hauser-Cram, P., Sirin, S. R., & Stipek, D. (2003). When teachers' and parents' values differ: Teachers' ratings of academic competence in children from low-income families. *Journal of Educational Psychology, 95*(4), 813-820. doi:10.1037/0022-0663.95.4.813

Hay, D. F., Pawlby, S., Sharp, D., Schmucker, G., Alice, M., Allen, H., & Kumar, R. (1999). Parents' judgements about young children's problems: Why mothers and fathers might disagree yet still predict later outcomes. *The Journal of Child Psychology and Psychiatry, 40*(8), 1249-1258. doi:10.1111/1469-7610.00541

Hayduk, L., Cummings, G. G., Boadu, K., Pazderka-Robinson, H., & Boulianne, S. (2007). Testing! testing! one, two, three – Testing the theory in structural equation models! *Personality and Individual Differences, 42*(5), 841-850. doi:10.1016/j.paid.2006.10.001

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment, 7*(3), 238-247.

Hayton, J. C., Allen, D. G., & Scarpello, V. (2004). Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational Research Methods, 7*(2), 191-205.

Hedges, L. V. (1992). Meta-analysis. *Journal of Educational and Behavioral Statistics, 17*(4), 279-296. doi:10.3102/10769986017004279

Hedges, L. V., & Becker, B. J. (1986). Statistical methods in the meta-analysis of research on gender differences. In J. S. Hyde & M. C. Linn (Eds.), *The psychology of gender: Progress through meta-analysis*. Baltimore: The Johns Hopkins University Press.

Hedges, L. V., & Olkin, I. (1980). Vote-counting methods in research synthesis. *Psychological Bulletin, 88*(2), 359-369. doi:10.1037/0033-2909.88.2.359

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3*(4), 486-504. doi:10.1037/1082-989X.3.4.486

Hedges, R., Drysdale, K., & Levick, W. R. (2015). The Children's Memory Questionnaire-Revised. *Applied Neuropsychology, 4*(4), 285-296. doi:10.1080/21622965.2014.925806

Helmke, A., & Schrader, F.-W. (1987). Interactional effects of instructional quality and teacher judgement accuracy on achievement. *Teacher and Teacher Education, 3*, 91-98. doi:10.1016/0742-051X(87)90010-2

*Helwig, R., Anderson, L., & Tindal, G. (2001). Influence of elementary student gender on teacher's perceptions of mathematics achievement. *The Journal of Educational Research, 95*(2), 93-102. doi:10.1080/00220670109596577

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement, 66*(3), 393-416. doi:10.1177/0013164405282485

*Herbert, J., & Stipek, D. (2005). The emergence of gender differences in children's perceptions of their academic competence. *Journal of Applied Developmental Psychology, 26*(3), 276-295. doi:doi:10.1016/j.appdev.2005.02.007

Hermida, R. (2015). The problem of allowing correlated errors in structural equation modeling: Concerns and considerations. *Computational Methods in Social Sciences, 3*(1), 5-17.

*Hernández-Torrano, D., Ferrándiz, C., Ferrando, M., & Prieto, L. (2014). The theory of multiple intelligences in the identification of high-ability students. *Anales De Psicología, 30*(1), 192-200. doi:10.6018/analesps.30.1.148271

Higgins, J. P. T. (2008). Commentary: Heterogeneity in meta-analysis should be expected and appropriately quantified. *International Journal of Epidemiology, 37*(5), 158–1160. doi:10.1093/ije/dyn204

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*(11), 1539-1558. doi:Higgins and Thompson (2002)

Hilton, A., & Skrutkowski, M. (2002). Translating instruments into other langauges: Development and testing processes. *Cancer Nursing, 25*(1), 1-7.

Hinkin, T. R., Tracey, J. B., & Enz, C. A. (1997). Scale construction: Developing reliable and valid measurement instruments. *Journal of Hospitality & Tourism Research, 21*(1), 100-120. doi:10.1177/109634809702100108

Hinnant, J. B., O'Brien, M., & Ghazarian, S. R. (2009). The longitudinal relations of teacher expectations to achievement in the early school year. *Journal of Educational Psychology, 101*(3), 662-670. doi:10.1037/a0014306

*Hodge, K. A., & Kemp, C. R. (2006). Recognition of giftedness in the early years of school: Perspectives of teachers, parents, and children. *Journal for the Education of the Gifted, 30*(2), 164-204.

Hoeppner, B. B., Kelly, J. F., Urbanoski, K. A., & Slaymaker, V. (2011). Comparative utility of a single-item vs. multiple-item measure of self-efficacy in predicting relapse among young adults. *Journal of Substance Abuse Treatment, 41*(3), 305-312. doi:10.1016/j.jsat.2011.04.005

Hofer, S. I. (2015). Studying gender bias in physics grading: The role of teaching experience and country. *International Journal of Science Education, 37*(17), 2879-2905. doi:10.1080/09500693.2015.1114190

Hofer, S. M., & Clouston, S. (2014). On the importance of early-life cognitive abilities in shaping later-life outcomes. *Research in Human Development, 11*(3), 21-246. doi:10.1080/15427609.2014.936173

*Hoge, R. D., & Butcher, R. (1984). Analysis of teacher judgments of pupil achievement levels. *Journal of Educational Psychology, 76*(5), 777-781.

Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research, 59*(3), 297-313.

Holmes-Smith, P. (2014). Structural equation modeling (using AMOS): From the fundamentals to advanced topics (pp. 14.11-14.18). Melbourne, VIC: School Research Evaluation and Measurement Services.

Holmes-Smith, P., & Rowe, K. J. (1994). *The development and use of congeneric measurement models in school effectiveness research: Improving the reliability and validity of composite and latent variables for fitting multilevel and structural equation models*. Paper presented at the International Congress for School Effectiveness and Improvement, Melbourne, VIC.

Hooper, D. (2012). Exploratory factor analysis. In H. Chen (Ed.), *Approaches to quantitative research - Theory and its practical application: A guide to dissertation students*. Cork, Ireland: Oak Tree Press.

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods, 6*(1), 53-60.

*Hopkins, K. D., George, C. A., & Williams, D. D. (1985). The concurrent validity of standardised achievement tests by content area using teachers' ratings as criteria.

Journal of Educational Measurement, 22(3), 177-192. doi:10.1111/j.1745-3984.1985.tb01056.x

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*(2), 179-185.

Horn, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review, 75*(3), 242-259.

Horn, J. L. (1998). A basis for research on age differences in cognitive abilities. In J. McArdle & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 57-91). Mahwah, NJ: Lawrence Erlbaum Associates.

Horn, J. L., & Cattell, R. B. (1982). Whimsy and misunderstanding of Gf-Gc theory: A comment on Gilford. *Psychological Bulletin, 91*(623-633). doi:10.1037/0033-2909.91.3.623

Horn, J. L., & Hofer, S. M. (1992). Major abilities and development in the adult period. In R. J. Sternberg & C. A. Berg (Eds.), *Intellectual development* (pp. 44-99). New York, NY: Cambridge University Press.

Hox, J. J., & Bechger, T. M. (1998). An introduction to structural equation modeling. *Family Science Review, 11*, 354-373.

Hoyt, W. T., Warbasse, R. E., & Chu, E. Y. (2006). Construct validation in counseling psychology research. *The Counselling Psychologist, 34*(6), 769-805. doi:10.1177/0011000006287389

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1-55). doi:10.1080/10705519909540118

Huedo-Medina, T., Sanchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychological Methods, 11*, 193-206.

*Hughes, J. N., Gleason, K. A., & Zhang, D. (2005). Relationship influences on teachers' perceptions of academic competence in academically at-risk minority and majority first grade students. *Journal of School Psychology, 43*, 303-320. doi:10.1016/j.jsp.2005.07.001

Hughes, K. A. (2004). *Comparing pretesing methods: Cognitive interviews, respondent debriefing, and behavior coding.* Paper presented at the Annual Meeting of the Federal Committee on Statistical Methodology, Arlington, VA.

Humphreys, L. G., & Ilgen, D. R. (1969). Note on a criterion for the number of common factors. *Educational and Psychological Measurement, 29*(3), 571-578. doi:10.1177/001316446902900303

Humphreys, L. G., & Montanelli Jr, R. G. (1975). An investigation of the parallel analysis criterion for determining the number of common factors. *Multivariate Behavioral REsearch, 10*(2), 193-205. doi:10.1207/s15327906mbr1002_5

Hunt, E. (1995). The role of intelligence in modern society. *American Scientist, 83*(4), 356-368.

Hunt, J. M., & Paraskevopoulos, J. (1980). Children's psychological development as a function of the inaccuracy of their mothers' knowledge of their abilities. *The Journal of Genetic Psychology, 136*(2), 285-298.

Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta-analysis models: Implication for cumulative research knowledge. *International Journal of Selection and Assessment, 8*(4), 275-292. doi:10.1111/1468-2389.00156

Hyde, J. S. (2005). The gender similarities. *American Psychologist, 60*(6), 581-592. doi:10.1037/0003-066X.60.6.581

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*(2), 139-155. doi:10.1037/0033-2909.107.2.139

Ilieva, J., Baron, S., & Healey, H. M. (2002). Online surveys in marketing research: Pros and cons. *International Journal of Market Research, 44*(3), 362-380. doi:10.1177/147078530204400303

*Imbrosciano, A., & Berlach, R. (2003). Teacher perceptions of the relationships between intelligence, student behaviour, and academic performance. *Issues in Educational Research, 13*(1), 1-12.

Isiksal, M., & Cakiroghi, E. (2008). Gender diferences regarding mathematics achievement: The case of Turkish middle school students. *School Science and Mathematics, 108*(3), 113-120. doi:10.1111/j.1949-8594.2008.tb17814.x

Izquierdo, I., Olea, J., & Abad, F. J. (2014). Exploratory factor analyis in validation studies: Uses and recommendations. *Psicothema, 26*(3), 395-400. doi:10.7334/psicothema2013.349

Jacobs, J. C. (1971). Effectiveness of teacher and parent identification of gifted children as a function of school level. *Psychology in the Schools, 8*(2), 140-142.

Jacobs, K. E. (2012). *Validly measuring cognitive ability by self-report: Relevance of response format and type of ability assessed.* (Master of Psychology (Educational & Developmental)/Doctor of Philosophy), Monash University.

Jacobs, K. E., Watt, D., & Roodenburg, J. (2013). Why can't Jonny read? Bringing theory into assessment. *InPsych: The Bulletin of the Australian Psychological Society Ltd, 35*(6), 16.

Jansen, M., Bodden, D. H. M., Muris, P., van Doorn, M., & Granic, I. (2017). Measuring anxiety in children: The importance of separate mother and father reports. *Child & Youth Care Forum, 46*(5), 643-659. doi:10.1007/s10566-017-9402-5

Jensen, A. R. (1981). *Straight talk about mental tests*. New York: Free Press.

Jensen, A. R. (1998). *The g factor: The science of mental ability*. New York, NY: Praeger Publishers.

*Johansson, S., Myrberg, E., & Rosén, M. (2012). Teachers and tests: Assessing pupils' reading achievement in primary schools. *Educational Research and Evaluation, 18*(8), 693-711. doi:10.1080/13803611.2012.718491

Johnson, G. B. J. (1953). Factors to be considered in the interpretation of intelligence-test scores. *The Elementary School Journal, 54*(3), 145-150.

*Johnson, S., Marlow, N., & Wolke, D. (2012). Assessing educational outcomes in middle childhood: Validation of the Teacher Academic Attainment Scale. *Developmental Medicine & Child Neurology, 54*(6), 544-551. doi:10.1111/j.1469-8749.2012.04264.x

Johnston, M., Dixon, D., Hart, J., Glidewell, L., Schröder, C., & Pollard, B. (2014). Discriminant content validity: A quantitative methodology for assessing content of theory-based measures, with illustrative applications. *British Journal of Health Psychology, 19*, 240-257. doi:10.1111/bjhp.12095

Jones, S., Murphy, F., Edwards, M., & James, J. (2008). Doing things differently: Advantages and disadvantages of web questionnaires. *Nurse Researcher, 15*(4), 15-26.

Jussim, L. (1989). Teacher expectations: Self-fulfilling prophecies, perceptual biases, and accuracy. *Journal of Personality and Social Psychology, 57*(3), 469-480. doi:10.1037/0022-3514.57.3.469

Jussim, L., & Eccles, J. S. (1992). Teacher expectations II: Construction and reflection of student achievement. *Journal of Personality and Social Psychology, 63*(6), 947-961. doi:10.1037/0022-3514.63.6.947

Kail, R. V., & Hall, L. K. (1994). Processing speed, naming speed, and reading. *Developmental Psychology, 30*(6), 949-954. doi:10.1037/0012-1649.30.6.949

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*(1), 141-151. doi:10.1177/001316446002000116

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*(1), 31-36.

Kaplan, D. (1990). Evaluating and modifying covariance structure models: A review and recommendation. *Multivariate Behavioral REsearch, 25*(25), 137-155. doi:10.1207/s15327906mbr2502_1

Kaufman, A. S. (1979). *Intelligent testing with the WISC-R*. New York, NY: Wiley.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children, Second Edition (KABC-II)*. Circle Pines, MN: AGS.

Kaufman, A. S., & Lichtenberger, E. O. (2002). *Assessing adolescent and adult intelligence*. Boston, MA: Allyn & Bacon.

Kaufman, A. S., Raiford, S. E., & Coalson, D. L. (2016). *Intelligent Testing with the WISC-V*. Hoboken, NJ: John Wiley & Sons, Inc.

Kaufman, S. B., Reynolds, M. R., Liu, X., Kaufman, A. S., & McGrew, K. S. (2012). Are cognitive g and academic achievement g one and the same g? An exploration on the Woodcock-Johnson and Kaufman tests. *Intelligence, 40*(2), 123-138. doi:10.1016/j.intell.2012.01.009

Keith, T. Z., & Reynolds, C. R. (2012). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, test, and issues* (3rd ed., pp. 758-799). New York, NY: Guilford Press.

Keith, T. Z., & Reynolds, M. R. (2010). Cattell-Horn-Carroll abilities and cognitive tests: what we've learned from 20 years of research. *Psychology in the Schools, 47*(7), 635-650. doi:10.1002/pits.20496

Kenny, D. A., & McCoach, D. B. (2009). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 10*(3), 333-351. doi:10.1207/S15328007SEM1003_1

*Kenny, D. T., & Chekaluk, E. (1993). Early reading performance: A comparison of teacher-based and test-based assessments. *Journal of Learning Disabilities, 26*(4), 227-236.

*Kikas, E., Silinskas, G., & Soodla, P. (2015). The effects of children's reading skills and interest on teacher perceptions of children's skills and individualized support.

International Journal of Behavioral Development, 39(5), 402-412. doi:10.1177/0165025415573641

Kimura, D. (1999). *Sex and cognition*. Cambridge, MA: MIT Press.

*Kirkcaldy, B., Noack, P., Furnham, A., & Siefen, G. (2007). Parental estimates of their own and their children's intelligence. *European Psychologist, 12*(3), 173-180. doi:10.1027/1016-9040.12.3.173

Klassen, R. M., Neufeld, P., & Munro, F. (2005). When IQ is irrelevant to the definition of learning disabilities. *School Psychology International, 26*(3), 297-316. doi:10.1177/0143034305055975

Kline, P. (1994). *An easy guide to factor analysis*. New York, NY: Routledge.

Kline, P. (1998). *The new psychometrics: Science, psychology and measurement*. New York, NY: Routledge.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford Press.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.

*Krkovic, K., Greiff, S., Kupiainen, S., Vainikainen, M.-P., & Hautamäki, J. (2014). Teacher evaluation of student ability: What roles do teacher gender, student gender, and their interaction play? *Educational Research, 56*(2), 244-257. doi:10.1080/00131881.2014.898909

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: Implications for graduate student selection and performance. *Psychological Bulletin, 127*(1), 162-181. doi:10.1037/0033-2909.127.1.162

Kunda, Z., & Spencer, S. J. (2003). When do stereotypes come to mind and when do they colour judgment? A goal-based theoretical framework for stereotype activation and application. *Psychological Bulletin, 129*(4), 522-544. doi:10.1037/0033-2909.129.4.522

Lamb, G. D. (2008). *Ratings of everyday academic and cognitive skills in evaluation of school learning and learning problems: Initial scale development and validation.* (Doctor of Philosophy), Texas A&M University.

Lance, C. E., Dawson, B., Birkelbac, D., & Hoffman, B. J. (2010). Method effects, measurement error, and substantive conclusions. *Organizational Research Methods, 13*(3), 435-455. doi:10.1177/1094428109352528

Landau, B., Smith, L., & Jones, S. (1998). Object perception and object naming in early development. *Trends in Cognitive Sciences, 2*(1), 19-24. doi:10.1016/S1364-6613(97)01111-X

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

Landy, F. (1986). Stamp collecting versus science: Validation as hypothesis testing. *American Psychologist, 41*(11), 1183-1192.

Lavy, V., & Sand, E. (2015). *On the origins of gender human capital gaps: Short and long term consequences of teachers' stereotypical beliefs*. Retrieved from http://www.nber.org/papers/w20909

Lefever, S., Dal, M., & Matthíasdóttir, Á. (2007). Online data collection in academic research: Advantages and limitations. *British Journal of Educational Technology, 38*(4), 574-582. doi:10.1111/j.1467-8535.2006.00638.x

Li, H., Pfeiffer, S. I., Petscher, Y., Kumtepe, A. T., & Mo, G. (2008). Validation of the Gifted Rating Scales-School Form in China. *Gifted Child Quarterly, 52*(2), 160-169. doi:10.1177/0016986208315802

*Limbos, M. M., & Geva, E. (2001). Accuracy of teacher assessments of second-language students at risk for learning disability. *Journal of Learning Disabilities, 34*(2), 136-151. doi:10.1177/002221940103400204

Lingard, H., & Rowlinson, S. (2006). Letter to the editor. *Construction Management and Economics, 24*(11), 1107-1109. doi:10.1080/01446190601001620

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications, Inc.

Livingston, R. B., Jennings, E., Reynolds, C. R., & Gray, R. M. (2003). Multivariate analyses of the profile stability of intelligence tests: High for IQs, low to very low for subtest analyses. *Archives of Clinical Neuropsychology, 18*(5), 487-507. doi:10.1016/S0887-6177(02)00147-6

*Lockl, K., Ebert, S., & Weinert, S. (2017). Predicting school achievement from early theory of mind: Differential effects on achievement tests and teacher ratings. *Learning and Individual Differences, 53*, 93-102. doi:10.1016/j.lindif.2016.11.007

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3*(3), 635-694. doi:10.2466/pr0.1957.3.3.635

Logan, S., & Johnston, R. (2010). Investigating gender differences in reading. *Educational Review, 62*, 175-187. doi:10.1080/00131911003637006

Logsdon, A. (2014). Understanding intelligence testing for children. Retrieved from http://learningdisabilities.about.com/od/glossar1/a/intelligencetes.htm

Lohman, D. F., Gambrell, J., & Lakin, J. (2008). The commonality of extreme discrepancies in the ability profiles of academically gifted students. *Psychology Science Quarterly, 50*(2), 269-282.

Louis, B., & Lewis, M. (1992). Parental beliefs about giftedness in young children and their relation to actual ability level. *Gifted Child Quarterly, 36*(1), 27-31. doi:10.1177/001698629203600107

Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) "'General intelligence,' objectively determined and measured". *Journal of Personality and Social Psychology, 86*(1), 96-111. doi:10.1037/0022-3514.86.1.96

Luo, D., & Petrill, S. (1999). Elementary cognitive tasks and their roles in g estimates. *Intelligence, 27*(2), 157-174. doi:10.1016/S0160-2896(99)00020-3

Luria, A. R. (1970). The functional organization of the brain. *Scientific American, 222,* 66-78.

Lynn, R., & Mikk, J. (2009). National IQs predict educational attainment in math, reading, and science across 56 nations. *Intelligence, 37*(3), 305-310. doi:10.1016/j.intell.2009.01.00

Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology, 67*(3), 280-296. doi:10.1037/0021-9010.67.3.280

MacCallum, R. C. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin, 100*(1), 107-120. doi:10.1037/0033-2909.100.1.107

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin, 111*(3), 490-504. doi:10.1037/0033-2909.111.3.490

MacCallum, R. C., & Widaman, K. F. (1999). Sample size in factor analysis. *Psychological Methods, 4*(1), 84-99. doi:10.1037/1082-989X.4.1.84

MacCorquodale, K., & Meehl, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review, 55*(2), 95-107. doi:10.1037/h0056029

Machek, G. (2003). The role of standardized intelligence measures in testing for giftedness. *Human Intelligence.* Retrieved from https://www.intelltheory.com/giftednessTesting.shtml

Machts, N., Kaiser, J., Schmidt, F. T. C., & Möller, J. (2017). Accuracy of teachers' judgments of students' cognitive abilities: A meta-analysis. *Educational Research Review, 19*, 85-103. doi:10.1016/j.edurev.2016.06.003

Mackintosh, N. J. (1998). *IQ and human intelligence*. Oxford: Oxford University Press.

Mackintosh, N. J. (2011). *IQ and human intelligence* (2nd ed.). New York, NY: Oxford University Press.

*Maguin, E., & Loeber, R. (1996). How well do ratings of academic performance by mothers and their sons correspond to grades, achievement test scores, and teachers' ratings? *Journal of Behavioral Education, 6*(4), 405-425. doi:10.1007/BF02110514

Marsh, H. W., Lüdtke O, Muthén, B., Asparouhov, T., Morin, A. J., Trautwein, U., & Nagengast, B. (2010). A new look at the Big Five factor structure through exploratory structural equation modelling. *Psychological Assessment, 22*(3), 471-491. doi: 10.1037/a0019227

Martin, J. L., Pérez, V., Sacristán, M., & Álvarez, E. (2005). Is grey literature essential for a better control of publication bias in psychiatry? An example from three meta-analyses of schizophrenia. *European Psychiatry, 20*, 550-553. doi:10.1016/j.eurpsy.2005.03.011

*Martin, S. D., & Shapiro, E. S. (2011). Examining the accuracy of teachers' judgments of DIBELS performance. *Psychology in the Schools, 48*(4), 343-356. doi:10.1002/pits.20558

Mascolo, J. T., Flanagan, D. P., & Alfonso, V. C. (2014). A systematic method of analysing assessment results for tailoring interventions (SMAARTI). In J. T. Mascolo, V. C. Alfonso, & D. P. Flanagan (Eds.), *Essentials of planning, selecting and tailoring interventions for unique learners* (pp. 3-55). Hoboken, NJ: John Wiley & Sons, Inc.

*Massa, J., Gomes, H., Tartter, V., Wolfson, V., & Halperin, J., M. (2008). Concordance rates between parent and teacher clinical evaluation of language fundamentals observational rating scale. *International Journal of Language & Communication Disorders, 43*(1), 99-110.

Mastoras, S. M., Climie, E. A., McCrimmon, A. W., & Schwean, V. L. (2015). A C.L.E.A.R approach to report writing: A framework for impoving the efficacy of psychoeducational reports. *Canadian Journal of School Psychology, 26*(2), 127-147. doi:10.1177/0829573511409722

Mather, N., & Woodcock, R. W. (2001). *Examiner's manual: Woodcock-Johnson III Tests of Cognitive Abilities*. Itasca, IL: Riverside Publishing.

Matt, G. E., & Cook, T. D. (2009). Threats to the validity of generalized inferences. In H. M. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis*. New York, NY: The Russell Sage Foundation.

McCandless, S., & O'Laughlin, L. (2007). The clinical utility of the Behavior Rating Inventory of Executive Function (BRIEF) in the Diagnosis of ADHD. *Journal of Attention Disorders, 10*(4), 381-389. doi:10.1177/1087054706292115

McCoach, D. B., Gable, R. K., & Madura, J. P. (2013). *Instrument development in the affective domain: School and corporate applications* (3rd ed.). New York, NY: Springer.

McCoach, D. B., & Newton, S. D. (2016). Confirmatory factor analysis. In D. Wyse, N. Selwyn, E. Smith, & L. E. Suter (Eds.), *The BERA/SAGE handbook of educational research* (Vol. 2, pp. 851-872). London, UK: SAGE Publications Ltd.

McDonagh, M., Peterson, K., Raina, P., Chang, S., & Shekelle, P. (2013). *Avoiding bias in selecting studies* (AHRQ Publication No. 13-EHC045-EF). Retrieved from Agency for Healthcare Research and Quality: www.effectivehealthcare.ahrq.gov/reports/final.cfm.

McGarty, C., Yzerbyt, V. Y., & Spears, R. (2002). Social, cultural and cognitive factors in stereotype formation. In C. McGarty, V. Y. Yzerbyt, & R. Spears (Eds.), *Stereotypes* *as explanations: The formation of meaningful beliefs about social groups* (pp. 1-15). Cambridge, UK: Cambridge University Press.

McGrath, K. F., & Van Bergen, P. (2017). Male teachers are an endangered species in Australia: New research. Retrieved from https://theconversation.com/male-teachers-are-an-endangered-species-in-australia-new-research-83464

McGrew, K. S. (1993). The relationship between the Woodcock-Johnson Psychoeducational Battery-Revised Gf-Gc cognitive clusters and reading achievement across the life span. *Journal of Psychoeducational Assessment, WJR Monograph*, 39 – 53.

McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (2nd ed., pp. 136-181). New York: Guildford Press.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*, 1-10. doi:10.1016/j.intell.2008.08.004

McGrew, K. S. (2016, January 28). "Intelligent" intelligence testing with the W IV Tests of Cognitive Ability #3: Within-CHC assessment trees - a Gf "tease" [Blog post]. Retrieved from http://www.iqscorner.com/2016/01/intelligent-intelligence-testing-with-w.html

McGrew, K. S., & Flanagan, D. P. (1998). *The intelligence test desk reference (ITDR): Gf-Gc cross-battery assessment*. Boston, MA: Allyn & Bacon.

McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). *Technical manual: Woodcock-Johnson IV*. Rolling Meadows, IL: Riverside.

McGrew, K. S., & Wendling, B. J. (2010). Cattell–Horn–Carroll cognitive-achievement relations: What we have learned from the past 20 years of research. *Psychology in the Schools, 47*(7), 651-675. doi:10.1002/pits.20497

*Meissel, K., Meyer, F., Yao, E. S., & Rubie-Davies, C. M. (2017). Subjectivity of teacher judgments: Exploring student characteristics that influence teacher judgments of student ability. *Teaching and Teacher Education, 65*, 48-60. doi:10.1016/j.tate.2017.02.021

Mervielde, I., & De Fruyt, F. (1999). Construction of the Hierarchical Personality Inventory for Children (HiPIC). In I. Mervielde, I. Dreary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe. Proceedings of the Eight European Conference on Personality Psychology* (pp. 107-127). Tilburg: Tilburg University Press.

Mervielde, I., & De Fruyt, F. (2002). Assessing children's traits with the Hierarchical Personality Inventory for Children. In B. De Raad & M. Perugini (Eds.), *Big Five Assessment* (pp. 129-146). Washington, DC: Hogrefe & Huber.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749.

*Miller, S. A. (1986). Parents' beliefs about their children's cognitive abilities. *Developmental Psychology, 22*(2), 276-284. doi:10.1037/0012-1649.22.2.276

Miller, S. A. (1988). Parents' beliefs about children's cognitive development. *Child Development, 59*(2), 259-285. doi:10.2307/1130311

*Miller, S. A., & Davis, T. L. (1992). Beliefs about children: A comparative study of mothers, teachers, peers, and self. *Child Development, 63*(5), 1251-1265. doi:10.1111/j.1467-8624.1992.tb01693.x

*Miller, S. A., Manhal, M., & Mee, L. L. (1991). Parental beliefs, parental accuracy, and children's cognitive performance: A search for causal relations. *Developmental Psychology, 27*(2), 267-276. doi:10.1037/0012-1649.27.2.267

Millman, J., & Greene, J. (1993). The specification and development of tests of achievement and ability. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed.). Phoenix, AZ: The Oryx Press.

Moore, D. S., Notz, W. I., & Flinger, M. A. (2013). *The basic practice of statistics* (6th ed.). New York, NY: W. H. Freeman and Company.

Morgado, F. F. R., Meireles, J. F. F., Neves, C. M., Amaral, A. C. S., & Ferreira, M. E. C. (2017). Scale dvelopment: Ten main limitations and recommendations to improve future research practices. *Psicologia: Reflexão e Crítica, 30*(3). doi:10.1186/s41155-016-0057-1

Mulholland, L. A., & Berliner, D. C. (1992). *Teacher experience and the estimation of student achievement*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Mulrow, C. D. (1994). Systematic reviews: Rationale for systematic reviews. *British Medical Journal, 309*(6954), 597-599. doi:10.1136/bmj.309.6954.597

Munck, I. M. E. (1979). *Model building in comparative education: applications of the LISREL method to cross-national survey data*. Stockholm: Almqvist & Wiksell.

Mungas, D., Heaton, R., Tulsky, D., Zelazo, P., Slotkin, J., Blitz, D., . . . Gershon, R. (2014). Factor structure, convergent validity, and discriminant validity of the NIH Toolbox Cognitive Health Battery (NIHTB-CHB) in adults. *Journal of the International Neuropsychological Society, 20*(6), 579-587. doi:10.1017/S1355617714000307

Murphy, K. R., & Davidshofer, C. O. (1994). *Psychological testing: Principles and applications* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Naglieri, J. A., & Bornstein, B. T. (2003). Intelligence and achievement: Just how correlated are they? *Journal of Psychoeducational Assessment, 21*(3), 244-260. doi:10.1177/073428290302100302

Naglieri, J. A., & Reardon, S. M. (1993). Traditional IQ is irrelevant to learning disabilities - intelligence is not. *Journal of Learning Disabilities, 26*(2), 127-133. doi:10.1177/002221949302600205

Nastasi, B. K., Varjas, K., Bernstein, R., & Pluymert, K. (1998). Mental health programming and the role of school psychologists. *School Psychology Review, 27*(2), 217-233.

Neisser, U., Boodoo, G., Bouchard, T. J. J., Boykin, A. W., Brody, N., Ceci, S. J., . . . Urbina, S. (1996). Intelligence: Known and unknowns. *American Psychologist, 51*(2), 77-101. doi:10.1037/0003-066X.51.2.77

Neugebauer, S. R., Kieffer, M. J., & Howard, E. R. (2015). Exploring multidimensionality and mediation in the roles of lexical knowledge in reading comprehension for spanish-speaking language minority learners. *Learning and Individual Differences, 39*, 24-38. doi:10.1016/j.lindif.2015.03.003

Newton, J. H., & McGrew, K. S. (2010). Introduction to the special issue: Current research in Cattell-Horn-Carroll based asssessment. *Psychology in the Schools, 47*(7), 621-634. doi:10.1002/pits.20495

Nguyen, H. H., & Ryan, A. M. (2008). Does stereotype threat affect test performance of minorities and women? A meta-analysis of experimental evidence. *Journal of Applied Psychology, 93*(6), 1314-1334. doi:10.1037/a0012702

Noland, R. M. (2017). Intelligence testing usnig a tablet computer: Experiences with using Q-Interactive. *Training and Education in Professional Psychology, 11*(3), 156-163. doi:10.1037/tep0000149

Norris, S. L., Holmer, H. K., Ogden, L. A., Fu, R., Abou-Setta, A. M., Viswanathan, M. S., & McPheeters, M. L. (2012). *Selective outcome reporting as a source of bias in reviews of comparative effectiveness* (AHRQ Publication No. 12-EHC110-EF). Retrieved from Agency for Healthcare Research and Quality: https://www.ncbi.nlm.nih.gov/books/NBK100613/pdf/Bookshelf_NBK100613.pdf

Nunnally, J. C. (1967). *Psychometric theory*. New York, NY: McGraw-Hill.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.

Nunnally, J. C., & Bernstein, I. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.

O'Conner, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, & Computers, 32*(3), 396-402.

Ockey, G. J. (2014). Exploratory factor analysis and structural equation modeling. In A. J. Kunnan (Ed.), *The companion to language assessment* (1st ed., Vol. 3): John Wiley & Sons, Inc.

Ortiz, S. O., & Lella, S. A. (2004). Intellectual ability and assessment: A primer for parents and eduators. In A. Canter, L. Paige, & S. Shaw (Eds.), *Helping children at home and school III*. Bethesda, MA: National Association of School Psychologists.

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics, 8*(2), 157-159.

Öst, L.-G. (2008). Efficacy of the third wave of behavioral therapies: A systematic review and meta-analysis. *Behaviour Research and Therapy, 46*(3), 296-321. doi:10.1016/j.brat.2007.12.005

Page, M. J., McKenzie, J. E., Kirkham, J., Dwan, K., Kramer, S., Green, S., & Forbes, A. (2014). Bias due to selective inclusion and reporting of outcomes and analyses in systematic reviews of randomised trials of healthcare interventions. *The Cochrane Database of Systematic Reviews*(10). doi:10.1002/14651858.MR000035.pub2

Pajares, F., & Valiante, G. (1997). Influence of self-efficacy on elementary students' writing. *Journal of Educational Research, 90*(6), 353-360. doi:10.1080/00220671.1997.10544593

*Paleczek, L., Seifert, S., & Gastieger-Klicpera, B. (2017). Influences on teachers' judgment accuracy of reading abilities on second and third grade students: A multilevel analysis. *Psychology in the Schools, 54*(3), 228-245. doi:10.1002/pits.21993

Pallant, J. (2011). *SPSS survival manual: A step by step guide to data analysis using the SPSS program* (4th ed.). Berkshire: Allen & Unwin.

Pässler, K., Beinicke, A., & Hell, B. (2015). Interests and intelligence: A meta-analysis. *Intelligence, 50*, 30-51. doi:10.1016/j.intell.2015.02.001

Patil, V., H., Singh, S. N., Mishra, S., & Donavan, T. D. (2008). Efficient theory development and factor retention criteria: A case for abandoning the 'eigenvalue greater than one' criterion. *Journal of Business Research, 61*(2), 162-170. doi:10.1016/j.jbusres.2007.05.008

Patrick, D. L., Burke, L. B., Gwaltney, C. J., Leidy, N. K., Martin, M. L., Molsen, E., & Ring, L. (2011). Content validity—establishing and reporting the evidence in newly developed Patient-Reported Outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force Report: Part 2—assessing respondent understanding. *Value in Health, 14*, 978-988. doi:10.1016/j.jval.2011.06.013

Paulhus, D. L., Lysy, D. C., & Yik, M. S. M. (1998). Self-report measures of intelligence: Are they as useful as proxy IQ tests? *Journal of Personality, 66*(4), 525-554.

Payne, B. K. (2006). Weapon bias: Split second decisions and unintended stereotyping. *Current Directions in Psychological Science*. doi:10.1111/j.1467-8721.2006.00454.x

Pearson, R. H., & Mundform, D. J. (2010). Recommended sample size for conducting exploratory factor analysis on dichotomous data. *Journal of Modern Applied Statistical Methods, 9*(2), 359-368. doi:10.22237/jmasm/1288584240

*Pedulla, J. J., Airasian, P. W., & Madaus, G. F. (1980). Do teacher ratings and standardized test results of students yield the same information? *American Educational Research Journal, 17*(3), 303-307.

Pervin, L. A., Cervone, D., & John, O. P. (2005). *Personality: Theory and research* (9th ed.). New York, NY: John Wiley & Sons.

*Pesu, L., Viljaranta, J., & Aunola, K. (2016). The role of parents' and teachers' beliefs in children's self-concept development. *Journal of Applied Developmental Psychology, 44*, 63-71. doi:10.1016/j.appdev.2016.03.001

Peterson, C. H. (2012). The individual regulation component of group emotional intelligence: Measure development and validation. *The Journal for Specialists in Group Work, 37*(3), 232-251. doi:10.1080/01933922.2012.686962

Peterson, C. H., Peterson, N. A., & Powell, K. G. (2017). Cognitive interviewing for item development: Validity evidence based on content and response process. *Measurement and Evaluation in Counselling and Development, 50*(4), 217-223. doi:10.1080/07481756.2017.1339564

Peterson, R. A., & Brown, S. P. (2005). On the use of beta coefficients in meta-analysis. *Journal of Applied Psychology, 90*(1), 175-181. doi:10.1037/0021-9010.90.1.175

Petrucelli, M. L., Fiorello, C. A., & Thurman, S. K. (2010). Comparison of teachers' and school psychologists' accuracy in assigning basic academic tasks to underlying CHC-model cognitive abilities. *Journal of Applied School Psychology, 26*(3), 230-246. doi:10.1080/15377903.2010.495921

Pett, M. A., Lackey, N. R., & Sullivan, J. J. (2003). *Making sense of factor analysis*. Thousand Oaks, CA: SAGE Publications Ltd.

Pettiti, D. B. (2000). *Meta-analysis, decision analysis, and cost-effective analysis* (2nd ed.). New York, NY: Oxford University Press, Inc.

Pfeiffer, S. I. (2003). Challenges and opportunities for students who are gifted: What the experts say. *Gifted Child Quarterly, 47*(2), 161-169. doi:10.1177/001698620304700207

Phillipson, S., & Phillipson, S. N. (2007). Academic expectations, belief of ability, and involvement by parents as predictors of child achievement: A cross-cultural comparison. *Educational Psychology, 27*(3), 329-348. doi:10.1080/01443410601104130

*Phillipson, S., & Phillipson, S. N. (2012). Children's cognitive ability and their academic achievement: The mediation effects of parental expectations. *Asia Pacific Education Review, 13*, 495-508. doi:10.1007/s12564-011-9198-1

Pigott, T. D. (2009). Handling missing data. In H. M. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 400-415). New York, NY: Russell Sage Foundation.

*Plucker, J. A., Callahan, C. M., & Tomchin, E. M. (1996). Wherefore art thou, multiple intelligences? Alternative assessments for identifying talent in ethnically diverse and low income students. *Gifted Child Quarterly, 40*(2), 81-91. doi:10.1177/001698629604000205

Podsakoff, P. M., Mackenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879-903. doi:10.1037/0021-9010.88.5.879

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing and Health, 29*, 489-497. doi:10.1002/nur.20147

*Politmou, N., Masoura, E., & Kiosseoglou, G. (2015). Working Memory Rating Scale's utility to identify children's memory difficulties in diverse educational environments: Can it work in every school? *Applied Cognitive Psychology, 29*, 291-298. doi:10.1002/acp.3107

Pomerantz, E. M., & Wei, D. (2006). Effects of mothers' perceptions of children's competence: The moderating role of mothers' theories of competence. *Developmental Psychology, 42*(5), 950-961. doi:10.1037/0012-1649.42.5.950

*Prawat, R. S., & Jarvis, R. (1980). Gender difference as a factor in teachers' perceptions of students. *Journal of Educational Psychology, 72*(6), 743-749.

Presser, S., Couper, M. P., Lessler, J. T., Martin, E. M., J., Rothgeb, J. M., & Singer, E. (2004). Methods for testing and evaluating survey questions. *Public Opinion Quarterly, 68*(1), 109-130. doi:10.1093/poq/nfh008

*Pretzlik, U., Olsson, J., Nabuco, M. E., & Cruz, I. (2003). Teachers' implicit views of intelligence predict pupils' self-perceptions as learners. *Cognitive Development, 18*(579-600). doi:10.1016/j.cogdev.2003.09.008

Proctor, B. E. (2012). Relationships between Cattell-Horn-Carroll (CHC) cognitive abilities and math achievement within a sample of college students with learning disabilities. *Journal of Learning Disabilities, 45*(3), 278-227. doi:10.1177/0022219410392049

Räty, H., Kasanen, K., Kiiskinen, J., Nykky, M., & Atjonen, P. (2004). Childrens' notions of the malleability of their academic ability in the mother tongue and mathematics.

Scandinavian Journal of Educational Research, 48(4), 413-426. doi:10.1080/0031383042000245807

Recker, J., & Rosemann, M. (2007). *Understanding the process of constructing scales inventories in the process modelling domain.* Paper presented at the European Conference on Information Systems.

Reeve, C. L., & Bonaccio, S. (2008). Does test anxiety induce measurement bias in cognitive ability tests? *Intelligence, 36*(6), 526-538. doi:10.1016/j.intell.2007.11.003

Reid, R., & Maag, J. W. (1994). How many fidgets in a pretty much: A critique of behavior rating scales for identifying students with ADHD. *Journal of School Psychology, 32*(4), 339-354. doi:10.1016/0022-4405(94)90032-9

Reilly, D., Numann, D. L., & Andrews, G. (2017). Gender differences in spatial ability: Implications for STEM education and approaches to reducing the gender gap for parents and educators. In M. S. Khine (Ed.), *Visual-spatial ability: Transforming research into practice* (pp. 195-224). Switzerland: Springer International.

*Reilly, T. P., Drudge, O. W., Rosen, J. C., Loew, D. E., & Fischer, M. (1985). Concurrent and predictive validity of the WISC-R, McCarthy scales, Woodcock-Johnson, and academic achievement. *Psychology in the Schools, 22*(4), 380-382.

Reise, S. P., Waller, N. G., & Comrey, A. L. (2000). Factor analysis and scale revision. *Psychological Assessment, 12*(3), 287-297. doi:10.1037/1040-3590.12.3.287

Reschly, D. J., Myers, T. G., & Hartel, C. R. (2002). *Mental retardation: Determining eligibility for social security benefits*. Washington D. C.: National Academy Press.

Reschly, D. J., & Wilson, M. S. (1995). School psychology practitioners and faculty: 1986 to 1991-92 trends in demographics, roles, satisfaction, and system reform. *School Psychology Review, 21*(1), 62-80.

Reynolds, C. R., & Kamphaus, R. W. (2015). *Behavior Assessment System for Children - Third Edition (BASC-3)*: Pearson Clinical Assessment.

Reynolds, C. R., Livingston, R. B., & Wilson, V. (2006). *Measurement and assessment in education*. Boston: Pearson/Allyn & Bacon.

Reynolds, M. R., Scheiber, C., Hajovsky, D. B., Schwartz, B., & Kaufman, A. S. (2015). Gender differences in academic achievement: Is writing an exception to the gender similarities hypothesis? *The Journal of Genetic Psychology, 176*(3-4), 211-234. doi:10.1080/00221325.2015.103683

Richardson, K., & Norgate, S. H. (2015). Does IQ really predict job performance? *Applied Developmental Science, 19*(3), 153-169. doi:10.1080/10888691.2014.983635

Rohde, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence, 35*, 83-92. doi:10.1016/j.intell.2006.05.004

Roid, G. H. (2003). *Stanford-Binet Intelligence Scales, Fifth Edition*. Itasca, IL: Riverside Publishing.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638-641. doi:10.1037/0033-2909.86.3.638

Rossiter, J. R. (2002). The C-OAR-SE procedure for scale development marketing. *International Journal of Research in Marketing, 19*(4), 305-335. doi:10.1016/S0167-8116(02)00097-6

*Rothenbusch, S., Zettler, I., Voss, T., Lösch, T., & Trautwein, U. (2016). Exploring reference group effects on teachers' nominations of gifted students. *Journal of Educational Psychology, 108*(6), 883-897. doi:10.1037/edu0000085

Russo, M. W. (2007). How to review a meta-analysis. *Gastroenterology & Hepatology, 3*(8), 637-642.

Sackett, P. R., & Larson, J. R. (1990). Research strategies and tactics in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (pp. 419-489). Palo Alto, CA: Consulting Psychologists Press.

Salazar, M. S. (2015). The dilemma of combining positive and negative items in scales. *Psicothema, 27*(2), 192-199. doi:10.7334/psicothema2014.266

Sarstedt, M., & Wilczynski, P. (2009). More for less? A comparison of single-item and multi-item measures. *Die Betriebswirtschaft, 69*(2), 211-227.

Sattler, J. M. (1990). *Assessment of children* (3rd ed.). San Diego: J. M. Sattler.

Sauro, J., & Lewis, J. R. (2011). *When designing usability questionnaires, does it hurt to be positive?* Paper presented at the Proceedings of the International Conference on Human Factors in Computing Systems, CHI 2011, Vancouver, BC, Canada.

Scammacca, N., Roberts, G., & Stuebing, K. K. (2014). Meta-analysis with complex research designs: Dealing with dependence from multiple measures and multiple group copmarisons. *Review of Educational Resaerch, 84*(3), 328-364. doi:10.3102/0034654313500826

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Bulletin, 115*. doi:10.1037/0033-295X.115.2.336

Schmenk, B. (2004). Language learning: A feminine domain? The role of stereotyping in constructing gendered learner identities. *TESOL Quarterly, 38*(3), 514-524. doi:10.2307/3588352

Schmidt, F. L., & Oh, I.-S. (2016). The crisis of confidene in research findings in psychology: Is replication the real problem? Or is it something else? *Archives of Scientific Psyhology, 4*, 32-37. doi: 10.1037/arc0000029

Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99-144). New York, NY: Guilford Press.

Schneider, W. J., & McGrew, K. S. (in press). The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan & E. M. McDonaugh (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed.). New York, NY: Guilford Press.

Schneider, W. J., & Newman, D. A. (2015). Intelligence is multidimensional: Theoretical review and implications of specific cognitive abilities. *Human Resource Management Review, 25*(1), 12-27. doi:10.1016/j.hrmr.2014.09.004

Schreiber, J. B., Amaury, N., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99*(6), 323-338. doi:10.3200/JOER.99.6.323-338

Schwab, D. P. (1980). Construct validity in organizational behavior. In B. M. Staw & L. L. Cummings (Eds.), *Research in Organizational Behavior* (Vol. 2, pp. 3-43). Greenwich, CT: JAI Press.

Scott, J. (2015). *Vocabulary knowledge as a multidimensional concept: A six factor model.* Paper presented at the Twenty-second Annual Meeting of the Society for the Scientific Study of Reading, The Big Island, Hawaii.

*Shapiro, E. S., Gebhardt, S., Flatley, K., Guard, K. B., Fu, Q., Leichman, E. S., . . . Hojnoski, R. (2017). Development and validity of the Rating Scales of Academic Skills for Reading Comprehension. *School Psychology Quarterly, 32*(4), 509-524. doi:10.1037/spq0000193

Sharma, S., Mukherjee, S., Kumar, A., & Dillon, W. R. (2005). A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models. *Journal of Business Research, 58*(7), 935-943. doi:10.1016/j.jbusres.2003.10.007

Sharp, C., Mosko, O., Chang, B., & Ha, C. (2010). The cross-informant concordance and concurrent validity of the Borderline Personality Features Scale for Children in a community sample of boys. *Clinical Child Pychology and Psychiatry, 16*(3), 335-349. doi:10.1177/1359104510366279

Sharpley, C. F., & Edgar, E. (1986). Teachers' ratings vs standardized tests: An empirical investigation of agreement between two indices of achievement. *Psychology in the Schools, 23*(1), 106-111. doi:10.1002/1520-6807(198601)23:1<106::AID-PITS2310230117>3.0.CO;2-C

Siegel, L. S. (1989). IQ is irrelevant to the definition of learning disabilities. *Journal of Learning Disabilities, 22*(8), 469-478. doi:10.1177/002221948902200803

Siegler, R. S. (2007). Cognitive variability. *Developmental Science, 10*(1), 104-109. doi:10.1111/j.1467-7687.2007.00571.x

Sievertsen, H. H., Gino, F., & Piovesan, M. (2016). Cognitive fatigue influences students' performance on standardized tests. *Proceedings of the National Academy of Sciences, 113*(10), 2621-2624. doi:10.1073/pnas.1516947113

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*(1), 107-120. doi: 10.1007/s11336-008-9101-0

Silverman, L. K., Chitwood, D. G., & Waters, J. L. (1986). Young gifted children: Can parents identify giftedness? *Topics in Early Childhood Eduation, 6*(1), 23-38. doi:10.1177/027112148600600106

Silvia, E. S. M., & MacCallum, R. C. (1988). Some factors affecting the success of specification searches in covariance structure modeling. *Multivariate Behavioral REsearch, 23*(3), 297-326. doi:10.1207/s15327906mbr2303_2

*Simone, A. N., Marks, D. J., Bédard, A.-C., & Halperin, J., M. (2017). Low working memory rather than ADHD symptoms predicts poor academic achievement in school-aged children. *Journal of Abnormal Child Psychology, 46*(2), 277-290. doi:10.1007/s10802-017-0288-3

Singer, E., Van Hoewyk, J., & Maher, M. P. (2000). Experiments with incentives in telephone surveys. *Public Opinion Quarterly, 64*(2), 171-188.

Singh, D. (2016). *Examining Canadian teacher's perceptions of the importance of cognitive abilities in the classroom.* (Master of Education in School and Clinical Child Psychology), University of Alberta.

Sink, C. A., Barnett, J. E., & Pool, B. A. (1993). Perceptions of scholastic competnce in relation to middle-school achievement. *Perceptual and Motor Skills, 76*(q), 471-478. doi:10.2466/pms.1993.76.2.471

*Sommer, U., Fink, A., & Neubauer, A. C. (2008). Detection of high ability children by teachers and parents: Psychometric quality of new rating checklists for the assessment of intellectual, creative and social ability. *Psychology Science Quarterly, 50*(2), 189-205.

*Souroulla, A. V., Panayiotou, G., & Kokkinos, C. M. (2009). The role of the teacher in identifying learning disabilities. *Journal of Learning Disabilities, 42*(6), 483-493. doi:10.1177/0022219409335217

Spearman, C. (1904). "General Intelligence", objectively determined and measured. *The American Journal of Psychology, 15*(2), 201-292. doi:10.2307/1412107

Spector, P. E. (1992). *Quantitative applications in the social sciences: Summating rating scale construction*. Newbury Park, CA: SAGE Publications Ltd.

Spector, P. E. (2006). Method variance in organizational research: Truth or urban legend? *Organizational Research Methods, 9*(2), 221-232. doi:10.1177/1094428105284955

*Speece, D. L., Ritchey, K. D., Silverman, R., Schatschneider, C., Walker, C. Y., & Andrusik, K. N. (2010). Identifying children in middle childhood who are at risk for reading problems. *School Psychology Review, 39*(2), 258-276.

*Spinath, B., & Spinath, F. M. (2005). Development of self-perceived ability in elementary school: The role of parents' perceptions, teacher evaluations, and inteligence. *Cognitive Development, 20*, 190-204. doi:10.1016/j.cogdev.2005.01.001

*Steinmayr, R., & Spinath, B. (2009). What explains boys' stronger confidence in their intelligence? *Sex Roles, 61*(9-10), 736-749. doi:10.1007/s11199-009-9675-8

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence.* New York, NY: Cambridge University Press.

Sternberg, R. J., Conway, B. E., Ketron, J. L., & Bernstein, M. (1981). People's conception of intelligence. *Journal of Personality and Social Psychology, 41*(1), 37-55.

Sternberg, R. J., & Grigorenko, E. L. (2004). Successful intelligence in the classroom. *Theory Into Practice, 43*(4), 274-280. doi:10.1207/s15430421tip4304_5

Sternberg, R. J., & Kaufman, J. C. (1998). Human abilities. *Annual Review of Psychology, 49*, 479-502. doi:10.1146/annurev.psych.49.1.479

Sterne, J. A. C., & Harbord, R. M. (2004). Funnel plots in meta-analysis. *The Stata Journal, 4*(2), 127-141.

Stokes, E. K. (2011). Validity. In E. K. Stokes (Ed.), *Rehabilitation outcome measures* (pp. 35-46). Edinburgh: Churchill Livingstone.

Stone, E. R., Dodrill, C. L., & Johnson, N. (2001). Depressive cognition: A test of depressive realism versus negativity using general knowledge questions. *The Journal of Psychology: Interdisciplinary and Applied, 135*(6), 583-602. doi:10.1080/00223980109603722

Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual Review of Psychology, 5*, 1-25. doi:10.1146/annurev.clinpsy.032408.153639

Streiner, D. L., & Norman, G. R. (1989). *Health measurement scales: A practical guide to their development and use*. New York, NY: Oxford Medical Publications.

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of teachers' judgments of students' academic achievement: A meta-analysis. *Journal of Educational Psychology, 104*(3), 743-762. doi:10.1037/a0027627

*Südkamp, A., Praetorius, A.-K., & Spinath, B. (2017). Teachers' judgment accuracy concerning consistent and inconsistent student profiles. *Teaching and Teacher Education*. doi:10.1016/j.tate.2017.09.016

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Pearson/Allyn & Bacon.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson Education.

Taub, G. E., Keith, T. Z., Floyd, R. G., & McGrew, K. S. (2008). Effects of general and broad cognitive abilities on mathematics achievement. *School Psychology Quarterly, 23*(2), 187-198. doi:10.1037/1045-3830.23.2.187

Taub, G. E., & Valentine, J. (2014). A critical analysis of time allocation in psychoeducational evaluations. *Contemporary Issues in Education Research, 7*(4), 285-290.

Taylor, T. R. (1994). A review of three approaches to cognitive assessment, and a proposed integrated approach based on a unifying theoretical framework. *South African Journal of Psychology, 24*(4), 184-193. doi:10.1177/008124639402400403

Teglasi, H., Simcox, A. G., & Kim, N.-Y. (2007). Personality constructs and measures. *Psychology in the Schools, 44*(3), 215-228. doi:10.1002/pits.20218

Teo, T., Tsai, L. T., & Yang, C.-C. (2013). Applying structural equation modeling (SEM) in educational research: An introduction. In M. S. Khine (Ed.), *Application of structural equation modeling in educational research and practice* (pp. 3-22). Rotterdam: Sense Publishers.

Thompson, S. G., & Sharp, S. J. (1999). Explaining heterogeneity in meta-analysis: A comparison of methods. *Statistics in Medicine, 18*(20), 2693-2708. doi:10.1016/S0197-2456(97)91051-5

Thurstone, L. L. (1938). *Primary mental abilities.* Chicago, IL: University of Chicago Press.

Tiedemann, J. (2000). Parents' gender stereotypes and teachers' beliefs as predictors of children's concept of their mathematical ability in elementary school. *Journal of Educational Psychology, 92*(1), 144-151. doi:10.1037/0022-0663.92.1.144

Tiedemann, J., & Steinmetz, A. (1998). *Gender-related teacher beliefs in mathematics from a cross-cultural perspective*. Paper presented at the Congress of the German Psychological Association, Dresden, Germany.

Tomasetto, C., Mirisola, A., Galdi, S., & Cadinu, M. (2015). Parents' math–gender stereotypes, children's self-perception of ability, and children's appraisal of parents' evaluations in 6-year-olds. *Contemporary Educational Psychology, 42*, 186-198. doi:10.1016/j.cedpsych.2015.06.007

Tourangeau, R. (1984). Cognitive science and survey methods. In M. Jabine, M. Straf, J. Tanur, & R. Tourangeau (Eds.), *Cognitive aspects of survey methodology: Building a bridge between disciplines* (pp. 73-100). Washington, DC: National Academy Press.

Tourangeau, R., Rips, L. J., & Rasinki, K. (2000). *The psychology of survey response*. New York, NY: Cambridge University Press.

Tractenberg, R. E. (2009). Analytic methods for factors, dimensions and endpoints in clinical trials for Alzheimer's disease. *The Journal of Nutrition, Health and Aging, 13*(3), 249-255.

*Triga, A. (2004). An analysis of teachers' rating scales as sources of evidence for a standardised Greek reading test. *Journal of Research in Reading, 27*(3), 311-320. doi:10.1111/j.1467-9817.2004.00234.x

Tripp, G., Schaughency, E., & Clarke, B. (2006). Parent and teacher rating scales in the evaluation of Attention-Deficit Hyperactivity Disorder. *Journal of Developmental & Behavioral Pediatrics, 27*(3), 209-218. doi:10.1097/00004703-200606000-00006

Trninić, V., Jelaska, I., & Štalec, J. (2013). Appropriateness and limitations of factor analysis methods utilised in psychology and kinesiology - Part 2. *Physical Culture, 67*(1), 5-17.

Truetler, C. M., & Epkins, C. C. (2003). Are discrepancies among child, mother, and father reports on children's behavior related to parents' psychological symptoms and aspects of parent-child relationships? *Journal of Abnormal Child Psychology, 31*(1), 13-27. doi:10.1023/A:1021765114434

Upadyaya, K., & Eccles, J. S. (2014). Do teachers' perceptions of children's math and reading related ability and effort predict children's self-concept of ability in math and reading? *Educational Psychology, 35*(1), 110-127. doi:10.1080/01443410.2014.915927

*Uppal, H. K., & Swanson, H. L. (2016). Teachers' ratings of working memory in English language learners: Do laboratory maesures predict classroom analogues? *Applied Cognitive Psychology, 30*, 871-884. doi:10.1002/acp.3273

*Urhahne, D., Chao, S.-H., Florineth, M. L., Luttenberger, S., & Paechter, M. (2011). Academic self-concept, learning motivation, and test anxiety of the underestimated student. *British Journal of Educational Psychology, 81*(1), 161-177. doi:10.1348/000709910X504500

van de Mortel, T. F. (2008). Faking it: Social desirability response bias in self-report research. *Australian Journal of Advanced Nursing, 25*(4), 40-48.

van Sonderon, E., Sanderman, R., & Coyne, J. C. (2013). Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. *PLOS ONE, 8*(9), e68967. doi:10.1371/journal.pone.0068967

Vanderwood, M. L., McGrew, K. S., Flanagan, D. P., & Keith, T. Z. (2002). The contribution of general and specific cognitive abilities to reading achievement. *Learning and Individual Differences, 13*(2), 159-188. doi:10.1016/S1041-6080(02)00077-8

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods, 10*(4), 428-443. doi:10.1037/1082-989X.10.4.428

*Vilhena, D. d. A., & Pinheiro, Â. M. V. (2016). Revised version of the Scale of Evaluation of Reading Competence by the teacher: Final validation and standardization. *Universitas Pscyhologica, 15*(4). doi:10.11144/Javeriana.upsy15-4.efvs

Visser, B. A., Ashton, M. C., & Vernon, P. A. (2006). Beyond g: Putting multiple intelligences theory to the test. *Intelligence, 34*(5), 487-502. doi:10.1016/j.intell.2006.02.004

Vogt, W. P. (1993). *Dictionary of statistics and methodology: A nontechnical guide for the social sciences*. Thousand Oaks, CA: SAGE Publications, Inc.

*Waschbusch, D. A., Daleiden, E., & Drabman, R. S. (2000). Are parents accurate reporters of their child's cognitive abilities? *Journal of Psychopathology and Behavioral Assessment, 22*(1), 61-77.

Waterhouse, L. (2006). Inadequate evidence for multiple intelligences, Mozart effect, and emotional intelligence theories. *Educational Psychologist, 41*(4), 247-255. doi:10.1207/s15326985ep4104_5

Watkins, M. W. (2000). Monte Carlo PCA for Parallel Analysis (Computer Software). State College, PA: Ed & Psych Associates.

Watkins, M. W. (2006). Determining parallel analysis criteria. *Journal of Modern Applied Statistical Methods, 5*(2), 344-346. doi:10.22237/jmasm/1162354020

Watkins, M. W., Glutting, J. J., & Lei, P.-W. (2007). Validity of the full-scale IQ when there is significant variability among WISC-III and WISC-IV factor scores. *Applied Neuropsychology, 14*(1), 13-20. doi:10.1080/09084280701280353

Webster, B. J., & Fisher, D. L. (2001). *A two-step approach to modelling student performance: A methodology that accounts for measurement and structural error*. Paper presented at the Australian Association for Research in Education, Fremantle, Perth. https://www.aare.edu.au/data/publications/2001/web01573.pdf

Weijters, B. (2006). *Response styles in consumer research.* (Doctor in Applied Economic Science), Ghent University, Ghent, Belgium. Retrieved from http://hdl.handle.net/1854/LU-4100284

Weijters, B., & Baumgartner, H. (2012). Misresponse to reversed and negated items in surveys: A review. *Journal of Marketing Research, 49*(5), 737-747.

Whittaker, T. A. (2012). Using the modification index and standardized expected parameter change for model modification. *The Journal of Experimental Education, 80*(1), 26-44. doi:10.1080/00220973.2010.531299

Wickens, T. D. (1994). *The geometry of multivariate statistics*. Mahwah, NJ: Lawrence Erlbaum Associates.

Wiggins, G. (1998). *Educative assessment: Designing assessments to inform and improve student performance*. San Francisco, CA: Jossey-Bass.

Williams, B., Brown, T., & Onsman, A. (2010). Exploratory factor analysis: A five-step guide for novices. *Australasian Journal of Paramedicine, 8*(3).

Williams, K. S., Ochs, J., Williams, M. J., & Mulhern, R. K. (1991). Parental report of everyday cognitive abilities among children treated for acute lymphoblastic leukemia. *Journal of Pediatric Psychology, 16*(1), 13-26. doi:10.1093/jpepsy/16.1.13

Willis, G. B. (1999). *Cognitive interviewing: A "how to" guide*. Reducing survey error through research on the cognitive and decision processes in surveys. Research Triangle Institute. 1999 Meeting of the American Statistical Association.

Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: SAGE Publications, Inc.

Willis, G. B., & Miller, K. (2011). Cross-cultural cognitive interviewing: Seeking comparability and enhancing understanding. *FIeld Methods, 23*(4), 331-341. doi:10.1177/1525822X11416092

*Wilson, J., & Wright, C. R. (1993). The predictive validity of student self-evaluations, teachers' assessments, and grades for performance on the verbal reasoning and numerical ability scales of the Differential Aptitude test for a sample of secondary school students attending rural Appalachia schools. *Educational and Psychological Measurement, 53*(1), 259-270.

Wood, D., Kurtz-Costes, B. E., Okeke-Adeyanju, N., & Rowley, S. J. (2010). Mothers' academic gender stereotypes and education-related beliefs about sons and daughters in African American families. *Journal of Educational Psychology, 102*(2), 521-530. doi:10.1037/a0018481

Woolley, M. E., Bowen, G. L., & Bowen, N. K. (2006). The development and evaluation of procedures to assess child self-report item validity. *Educational and Psychological Measurement, 66*(4), 687-700. doi:10.1177/0013164405282467

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counselling Psychologist, 34*(6), 806-838.

*Wright, D., & Wiese, M. J. (1988). Teacher judgment in student evaluation: A comparison of grading methods. *The Journal of Educational Research, 82*(1), 10-14.

Wright, K. B. (2005). Researching internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated Communication, 10*(3), JCMC1034. doi:10.1111/j.1083-6101.2005.tb00259.x

Yen, C.-J., Konold, T. R., & McDermott, P. A. (2004). Does learning behavior augment cognitive ability as an indicator of academic achievement? *Journal of School Psychology, 42*(2), 157-169. doi:10.1016/j.jsp.2003.12.001

Yong, A. G., & Pearce, S. (2013). A beginner's guide to factor analysis: Focus on exploratory factor analysis. *Tutorials in Quantitative Methods for Psychology, 9*(2), 79-94. doi:10.20982/tqmp.09.2.p079

Zakharov, A., & Carnoy, M. (2015). Are teachers accurate in predicting their students' performance on high stakes' exams? The case of Russia. *International Journal of Educational Development, 43*, 1-11. doi:10.1016/j.ijedudev.2015.04.007

*Zsoldos, M. (2003). Low-achieving (learning disabled) pupils in year 2 of the primary school in Hungary. *European Journal of Special Needs Education, 18*(2), 243-250. doi:10.1080/0885625032000079014

Zucker, S. (2004). *Administration practices for standardized assessments*: Pearson Assessment Report.

Appendix

Appendix A

Self- and Other-report Measures of Cognitive Abilities Reviewed During the Development of the Estimates of Children's Cognitive Abilities

Chan, D. W. (2001). Assessing giftedness of Chinese secondary students in Hong Kong: A multiple intelligences perspective. *High Ability Studies, 12*, 215-234.

Fiorello, C. A., Thurman, S. K., Zavertnik, J., Sher, R., & Coleman, S. (2009). A comparison of teachers' and school psychologists' perceptions of the importance of CHC abilities in the classroom. *Psychology in the Schools, 46*(6), 489-500. doi:10.1002/pits.20392

Hedges, R., Drysdale, K., & Levick, W. R. (2015). The Children's Memory Questionnaire-Revised. *Applied Neuropsychology, 4*(4), 285-296. doi:10.1080/21622965.2014.925806

Hernández-Torrano, D., Ferrándiz, C., Ferrando, M., & Prieto, L. (2014). The theory of multiple intelligences in the identification of high-ability students. *Anales De Psicología*, 30(1), 192-200. doi:10.6018/analesps.30.1.148271

Jacobs, K. E. (2012). *Validly measuring cognitive ability by self-report: Relevance of response format and type of ability assessed.* (Master of Psychology (Educational & Developmental)/Doctor of Philosophy), Monash University.

Lamb, G. D. (2008). *Ratings of everyday academic and cognitive skills in evaluation of school learning and learning problems: Initial scale development and validation*. (Doctor of Philosophy), Texas A&M University.

Petrucelli, M. L., Fiorello, C. A., & Thurman, S. K. (2010). Comparison of teachers' and school psychologists' accuracy in assigning basic academic tasks to underlying CHC-model cognitive abilities. *Journal of Applied School Psychology, 26*(3), 230-246. doi:10.1080/15377903.2010.495921

Seidenberg, M., Haltiner, A., Taylor, M. A., Hermann, B. B., & Wyler, A. (1994). Development and validation of a multiple ability self-report questionnaire. *Journal of Clinical and Experimental Neuropsychology, 16*, 93-104. doi: 10.1080/01688639408402620

Steinmayr, R., & Spinath, B. (2009). What explains boys' stronger confidence in their intelligence? *Sex Roles, 61*(9-10), 736-749. doi:10.1007/s11199-009-9675-8

Waschbusch, D. A., Daleiden, E., & Drabman, R. S. (2000). Are parents accurate reporters of their child's cognitive abilities? *Journal of Psychopathology and Behavioral Assessment, 22*(1), 61-77.

Woodhouse, B., & Jackson, P. H. (1977). Lower bounds for the reliability of the total score on a test composed of non-homogenous items: II: A search procedure to locate the greatest lower bound. *Psychometrika, 42,* 579-591.

Williams, K. S., Ochs, J., Williams, M. J., & Mulhern, R. K. (1991). Parental report of everyday cognitive abilities among children treated for acute lymphoblastic leukemia. *Journal of Pediatric Psychology, 16*(1), 13-26. doi:10.1093/jpepsy/16.1.13

Appendix B

Invitation to Participate in Expert Panel Review and Survey Questions

Dear [name],

You are invited to take part in an expert panel to review a newly developed parent- and teacher-report measure of Cattell-Horn-Carroll (CHC) cognitive abilities.

My name is Lydia Soh and I am currently completing my PhD in Educational & Developmental Psychology at Monash University under the supervision of Dr Kate Jacobs. The aim of my thesis is to determine the extent to which parents and teachers can accurately report on a child's level of cognitive functioning on 10 of the broad cognitive ability areas from the Cattell-Horn-Carroll (CHC) Model of Cognitive Abilities. To achieve this, I am developing parent- and teacher-report scales measuring each of the CHC broad cognitive ability areas, called Estimates of Children's Cognitive Abilities (ECCA).

You have been invited to be a part of the expert panel due to your expert knowledge and understanding of the CHC model. Your collaboration with the scale item writing process would involve sharing your professional knowledge and experience through the completion of an online survey (expected to take 20 to 30 minutes). At no time will you be asked to divulge information of a personal or sensitive nature.

If you are able to be involved, I ask that you please respect the intellectual property. If and when a commercially available parent- and teacher-report measure is available, your help will be recognised through privileged access to the measure, which you can request for in writing.

Please provide your contact details on the questionnaire so I may contact you if I have follow-up questions. Please try to respond to every item. There are no right or wrong answers – I am simply aiming to gather the opinions and perceptions of others in the field. You are not expected to do any research or reading. Please just use the knowledge and clinical experience that you already possess. If you have any questions, please feel free to contact me.

You can access the questionnaire via http://tinyurl.com/eccaip.

Thank you for your participation.

Kind regards,

Lydia Soh

Thank you for participating in the expert review of the Estimates of Children's Cognitive Abilities (ECCA) questionnaire, developed to obtain parent- and teacher-reports of children's Cattell-Horn-Carroll abilities.

You have been invited to be a part of the expert panel due to your expert knowledge and understanding of the CHC model. Your collaboration with the scale item writing process would involve sharing your professional knowledge and experience through the completion of an online questionnaire. This is expected to take 20 to 30 minutes. At no time will you be asked to divulge information of a personal or sensitive nature.

Please respect the intellectual property. If and when a commercially available parent- and teacher-report measure is available, your help will be recognised through privileged access to the measure, which you can request for in writing.

Please try to respond to every item. There are no right or wrong answers - I am simply aiming to gather the opinions and perceptions of others in the field. You are not expected to do any research or reading. Please just use the knowledge and clinical experience that you already possess.

If you have any questions, please feel free to contact me:

Lydia Soh
lydia.s.soh@monash.edu
+61 03 9902 4891

These following scale items have been written to reflect behavioural manifestations that are indicative of the following eleven CHC broad ability areas; Fluid Reasoning (Gf), Crystallised Knowledge (Gc), Short-term Working Memory (Gwm), Learning Efficiency (Gl), Retrieval Fluency (Gr), Auditory Processing (Ga), Processing Speed (Gs), Visual Processing (Gv), Reading (Grw-R), Writing (Grw-W), and Quantitative Knowledge (Gq). Some commonly used definitions of the broad cognitive ability areas will be listed.

Please indicate the **construct-relevance** of the item to the corresponding CHC broad ability, and the **clarity and conciseness** of the item. These ratings will be based on a scale of 1 to 5.

Please provide your comments (if any) on how to improve the wording of the current items, and if there are any other **observable behaviours and/or external expressions** you believe is important to include.

Please provide your contact details below so I may contact you if I have any follow-up questions.

Name:

Email Address:

FLUID REASONING (Gf): the use of deliberate and controlled focused attention to solve novel, "on the spot" problems that cannot be solved solely by using prior knowledge; reasoning that depends minimally on learning and acculturation

Scale item	Relevance to Gf	Clarity and conciseness
Gf 1. Identify what comes next in a pattern	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gf 2. Understand how parts fit together in a puzzle	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gf 3. Know what is missing from an incomplete logic puzzle	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gf 4. Know the number that comes next in a number series (e.g., 1, 2, 4, 7, 11, ...?)	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gf 5. Form and recognise concepts (e.g., how are a dog, cat and cow alike?)	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gf 6. Identify and perceive relationships (e.g., sun is to morning as moon is to ...?)	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gf 7. Select one of several pictures to complete a puzzle	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gf 8. Make connections between new material and acquired knowledge	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gf 9. Demonstrate problem solving skills in new and everyday situations	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required

Do you have any comments about the existing items of Gf?

Are there any other **observable behaviours** or **external expressions** of Gf that you think would be important to include?

COMPREHENSION-KNOWLEDGE (Gc): to have a large amount of acquired knowledge language ability, and to be able to effectively use that knowledge; comprehension of language, words and general knowledge developed through experience, learning, and acculturation

Scale item	Relevance to Gc	Clarity and conciseness
Gc.1. Understand conversations	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gc.2. Provide correct responses to questions of fact (e.g., in what direction does the sun set?)	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gc.3. Provide synonyms and antonyms of a word (e.g., word "happy; synonym "cheerful"; antonym "sad")	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gc.4. Point to pictures as orally instructed	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gc.5. Convey precise what he/she is tying to say	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gc.6. Participate in conversations with other children	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gc.7. Display the extent of his/her general knowledge	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gc.8. Provide oral definitions for words	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gc.9. Understand verbal directions	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required

Do you have any comments about the existing items of Gc?

Are there any other **observable behaviours** or **external expressions** of Gc that you think would be important to include?

SHORT-TERM WORKING MEMORY (Gwm): the ability to apprehend and hold information in immediate awareness and then use it within a few seconds

Scale item	Relevance to Gwm	Clarity and conciseness
Gwm.1. Follow multi-step verbal instructions	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gwm.2. Writing down dictated information	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gwm.3. Repeat numbers or words orally in the same order as presented (e.g., 7, tree, 4, boy)	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gwm.4. Remember a phone number long enough to dial it	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gwm.5. Remember a series of related words (e.g., red, orange, yellow)	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gwm.6. Remember a series of unrelated words (e.g., dog, bottle, umbrella)	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gwm.7. Retrieve all items on a list as instructed orally	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gwm.8. Remember all materials needed for a task	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required

Do you have any comments about the existing items of Gwm?

Are there any other **observable behaviours** or **external expressions** of Gwm that you think would be important to include?

LEARNING EFFICIENCY (Gl): the ability and efficiency to learn, store, and consolidate new information in long-term memory

Scale item	Relevance to Gl	Clarity and conciseness
Gl.1. Accurate recall names paired with pictures of faces	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gl.2. Recall specific words or facts	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gl.3. Recall basic math facts	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gl.4. Recall mathematical procedures	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gl.5. Acquire new skills	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gl.6. Relate and link learned information together	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gl.7. Absorb and retain information	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gl.8. Relate new material to previous knowledge	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required

Do you have any comments about the existing items of Gl?

Are there any other **observable behaviours** or **external expressions** of Gl that you think would be important to include?

RETRIEVAL FLUENCY (Gr): the rate and fluency at which individuals are able to access information stored in long-term memory

Scale item	Relevance to Gr	Clarity and conciseness
Gr.1. Retrieve needed information from long-term memory	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gr.2. Remember previously learnt spelling words	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gr.3. Recall facts for end-of-term tests and/or exams	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gr.4. Rapidly name letters presented on a page	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gr.5 Remember the order in which events have occurred	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gr.6. Retell a story he/she has heard before	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gr.7. Recall facts about what he/she has read in a book	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required

Do you have any comments about the existing items of Gr?

Are there any other **observable behaviours** or **external expressions** of Gr that you think would be important to include?

VISUAL-SPATIAL PROCESSING (Gv): the ability to think with visual patterns and stimuli. Includes the ability to mentally rotate, reverse, and manipulate spatial configurations, and spatial orientation.

Scale item	Relevance to Gv	Clarity and conciseness
Gv.1. Differentiate between similarly shaped symbols (e.g., d and b; 3 and 5)	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gv.2. Identify information from pictures, charts, graphs, maps, etc.	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gv.3. Organise information from different sources into one document	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gv.4. Find specific information on a printed page (e.g., getting a number out of a phone book)	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gv.5. Put parts together to form a whole (e.g., three dimensional objects)	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gv.6. Build a model (with Lego or blocks) from a picture of the completed model	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gv.7. Not be distracted by irrelevant visual information	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gv.8. Accurately judge distances (e.g., place objects not too close to an edge)	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required

Do you have any comments about the existing items of Gv?

Are there any other **observable behaviours** or **external expressions** of Gv that you think would be important to include?

AUDITORY PROCESSING (Ga): the ability to notice, compare, discriminate, and distinguish distinct and separate sounds.

Scale item	Relevance to Ga	Clarity and conciseness
Ga.1. Blend letter sounds together fluently to form words	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Ga.2. Know which sounds go with which letters	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Ga.3. Understand what is being said when background noise is present	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Ga.4. Hear the difference between similar sounding words (e.g., rhyming words)	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Ga.5. Recognise a word only when parts of it are pronounced	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Ga.6. Differentiate speech sounds (e.g., tell the difference between /ch/ and /sh/, similar vowel and consonant sounds, etc.)	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Ga.7. Hear different sounds in words	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Ga.8. Filter out background noise to listen to teacher's voice in the classroom	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required

Do you have any comments about the existing items of Ga?

Are there any other **observable behaviours** or **external expressions** of Ga that you think would be important to include?

PROCESSING SPEED (Gs): the ability to perform simple repetitive cognitive tasks quickly and fluently

Scale item	Relevance to Gwm	Clarity and conciseness
Gs.1. Work quickly and efficiently on tasks already mastered	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gs.2. Quickly compare how different or similar two objects are	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gs.3. Complete a series of arithmetic problems within a time limit	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gs.4. Quickly and accurately copy information from the whiteboard	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gs.5. Complete mastered tasks in a timely manner	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gs.6. Solve math questions automatically	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gs.7. Scan and quickly determine important information on a page	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gs.8. Complete classwork in a timely manner	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required
Gs.9. Display automaticity of rote information	☐ Not at all relevant ☐ Of little relevance ☐ Somewhat relevant ☐ Relevant ☐ Very relevant	☐ Poor; did not understand ☐ Fair; understood a little ☐ Good; understood most of it ☐ Very good; understood very well ☐ Excellent; no changes required

Do you have any comments about the existing items of Gs?

Are there any other **observable behaviours** or **external expressions** of Gs that you think would be important to include?

Appendix C

Transcript of Cognitive Interviews with Psychologist

D= psychologist; L= interviewer

L: What sort of behaviours would indicate that a child has high levels of fluid reasoning?

D: Quite good at problem solving, in a social setting, I'm just thinking of kids in school, they'd be good at reading social cues, they would obviously do well with mathematics and high level comprehension and they probably would have "street smarts" as in... Maybe someone from an impoverished background with high level of Gf would still be quite capable at surviving in the playground because they can use problem solving skills as opposed to their learnt skills from environment... does that make sense?

L: Yeah it does, so basically just sort of adapting to new situations and applying problem solving skills to solve social situations… What about behaviours that a child might express if he has low levels of Gf?

D: From an academic POV they would struggle mathematically, with high level comprehension... sort of thinking from social POV they would have poor problem solving skills. They would struggle in a playground setting and they might find that they react emotionally as compared to someone else who would be able to work through things

L: See some of the issues that I faced with including the social aspects of how cog skills apply in these sort of things is that some parents may actually confuse that with indicators of behaviours to do with ASD

D: I'm not actually indicating that at all, but kids with low Gf really don't get it in the playground situation! They just don't have the skills to actually read social situations or to apply what they've learnt in one social situation to the next. And this doesn't mean a diagnosis of ID or ASD. Quite interesting when you do some profiles of S&W, kids who do struggle in the playground often do have issues with Gf because they just can't do that at all!

L: I guess just the balance of trying to word it in such a way that parents won't think that "oh like is this questionnaire actually asking me for ASD indicators" just something that I need to work on

D: Something you should talk with your supervisor about but it certainly does impact on them, more than… well there is a level of needing to understand that it can impact socially. And you know you can have social problems without meeting the criteria for autism. The trick is to word it I think!

D: The other thing too is that when people look at cognitive abilities the first thing they think of is academic interface, and it's not always just that. So maybe I don't know, you could look at academic and social impact too.

L: We might go on to the ECCA items now – so the first one is "know what is missing from an incomplete logic puzzle" how would you rate this item, is it clear or unclear?

D: I think that's tricky for the parent. They may say "what's a logical puzzle?"

L: so the term "logical puzzle" might make it difficult for parents to understand?"

D: you would have to have examples - what is a logical puzzle

L: so do you think the word "picture pattern, visual puzzle" might be better?

D: yeah, I don't like "logical puzzle" as a parent. Some teachers might also go "what does that mean"? Yeah.. and "incomplete logical puzzle".. I'm just trying to think, if i'm looking at that without knowing what Matrix Reasoning is... I would think of jigsaw.

L: Yeah but see I had one of the other items on my list “select one of several pictures to complete the puzzle” so in that sense it might seem repetitive for parents so I think the issue here is the “Logical puzzle” term that I need to reword it in some different way, that it’s obvious that I mean visual puzzle in some way

D: even selecting one of several pictures to complete the puzzle can mean to be jigsaw

L: yes so if we just go to that item I guess do you think that item is written clearly or should I be adding a descriptor to puzzle in front of it which might make it a bit clearer as to what I mean by puzzle

D: I actually think that examples would be useful for most of them.. then maybe just say to complete a jigsaw puzzle. do you actually mean to find a picture that fits into a jigsaw puzzle or do you mean like that things that might go together you know like how you've got the Matrix reasoning on the wechsler scale like you've got the car placed in different ways and in different colours is that what you mean there?

L: No I think I actually did mean jigsaw puzzle

D: as in so that would be like a section or a shade that fills in the hole in the puzzle if you'd like. Does that measure fluid reasoning?

L: that actually sounds like visual processing now if you word it that way. well I'll put a question mark next to that since it's obviously might not be exploring Fluid reasoning very well.

D: or if you are still Referring to visual puzzles, sequencing that sort of thing that might still be relevant if it's a jigsaw puzzle that to me is visual processing because what you want to measure here is patterns like 1 square, 2 squares, what comes next. Like what’s the rule here

L: we might go to the second item now, identify and perceive relationships between new concepts. I didn't put an example here just because I wanted to see what was the first thing that comes to mind when a test-taker might read this item so how would you rate this item based on whether it is clear or not?

D: I think it’s clear in my language but who is your target audience?

L: Still parents and teachers. in the original questionnaire there was an example next to this item sun is to morning as moon is to blank. a lot of the expert panel said that this has to do with Gc because of the words so I re-worded it to make it seem not so constricted to words itself. So for example in WJ Analysis-Synthesis, Concept Formation - that was the concepts that I was trying to go for.

D: I think it does, but the word concept the word is a hard thing for people to wrap their heads around, it’s a word that people don't use very often. a lot of parents of find it hard to understand teachers would understand that some parents might have difficulty this is. this is a report on my child's cognitive skills is that right

L: yeah the parent is meant to estimate how easy or difficult a child would find the following tasks

D: Is it possible to have picture diagrams on the questionnaire as examples?

L: No I haven't thought of that but do you mean like pictures of examples of what tasks would look like what the child would have to do? Kate and I actually discussed maybe describing the actual tasks that children would have to do on intelligence tests like the Wechsler or Woodcock Johnson and then see whether the parent would be able to figure out if your child could accomplish the task or not. But at this stage we have gone with behavioural manifestations of the narrow abilities that make up the broad abilities instead. But yea, picture diagrams of the task might be useful

D: sun is to moon is verbal, but maybe you can find some similar non-verbal concepts which would have to be a diagram

L; so something like what concept formation does, but a simplified version (and avoiding copyright).. That might be something to think about, talk about. Ok so with the 4th item - sort and classify objects according to predetermined categories

D: that's clear… you probably need to get a different word for predetermined that might be a bit too hard for lay people I'm just trying to think the audience that you've got it's quite a large variety like parents have a huge range of intellect.

L: ok me last item making connections between new material and what he/she already knows. Is that a clearly worded item and captures Gf reasonably?

D: yes… but it's a bit of long-term.. Association isn't it? But still that rule-making ability

L: It's sort of like I guess using previously acquired knowledge to solve new problems. That is what this item is trying to capture.
L: So let's move onto crystallised knowledge now. What sort of behaviours would a child express if he has high or low level of Gc?

D: high - quite verbal, large vocab, ability to reason verbally, quite likely to read well, and write well unless they have an SLD. low - "fillers" in language (e.g., um ah) might speak a lot but have lack of content in language, low vocab, probably difficulty with written and spoken comprehension. Maybe lack of knowledge about the world (social cues, general knowledge)

L: So moving onto the items, the first one - understand conversations of others - is that clear?

D: clear and explores intended construct. Reword to understand others' conversations.

L: 2nd item - communicate effectively what he/she is trying to say. D: all good. Trying to think from parent POV all the time as opposed to teacher

L: 3rd item - provide correct responses to questions of fact e.g., in what direction does the sun set. D: all good

L: participate in conversations with other children - all good

L; Display extent of his/her general knowledge - display general knowledge of learnt social rules and concepts, (e.g. what should you do when you see smoke coming from neighbour's house)

D: again, the word concept is hard. Think of a different word to concept.

L: what about if i just remove the word concepts, just have learnt social rules, and leave the examples? Yes good

L: point to pictures as orally instructed - what about this item?

D: as in “show me the dog, show me the nose?” I think that would capture the smaller aged people, as the child gets older you would get into trouble. Are you trying to capture vocab? You only point to pictures for the little people right?

L: I have another item here - name familiar pictured objects - do you think that can apply to the little people?

D: I like that more than orally instructed, but it depends on the age, would be harder for some of them. Will the questions cut out as you get older?

L: That’s something that Kate and I said that we’d have to factor analyze and see if we can have diff age groups answer diff questions.

D: the other thing to, is that when you’re pointing to it is you’re showing that you understand.. You would have to have a higher level for older kids. The difference to pointing something and being able to define something

L: I haven’t included today “providing definitions of words” or things like that but I do have items that address defining words “provide antonyms and synonyms of words”

D: you will run out of that ceiling very quickly for “point to pictures”

L: so I have to think about whether to retain this item just for little people or axe it altogether

D: how are you going to differentiate between the 15yo that can do it and the little person? [30:48] Vineland - if the child performed the behaviour in the past, then you can score “2 points” like as in “show me your nose” for a parent of a 15yo they’ll go “omg, they should be able to do that” so if they’ve already learnt the skill then they can score full points. ADOS has the same concept as well - if they’ve done it but aren’t doing it currently they get full points

L: let’s move onto Gwm - what behaviours would indicate to you that a child has high/low level of Gwm

D: High - ability to listen to instruction, like complex multi-step… would be able to comprehend better because they would remember the whole sentence. LIkewise with WM and ability to do mental mathematics and to be able to work in their head… Low WM would have academic diff across the board, inability to follow stepped instruction, distractible in classroom setting bc wouldn’t really know what was going on in the classroom

L: remember series of related words in correct order, related like colours of the rainbow.. How would you rate this item

D: it’s quite clear, but i’m wondering why it has to be related words

L: capturing the essence of the WJ tasks, child has to remember related words, and then repeat the list of words in the correct order.. You had a question if it had to be related or not

D: yeah.. Like sometimes it might just be “dog, tree, shell”, as opposed to “dog, chihuahua, daschund”. Doesn’t necessarily have to be related. Still captures WM whether it’s related or not. Might be a better reflection actually because related words.. With semantics might actually help with WM

L: remember a verbally spoken phone number for long enough to write it down

D: well people don’t actually remember phone numbers anymore, but yeah that’s clear. That’s a good older child one rather than a younger one, so that’s something you’ll have to work out.

L: is there an item.. That i guess is similar, that would apply for younger kids?

D: well it's like a sequence of numbers isn't it?

L: short of just rephrasing the Digit Span task, are there opportunities in the classroom for children to remember a list of digits?

D: maybe remember a sequence of… or well, not in a natural setting I suppose

L: follow spoken directions or instructions that come one after another - is this clear or unclear

D: maybe "several spoken directions" .. how many do you want? Is it two or three step direction or what?

L: follow spoken multi-step directions?

D: again, what is multi-step? Is it like three or more, or two or more tasks that you're after?

L: i guess, at least two?

D: follows instructions to do two or more tasks… go and pick up your shoes and put your clothes in the basket - so have an example

L: write down dictated sentences accurately

D: it's clear… but is it testing more than Gwm? What if I spell it incorrectly, I'm not accurate?

L: I guess the point of the task is to hold what the teachers says in memory, and be able to write it down

D: so how important is spelling in that? I guess the issue is what does accurately mean?

L: how would you ask this item to make it clearer? In my mind it doesn't necessarily have to be spelling error free, but needs to reflect the meaning

D: verbatim is the word you're after, but it's a bad word to have in the item because it's technical. So accurately means.. If you want to make sure the punctuations correct, the spelling's correct, then you're not actually assessing Gwm anymore

L: i guess finding a more user-friendly version of "verbatim"

D: or ignoring spelling and punctuation errors might work

L: so the next item I have is (last item Gwm) basically a description of a subtest of a cog test. My issue here is that i don't know a real-world classroom example of where a child would do this, but it is a key aspect of Gwm so…
D: I actually think… listen to a series of numbers of words, then repeat the words first in alphabetical order, then numbers from lowest to highest. I know what you're capturing… but whether it's relatable to the child or whether the parent would know.. If you've never seen the child do such a task, then how will the parent estimate if the child can do the task or not? It's a performance-based question.. You'd actually have to see them do it. It makes me think of "here is the info, now i have to manipulate it in my head and give it back to you in a different way."

D: You know that game "Alphabet" - I'm A, I say Apple, you're B, you say Banana Apple, C says Chook Banana Apple.. It's a game kids play in a ring.. A word game. They have to remember the information and manipulate it such that its reversed.. And they have to come up with a new word too. Not a working memory task if I go first, but it is for the person that goes next.

D: I don't actually know how you would operationalise that task in a real world situation.

L: we might leave that Gwm item. So let's move on to Gl - what sort of behaviours would a child display if they had a high or low level of Gl?

D: High Gl - they'd be quick learners, wouldn't need a lot of trials to learn, would get the sound-symbol thing very quickly, like that alphabet principle, connecting sounds and symbols… because their associative memory is strong so they can link sounds and symbols which is key for literacy. They'd need less trials to comprehend and they'd be able to bring all the information to new situations fairly quickly… so it would just decrease, the speed at which, or the number of times they'd have to be exposed to something before they could get it

L: link unrelated concepts to learn new information.. I guess that if we never learnt that the colour red means stop, and the colour green means go. These colours and these concepts are unrelated, but it's been ingrained in us that they are related

D: i think you would need to explain it to me to understand as a parent… red = stop and green = go, as a parent, it's almost because the parent would assume that red is stop and green is go that it doesn't make sense that it is an "unrelated concept"

L: So you think may be a better example of unrelated concepts…

D: It's a good example but it's already so entrenched in our culture that it might not be immediately obvious to parents that the colours and "stop and go" are actually not related things. So are you looking at associative memory there

L: Yes so using one concept to learn another concept. Lot of the reading that I've been doing shows associative memory in the contact stuff you know using smell or a song to associate with a past memory which is not academic, so I couldn't use that example for associative memory

D: So like in the academic tests, for WJ like one particular symbol is associated with a word is that what you are going for? So it's sort of like sounds and symbols. In actual learning of literacy, that symbol, that formation of the "a" symbol, that formation = b.

D: So that's really how associative memory helps with reading, because you associate that particular formation of the letter, with that particular sound, does that make sense?

L: I'm just thinking whether that's also part of auditory processing as well.
D: Well you have to be able to hear it, processing, that apple starts with a. I dont have to know what it has to look like to write it. If i draw what we know as a, I know what I'm drawing is associated with the sound /a/. When I draw p, I know it's associated with the sound /p/. So when i see "ap" together, i know what that sounds like. But to have Ga i don't need to know what it looks like to know it.

L: so it's the visual of the symbol that triggers the memory of the sound

D: im associating the symbol with the sound, so i have to think of both of them together.. E.g., knowing the symbol "a" has an /a/ sound. So that's how associative memory is used in the classroom. Maybe have a young person example and older person example

L: associate novel information with existing information (e.g. mnemonics)

D: can't have the word mnemonics there. Provide an example instead. It's quite hard to operationalise this item, is it still associative memory? I'm thinking if associative memory is strong… if i'm good at it, then i only need to be exposed to it a couple of times before I learn it. Nothing captured there about how many trials someone might need to grasp a concept.

L: so you think i need an item that says something about how many trials a child requires to learn something?

D: yes. Because i might be exposed to five new sight words, i have good Gl so i can remember those words tomorrow. But some parents might say “i showed my kid these words yesterday but he can’t remember them today” so some kids require more exposure and learning trials to remember something, or to consolidate it if you’d like. Maybe talk to Kate about it.

L: I do think the exposure and learning trials is an important aspect that I haven’t addressed in the current items so I’ll definitely look into it

D: Can’t have “principal components” - have changed it.

L: (item) Remember the order in which events of the day have occurred

D: what’s that assessing?

L: still meaningful memory, in the way that events of the day is a narrative.

D: Is the sequence important? (yes) so is it more like an important event happened? Can you tell me the sequence of your day? Are you addressing MM here? (yes)

L: (definition of MM)

D: events of the day seems too broad to me, but that’s just me. Because some parts of the day might be more meaningful than others

L: how would you ask this question then?

D: is it the same thing as perhaps a plot of a movie? As in tell me the plot of a movie in an ordered way?

L: well yeah it’s still a narrative that a child would remember if he or she has seen it before and it’s different to an orally presented story, in that the child has received the information in a different way (visual rather than verbal)

L: ok last one – this was another associative memory issue I was trying to address

D: I like that, but you might have to provide examples for the acronyms, like ROYGBIV for colours of the rainbow – I think parents would get that then. And mental imagery.. have to give example for that too

L: so that’s all we have for Learning Efficiency… let’s move onto Retrieval fluency. What sort of behaviours would you expect to see from a child who has high or low levels of Gr?

D: I think with low-level… they’d be kids with fillers in their language and there would be a delay in response when you ask them a question. They may take a long time to copy from the white board um they may take a long time to get their books etc out. I think some teachers would probably think they might be in a dream a lot of the time.. and unfortunately they are kids that may know things but don’t get the chance to show them because someone else would jump in to answer before them.

D: My mum had dementia and I had picked her up from the hostel, and asked her what she had for lunch. I lived about 40km from the hostel, halfway home, about 20k home she would say “chicken sandwich” and I’d go “… what?” and then I realised that she was processing and retrieving information in her memory to answer the question.

L: Quickly name objects presented on a page

D: I know what you're getting at there, and I think a parent would probably work out if their kid can do that or not... they may not do it right now

L: I have another item that is sort of related to that, instead of objects, it's rapidly name letters.

D: yep that's fine

L: another one that I have – rapidly name examples from a given category within a time limit

D: that's good... given category, can we change that? Oh.. nah that will do. I should be able to understand that if I was a parent. Same category could be nicer. And "things you can eat and drink" sounds a bit like WJ COG so change that to maybe "fruits and vegetables"

D: this next one is funny… what is a "known prime", what does that mean?

L: in my mind I've worked it out to be, if a child were to see an image or hear a word that relates to something that they've already learnt, they'd be able to retrieve that information

D: ok so come up with an example for that.

L: I don't actually know if this happens in the classroom… but a sort of conceptual mind-map is what I was going for. If the centre word was "dinosaur" and it branches out to things like the different periods, the different kinds of dinosaurs, etc.

D: could you say "quickly recall information when provided with a ..cue?" so you're asking them to recall… how quickly are you asking them to recall, does it matter? It's like word associations isn't it? Like… that's what you're after. If I say to you, a colour, how quickly can you name a heap of colours, or how many gemstones can you name…

L: yes, so the word you give to the child is the cue

D: "quickly retrieve connected information when presented with a cue" would a parent know what a cue was? E.g., weather – thunderstorm, sun, cloudy, rain, temperature, seasons

L: ok let's move onto visual processing.. so what behaviours would a child display if they had high/low Gv

D: high – often they might be quite artistic or good with visual-motor skills, they might be good at construction, if not artwork. Often they would be good at map reading and many aspects of mathematics. Perhaps less impact on literacy as we know it at the moment although orthographics is a very interesting thing. Um.. low I think, often clumsy kids would have low Gv, can't judge distances between you and the door and push right through you, and they'd also be poor at mathematics. They may not be very good at judging faces and not good at reading social cues.. and obviously they'd be bad at construction, map-reading, things like that.

L: not be distracted by irrelevant visual information – what do you think?

D: need example, what is irrelevant visual information

L: what I was trying to look for is.. is the child able to like, look at a page that is full of pictures and words, and be able to look for a certain specific information, and not be distracted by "oh that's a nice picture"

D: what about "Where's Wally?" that's a good one. I think 'irrelevant visual information" is hard to wrap your head around as a parent

D: Put parts together to form a whole – what does that mean

L: I was re-wording Block Design in a different way (I have an item that addresses that already) so not just using blocks to put things together. If there was an image for example that was in parts, whether the child would be able to put the parts of the image together to form the whole image

D: see that to me is a jigsaw – that task is essentially putting a jigsaw together. I have the picture of the completed jigsaw, I have the pieces, I need to put the pieces together to form the picture.

L: so if I put e.g., complete a jigsaw puzzle?

D: or put parts together to form a whole picture… yeah that doesn't make sense still. Or put the parts together to form the whole… but a whole what? I mean you might be doing it in your head as well. But the parent can't judge that anyway because it's happening in the child's mind

L: I think "completing a jigsaw puzzle" would be a good representation of how the parent might be able to judge the child's ability in that

D: or wooden puzzles, construction puzzles… or those metal things that you have to put together and take apart… or rubic's cubes..

L: understand information presented in a visual format – all good

L: mentally rotate images in his/her mind

D: how would the parent know that?

L; unless there is a task that shows the ability of the child being able to mentally rotate images…

D: or when they have to… oh you know! Tetris? Because sometimes in tetris you have to rotate or flip the blocks over as they're coming down for it to be able to fit. And I think a parent would get that

L: find specific information on a printed page – had comments from expert panel that people don't really use phonebooks these days.

D: or dictionaries either.

L: in those like "reading comprehension" questions where they ask you to find the word in the passage and figure out its meaning within the context

D: you need to emphasise that you mean "find the word" not "tell me the meaning of the word" for it to be Gv. What about "wonderwords" like where you've got words within a grid and you have specific words to find within random letters. Teachers sometimes give them wonderwords – maybe if you Google it – you've actually got to identify the word, the words can go across, diagonally or vertically. That to me is Gv.

L: so what sort of behaviours would a child display if he had high/low levels of Ga?

D: high – quite good at music, they would maybe have perfect pitch? Phonics would be very sound, may have early literacy acquisition, might gain literacy very quickly. Guessing they would be good speakers as well. Low – struggle with phonics big time, impact on auditory working memory as well, struggle in the classroom..

L: so the items that I had on the previous version of the ECCA got mostly good ratings, so I only had a few items for us to talk about today. (first item) I know that sounds quite tricky and technical. I originally had "blend letter sounds fluently together to form words" that some panel members commented wasn't clear, so I was trying to emphasise that the child had to hear these letter sounds, blend them tgt, and then form words.

D: what you could do is find out how kindergarten or prep teachers call that. And they would probably have specific names for that. And you call it sound blending…

L: the trick is to make sure parents understand that as well, not just a technical term that teachers would use.

D: yeah the item wording as of now, I don't like that. Able to hear sounds and blend them into words? E.g. kindy parents at least would sort of understand where you're going with this because they would hear it a lot from teachers. And auditory processing is really going to tap into younger ones more than older ones isn't it. You are going to hit the ceiling with Ga quicker than some of the other ones I believe.

L: hear the internal structure of sounds in words

D: wow… how are the parents going to know that?

L: this was "hear the difference in sounds in words" which is still addressing the very basic phonics structure of words. I figured that phonics is a technical term so tried to change it to internal structure

D: how do you know if they hear it or not? Like.. break words into simpler sections maybe. What you're asking for there is – if I go "basket", I can actually break it down into bas-ket, and then b-as-k-et. That's what you're asking me? Maybe.. identify the different sounds in words? Teachers would know this item well because they often ask kids to identify the beginning, middle and ending sounds in words, parents would struggle more with that.

L: so if I were to change that "identify the beginning, middle and ending sounds in words" would that be much clearer?

D: for teachers definitely.

L: identify a word when only parts of the word is pronounced

D: yep I know where you're trying to go with that one too, but it's really hard! This is a performance task.

L: do you think it's still relevant in this questionnaire? Would people still be able to answer this question?

D: I wouldn't know… unless they actually say "para-ute" what am I saying here? Do you have items that address rhyming words and things (yes) Ok I really don't think there is a real-life example of this task here

L: yep fair enough I don't know if parents and teachers can answer this question

L: ok so high and low levels of Gs, what does that look like?

D: similar to Gr in a way… I think they overlap a bit. I think RAN has a bit of Gs in it. I think it is similar to Gs. And.. I talked about the chicken sandwich thing. Getting their questions answered before getting the chance to answer them. They might learn well, but they take it a while to get it out. High Gs – high visual-motor skills.

L: (1st item) D: sounds ok L: just with our discussion about Gv, I feel now that this item has something to do with Gv

D: it's all going to be related a little bit, but it's still that ability to do stuff, routine things that you already know how to do, but to do it quickly. Maybe include an example

D: 2nd item – I'd reword basic mathematical calculations, to a series of maths sums… or something like that. Simple maths problems or sums, or whatever. E.g. 3+2

D: 3rd item – yeah what does simple visual patterns mean? How am I measuring that? Is this like coding... or the circle and the star together... or pairs of digits together

L: again it's a very performance-based task so I don't know if it happens in the classroom. That might be an item that's not very relevant maybe... will have to think about it

D; it's like speed stuff isn't it, like speed drills. I wonder… have you got any other questions for Gs?

L: (lists other Gs items) change whiteboard to smartboard. Most teachers would use smartboards

D: also include 'respond to simple questions in a timely fashion" tapping into that young person who takes a lot longer respond to a question than other children

Appendix D

Definitions of CHC Broad Cognitive Ability Areas with Corresponding Narrow Ability Areas

These definitions presented here are adapted from Schneider, W. J., & McGrew, K. S. (in press). The Cattell-Horn-Carroll Theory of Cognitive Abilities. In D. P. Flanagan & E. M. McDonough (Eds.), Contemporary intellectual assessment: Theories, tests and issues (4th ed.,) New York: Guilford Press.

***Fluid Reasoning (Gf)*:** The use of deliberate and controlled procedures (often requiring focused attention) to solve novel "on the spot" problems that cannot be solved by using previously learned habits, schemas, and scripts

- Induction (I): The ability to observe a phenomenon and discover the underlying principles or rules that determine its behavior. This ability is also known as rule inference.
- General sequential reasoning (RG): The ability to reason logically using known premises and principles. This ability is also known as deductive reasoning or rule application.
- Quantitative reasoning (RQ): The ability to reason with quantities, mathematical relations, and operators.
- Reasoning Speed (RE): The ability to reason with quantities, mathematical relations, and operators.
- Piagetian Reasoning (RP): Seriation, conservation, classification and other cognitive abilities as defined by Piaget's developmental theory.

Comprehension-Knowledge (Gc): The ability to comprehend and communicate culturally-valued knowledge. Gc includes the depth and breadth of both declarative and procedural knowledge and skills such as language, words, and general knowledge developed through experience, learning and acculturation

- Language Development (LD): An intermediate stratum ability to comprehend and communicate using language. The general understanding of spoken language at the level of words, idioms, and sentences.
- Lexical knowledge (VL): The knowledge of the definitions of words and the concepts that underlie them. Vocabulary knowledge.
- General (verbal) information (K0): The breadth and depth of knowledge that one's culture deems essential, practical, or worthwhile for everyone to know.
- Listening ability (LS): The ability to understand speech. This ability starts with comprehending single words and increases to long complex verbal statements.
- Communication ability (CM): The ability to use speech to communicate effectively.
- Grammatical sensitivity (MY): The awareness of the formal rules of grammar and morphology of words in speech.

***Short-Term Memory (Gsm)*:** Refers to the ability to apprehend and hold information in immediate awareness and then use it within a few seconds.

- Auditory short-term storage (Wa): The ability to encode and maintain verbal information in primary memory.
- Visual-spatial short-term storage (Wv): The ability to encode and maintain visual information in primary memory.
- Attentional Control (AC): The ability to manipulate the spotlight of attention flexibly to focus on task-relevant stimuli and ignore task irrelevant stimuli. Sometimes referred to as spotlight or focal attention, focus, control of attention, executive controlled attention, or executive attention.
- Working memory capacity (WM): The ability to manipulate information in primary memory. Technically not a narrow ability (WMC = short-term storage + AC).

Learning efficiency (Gl): the ability to learn, store and consolidate new information over periods of time measured in minutes, hours, days, and years

- Associative memory (MA): The ability to form a link between two previously unrelated stimuli such that the subsequent presentation of one of the stimuli serves to activate the recall of the other stimuli.
- Meaningful memory (MM): The ability to remember narratives and other forms of semantically related information.
- Free recall memory (M6): The ability to recall lists in any order.

Retrieval fluency (Gr): the rate and fluency at which individuals can access information stored in long-term memory

- Ideational fluency (FI): The ability to rapidly produce a series of ideas, words, or phrases related to a specific condition or object.
- Expressional fluency (FE): The ability to rapidly think of different ways of expressing an idea.
- Associational fluency (FA): The ability to rapidly produce a series of original or useful ideas related to a particular concept.
- Sensitivity to problems/alternative solution fluency (SP): The ability to rapidly think of several alternative solutions to a practical problem.
- Originality/creativity (FO): The ability to rapidly produce original, clever, and insightful responses (expressions, interpretations) to a given topic, situation, or task.
- Speed of lexical access (LA): The ability to rapidly retrieve words from an individual's lexicon. Verbal efficiency or automaticity of lexical access. An intermediate stratum level ability.
- Naming facility (NA): The ability to rapidly call objects by their names.
- Word fluency (FW): The ability to rapidly produce words that share a phonological (e.g., fluency of retrieval of words via a phonological cue) or semantic feature (e.g., fluency of retrieval of words via a meaning-based representation).
- Figural fluency (FF): The ability to rapidly draw or sketch as many things (or elaborations) as possible when presented with a nonmeaningful visual stimulus (e.g., a set of unique visual elements).
- Figural flexibility (FX): The ability to rapidly draw different solutions to figural problems

***Visual Processing (Gv)*:** the ability to make use of simulated mental imagery to solve problems. Perceiving, discriminating, and manipulating images in the 'mind's eye'

- Visualization (Vz): The ability to perceive complex visual patterns and mentally simulate how they might look when transformed (e.g., rotated, changed in size, partially obscured, and so forth).
- Speeded rotation (SR): The ability to solve problems quickly using mental rotation of simple images. This ability is similar to Vz but is distinct because as it involves the speed at which mental rotation tasks can be completed.
- Imagery (IM): The ability to voluntarily mentally produce very vivid images of objects, people or events that are not actually present.
- Closure speed (CS): The ability to quickly identify and access a familiar, meaningful visual object stored in long-term memory from incomplete or obscured (e.g., vague, partially obscured, disguised, disconnected) visual cues of the object without knowing in advance what the object is.
- Flexibility of closure (CF): The ability to identify a visual figure or pattern embedded in a complex distracting or disguised visual pattern or array, when one knows in advance what the pattern is.
- Visual memory (MV): The ability to remember complex visual images over short periods of time (less than 30 seconds).
- Spatial scanning (SS): The ability to quickly and accurately survey (visually explore) a wide or complicated spatial field or pattern with multiple obstacles and identify a target configuration or identify a path through the field to a target end point.
- Serial perceptual integration (PI): The ability to recognize an object after only parts of it are shown in rapid succession.
- Length estimation (LE): The ability to visually estimate the length of objects (without using measurement instruments).
- Perceptual illusions (IL): The ability to not be fooled by visual illusions.
- Perceptual alternations (PN): Consistency in the rate of alternating between different visual perceptions.
- Perceptual speed (P): See definition under Gs. P has a secondary loading on Gv.

***Auditory Processing (Ga)*:** ability to discriminate, remember, reason, and work creatively (on) auditory stimuli, which may consist of tones, environmental sounds, and speech units

- Phonetic coding (PC): The ability to distinctly hear phonemes, blend sounds into words, and segment words into parts, sounds, or phonemes.
- Speech sound discrimination (US): The ability to detect and discriminate differences in speech sounds (other than phonemes) under conditions of little or no distraction or distortion.
- Resistance to auditory stimulus distortion (UR): The ability to hear words or extended speech passages correctly under conditions of distortion or background noise.
- Maintaining and judging rhythm (U8): The ability to recognize and maintain a musical beat.
- Memory for sound patterns (UM): The ability to retain (on a short-term basis) auditory codes such as tones, tonal patterns, or speech sounds.
- Musical discrimination and judgment (U1 U9): The ability to discriminate and judge tonal patterns in music with respect to melodic, harmonic, and expressive characteristics (phrasing, tempo, harmonic complexity, intensity variations).
- Absolute pitch (UP): The ability to perfectly identify the pitch of tones.
- Sound localization (UL): The ability to localize heard sounds in space.

Processing speed (Gs): the ability to control attention to automatically, quickly and fluently perform relatively simple repetitive cognitive tasks. Attentional fluency or attentional speediness

- Perceptual speed (P): An intermediate stratum level ability that can be defined as the speed and fluency with which similarities or differences in visual stimuli (e.g., letters, numbers, patterns, etc.) can be searched and compared in an extended visual field.
- Perceptual speed-search (Ps): The speed and fluency of searching or scanning an extended visual field to locate one or more simple visual patterns
- Perceptual speed-compare (Pc): The speed and fluency of looking up and comparing visual stimuli that are side-by-side or more widely separated in an extended visual field.
- Number facility (N): The speed, fluency and accuracy in manipulating numbers, comparing number patterns, or completing basic arithmetic.
- Reading speed (fluency) (RS): The speed and fluency of reading text with full comprehension. Also listed under Grw.
- Writing speed (fluency) (WS): The speed and fluency of generating or copying words or sentences. Also listed under Grw and Gps

Appendix E

Summary of Results of Cognitive Interviewing with Educational Specialists

General comments

- Ambitious to cover primary and secondary in the same survey. Having separate primary and secondary versions is a good idea because primary teachers are usually very generalist, while secondary teachers are specialists
- Teachers are not taught about developmental disorders like SLD or CAPD, if they don't know about it, they're not going to be observing the behaviours that are characteristic of the disorders or they would be interpreting them differently
- Suggestion to change rating scale – teachers would find it easier if you said 1 = below average ability, 3 = average ability, 5 = above average ability. Neither difficult nor easy is not the same as average.

Table E1

Responses from Cognitive Interviewing with Educational Specialists (N=2)

Survey question	**Probes**	**Responses (A)**	**Responses (K)**	**Changes**
FLUID REASONING (Gf): the use of deliberate and controlled focused attention to solve novel, "on the spot" problems that cannot be solved solely by using prior knowledge; reasoning that depends minimally on learning and acculturation **General comments:** Pictorial diagrams were useful. Gf 3 and 5 are very "performance-based", IQ test-like items, not sure if teachers could answer that.				
Demonstrate problem solving skills when faced with a problem	a) Can you repeat the question in your own words? b) What, to you, is "problem solving skills"? c) How sure of you that the student can display this behaviour?	I'm very much maths problem solving skills. I tried to generalise so I thought I'd go neutral. Wasn't sure about the answer. Maybe give an easy Raven's matrix type question as an example like what you did in the other items.	has ability to confronted with problem, come up with a solution is good at finding strategies when confronted with a problem	Come up with a solution when confronted with a problem

Table E1 continued				
Figure out the number that comes next in a number series	a) How sure are you of your answer? b) How hard was this to answer?	No problems. The child I'm thinking of wouldn't have a clue how to do that.	Easy to answer, teachers would find that easy too.	No change
Select the correct missing piece to complete a logical puzzle	a) What, to you, is a logical puzzle? b) Is the illustration here an example of a logical puzzle? c) How sure are you of your answer?	Not sure if teachers would know if a child is able to do the task.	Too much like an assessment task – not the kinds of things teachers do with kids	Can think logically to solve a problem e.g. could select the right piece in the puzzle below
Identify and perceive relationships between concepts	a) Can you rephrase this question in your own words? b) How hard was this to answer?	Same issue as previous	Same issue as previous. Teachers don't give them questions like that.. they might say the child is able to see the relationship between, or can classify"	Identify relationships between objects and concepts (see example below)
Sort and classify objects according to a given category	a) What, to you, does "given category" mean b) How sure are you of your answer?	Not difficult to answer at all	Wording is fine. This item should come before number 4 because it's easier, and can use it as a lead on	No change
Use previously acquired knowledge to solve new problems	a) How well do you think this item explores the intended construct? b) How did you get the answer?		I think teachers would know how to answer that.	No change
CRYSTALLISED KNOWLEDGE (Gc): to have a large amount of acquired knowledge language ability, and to be able to effectively use that knowledge; comprehension of language, words and general knowledge developed through experience, learning, and acculturation General comments: All items were easy to answer, teachers should be able to provide responses easily. Behaviours should be easily observable in the classroom.				
Provide synonyms and antonyms of a word	a) Was this item easy to answer? b) How sure are you of your answer?			No change
Provide oral definitions for words	a) Can you re-phrase the question in your own words? b) How do you remember this?			No change

Table E1 continued

Provide correct responses to questions of fact	a) What does “questions of fact” mean to you?		“Questions of fact” seems very vague. Example also restricted to art which not everyone is interested in so they might not know	Demonstrate good general knowledge about the world around them (e.g., able to answer "what is used to make paper")
Correctly name pictures of familiar objects	a) What does “pictures of familiar objects” mean to you? b) How sure are you of your answer?			No change
Understand others’ conversations	a) Was this item easy to answer? b) Is this item is important to the construct of crystallised knowledge?		Understanding conversation isn’t enough... the child needs to engage, it’s being able to reciprocate and participate in the conversation	Engage in conversation with others (peers and adults)
Display general knowledge of learnt social rules	a) Are there any words or phrases that make this item difficult to understand? b) Can you repeat the question in your own words?			No change
Communicate effectively what he/she is trying to say	a) What, to you, is “communicate effectively”? b) How did you get the answer for this item?			No change
Identify similarities between two common objects	a) How sure of you are your answer? b) How do you remember this?			No change

Table E1 continued				
SHORT-TERM WORKING MEMORY (Gwm): the ability to apprehend and hold information in immediate awareness and then use it within a few seconds				
Repeat numbers or words orally in the same order as presented	a) How did you arrive at your answer? b) What are some examples of classroom activities whereby the student would have to complete this task?	She's got dyscalculia so she can't count, can't remember numbers or words at all. I knew this because I tutored her. Primary school teachers play a game, Peter goes to the shop and buys an orange (next kid) Peter goes to the shop and buys an orange and a banana (and so on)	I knew you were asking about working memory and I knew the student has difficulty in this area so I just marked it low. Teachers wouldn't necessarily know what you're asking though. the only time they would be asked to repeat numbers or words orally (Gwm1) would be in a test situation, teachers don't normally do this in the classroom.	Discard
Remember a verbally spoken phone number for long enough to note it down on paper	a) Was this hard or easy to answer? b) How sure are you of your answer?	Fine	You don't give students mobile numbers so you need to change the example here	Remember the numbers given in a verbally spoken math problem for long enough to note it down on paper
Listen to and repeat a list of words in correct order	a) How did you arrive at your answer?	Teachers would do this in class.	I have seen this behaviour in classrooms so it would be ok for teachers to answer this	No change
Follow instructions to do two or more tasks (e.g., pick up your jacket and put away your lunch)	a) How sure are you of your answer? b) How do you remember this?	Yes this is something that teachers would definitely see in the classroom	Easy to answer, easy to pick out kids who have issues with this	No change
Write down dictated sentences accurately, ignoring spelling and punctuation errors	a) Was this hard or easy to answer? b) What, to you, is "writing down dictated sentences accurately"?	That was relatively easy to answer for me, because I know that the child would find it difficult since she has to remember all the different sentences	Fine	No change

Table E1 continued				
Accurately solve verbally presented maths sums by using mental arithmetic (i.e., no pencil and paper)	a) Were there any words in this item that made it difficult to answer?	Teachers equate learning math with quick mental math so be careful not to emphasise speed here	Fine	No change
LEARNING EFFICIENCY (Gl): the ability and efficiency to learn, store, and consolidate new information in long-term memory General comments: can use a range of strategies to solve problems in different situations, or to help them with their learning. associative strategies to learn new information is one thing, but when they're presented with something, do they know what strategies to apply?				
Listen to and retell important parts of an orally presented story	a) Was this question easy to answer? b) What, to you, are "important parts of a story"	Yes this was easy.	Easy, teachers would do this a lot	No change
Remember the order of events that happened in a movie or TV show	a) How sure are you of your answer? b) Is this item relevant to the construct of learning efficiency?	Fine to answer	I think kids watch movies or parts of movies often in an English lesson, the teachers will show them a clip of something and ask them questions about it	No change
Link unrelated concepts to learn new information (e.g., linking language symbols and sounds to learn how to read)	a) Were there any words/phrases that made this item difficult to understand? b) What opportunities do students have to perform this behaviour, aside from in the given example?	Better example?	In schools this is known as print knowledge, I would say here "has poor print knowledge, or has poor sound-symbol relationship"	Learn and remember the relationship between unrelated objects (e.g., associating a face with a name; associating the visual symbol "a" with its corresponding sound "ah")

Table E1 continued

Use associative strategies to learn new information (e.g., acronym ROYGBIV for colours of the rainbow)	a) Can you rephrase this item in your own words? b) How sure are you of your answer?	Include another example for visual learners like using mindmaps for example	teachers use the word mnemonics, so they would recognise ROYGBIV as mnemonics Can apply effective strategies to learn new information - I would say that - for example colours of the rainbow, or ones that are often used in class are things like bears eat apples under trees for remembering the beginning of beautiful or ants never yawn for "any"	Apply effective strategies to learn new information (e.g., using mnemonics to learn difficult spellings, using mindmaps)
Remember new information from a small number of learning trials (e.g., able to remember 5 new sight words in 2 days, rather than a week)	a) Was this item easy to answer? b) Can you rephrase this in your own words? c) How did you arrive at your answer?	I thought that had to do with phonemic awareness and sigh reading. Give an example for visual, but I can't think of one.	I'm not sure if this is related to learning efficiency.. if the child memorises the sight words on Thursday for Friday's spelling test, he might do really well on it. But then when you test him on Monday it's gone. A person might seem to have learnt something after a couple of trials, but doesn't mean it has been encoded into LTM	Retain new information after only a few learning trials (e.g., learning a spelling test)
Retell narratives and other forms of semantically related information	b) Were there any words in this item that made it difficult to answer?	Didn't understand what semantically related information was	What is the relation between retelling narratives and "semantically related information"? Very technical and jargony	Discard
Recall facts about what he/she has read in a book	a) How sure are you of your answer? b) Do you think this is important in measuring learning efficiency?	Sure of my answer, it's a relatively simple behaviour to observe. Yes important in learning.	No issues with responding to item	No change

Table E1 continued

RETRIEVAL FLUENCY (Gr): the rate and fluency at which individuals are able to access information stored in long-term memory General comments				
Quickly recognize letters presented on a page	a) Was this item easy to answer? b) How did you arrive at your answer? c) Do you think these items are relevant to the construct of retrieval fluency?	No issues	Change to quickly recognise and name	Quickly recognise and name objects presented on a page
Quickly recognize objects presented on a page				Quickly recognise and name letters presented on a page
Quickly recognize whole words presented on a page				Quickly recognise and name whole words presented on a page
Rapidly give numerous examples from the same category (e.g., fruits and vegetables)	a) Examples of classroom activities similar to this task? b) How did you arrive at your answer?	It's a game that teachers might play with kids so it was ok to anwer	yes teachers would understand that. I do this often with teachers, good to give kids practice of saying "in one minute name as many items you can beginning with letter s" and parents do that a lot with their kids	No change
Quickly retrieve connected information when presented with a visual or verbal cue (e.g., seeing an image, hearing a word)	a) Were there any words or phrases in this item that made it difficult to answer? b) How do you remember this?	Difficult to answer… provide an example "concept map"	I think it's okay.. I understood what you meant when I read it but I had to read it a couple of times.	Quickly name connected information when presented with a visual or verbal cue (e.g., seeing an image or hearing a word in a concept map)
Quickly give examples of words that share a semantic feature (e.g., fruits that grow on trees)	a) What, to you, is "semantic feature"? b) Can you rephrase this item in your own words?	I see that as a neuropsych question.. a primary school teacher would not know if a child can answer that	You've given the e.g., so that's fine.	No change

Table E1 continued				
Quickly recall information related to a learnt topic	a) Were there any words or phrases that made it difficult to answer this item?	No, it was fine	That's fine	No change
VISUAL PROCESSING (Gv): the ability to think with visual patterns and stimuli. Includes the ability to mentally rotate, reverse, and manipulate spatial configurations, and spatial orientation. General comments: extract visual information from graphs or a handout (added item)				
Build a model (e.g., with Lego or blocks) from a picture of a completed model	a) Was this item easy to answer? b) How did you arrive at your answer?	Easy to answer	Easy to answer	No change
Put parts of an image together to form the whole image (e.g., jigsaw puzzle)	a) Can you think of other classroom activities that require the successful completion of this behaviour?	Easy to answer	Easy to answer	No change
Differentiate between similarly shaped symbols	a) What, to you, are examples of "similarly shaped symbols"? b) Was this question easy or difficult to answer?	Letters like b and d maybe. No issues with answering	Plus and multiply signs. Easy to answer	Differentiate between similarly shaped symbols (e.g., plus and multiply signs)
Not be distracted by irrelevant visual information (e.g., playing "Where's Wally")	a) What, to you, is "irrelevant visual information" in a typical classroom task? b) Was this item clear and easy to understand?	I guess like… some teachers make their own handouts with a lot of pictures in the borders for example, that would be irrelevant visual information	I saw it as being like a visual search, find the things hidden in the picture. As opposed to being worded negatively, I would say "can pick out relevant information/objects in a picture e.g. where's wally"	Pick out relevant information in the presence of distracting visual stimuli (e.g., playing "Where's Wally")
Interpret information presented in a visual format (e.g., reading graphs, charts, maps, etc.)	a) Was this item easy to answer? b) Were there words/phrases here that made it difficult to understand?	Easy to answer	Extracting information and interpretation information is different	No change

Table E1 continued				
Accurately judge distances between objects (e.g., not placing objects too close to an edge)	a) How did you arrive at your answer? b) Can you re-phrase this item in your own words?	I guess if you saw the kid bumping into things a lot you would say that he has difficulties in this area.	What I thought you might have been getting at here was like.. if you had a bunch of beans clumped together, counted them and there's 6. And then if you spaced them apart and asked "how many are there now, are there more or less or the same" they would say more, because of the space between the objects.	No change
Find specific information on a printed page or computer screen (e.g., finding a specific word in a passage of text)	a) Was this item easy/difficult to answer? b) How sure are you of your answer?	Easy to answer	Yes that's an easy one	No change
AUDITORY PROCESSING (Ga): the ability to notice, compare, discriminate, and distinguish distinct and separate sounds. **General comments**: Do you want to know anything about sound manipulation? Something like say rip, what would you have when you added "t" to the beginning of rip? Say "split" without the /l/ the ability to manipulate sounds within words? *Is it something that teachers do with their students?* No, but it is relevant to phonological processing. So maybe you're right it might not be a good thing to ask the teachers cause they might not know. They should be doing it though as part of phonemic awareness.				
Understand what is being said when background noise is present	a) Was this item easy to answer? b) How did you arrive at your answer?	These items are all to do with phonological awareness and auditory discrimination. Because I know this child has auditory processing difficulties, I found these questions really easy to do. But her teacher didn't, she was diagnosed with CAPD but before that her teacher just thought she was a very quiet child. I wouldn't think these items would be picked up by the average teacher who	Easy to answer, most teachers would know	No change
Hear the difference between similar sounding words (e.g., rhyming words)	a) Can you rephrase this item in your own words? b) How sure are you of your answer?		I thought if it was auditory, it might be similarly sounding letters like /g/.. oh but you have that in 44 so that's ok	No change
Differentiate between speech sounds (e.g., difference between /ch/ and /sh/)	a) What, to you, are examples of "speech sounds"? b) Was this question easy or difficult to answer?		Speech sounds might not be a term that they know.. but since you have the example there it should be fine.	No change

Hear letter sounds and blend them fluently to form words	a) How did you arrive at your answer? b) Was this item easy to answer?	isn't taught about phonological awareness or dyslexia. They probably would still observe these behaviours but they'd probably think "oh she will grow out of it" because there is poor awareness about developmental disorders and SLDs initially, in grades 1-2, before they have any kind of alert, they would just think "kids progress at different rates" so the kids go under the radar.	What did you mean by "hear" though? So if you say "can blend letter sounds into words" I think they'd just understand that as opposed to "hear". *Hearing is a bit redundant?* Are you wanting to know if they can blend? If so then that's enough.	Blend letter sounds fluently to form words
Know which sounds go with which letters	a) Can you rephrase this item in your own words? b) How do you know if the child is capable of this task?		Something to do with the letter-sound relationship? But what you have here is fine.	No change
Filter out background noise to listen to my voice in the classroom	a) How did you arrive at your answer? b) Can you re-phrase this item in your own words?			No change
Understand what is being said when background noise is present	a) Was this item easy/difficult to answer? b) How sure are you of your answer?			No change
Identify the beginning, middle, and ending sounds in words	a) Was this item easy/difficult to answer? b) How did you arrive at your answer?		I'd put an example there	Identify the beginning, middle, and ending sounds in words (e.g., what's the first sound in cat? What's the last sound in run)
Sound out unfamiliar words	a) How sure are you of your answer? b) Is this item important to the construct of auditory processing?		I was very sure, I might put in brackets, ability to decode. *So teachers would know decode as a term?* Yes decode is a word they use a lot.	Sound out unfamiliar words (ability to decode)

Table E1 continued				
VISUAL PROCESSING (Gv): the ability to think with visual patterns and stimuli. Includes the ability to mentally rotate, reverse, and manipulate spatial configurations, and spatial orientation. General comments: could add "completes tasks at the same time as their peers"				
Quickly complete a series of simple maths sums (e.g., 3+2, 5+4, etc.)	a) Was this item easy to answer? b) How well do you remember this?	Teachers wouldn't have difficulty answering that	Easy to answer, but did you mean written down or given verbally?	Quickly complete a series of simple maths sums on a worksheet (e.g., 3+2, 5+4, etc.)
Quickly and accurately copy information from the board	a) How did you arrive at your answer?			No change
Respond to simple questions quickly	a) What, to you, is a "simple question"? b) How did you arrive at your answer?	Provide an example. Like what did you have for breakfast?		Respond to simple questions quickly (e.g., what did you have for breakfast)
Quickly identify differences and similarities between two objects	a) When would students encounter this task in the classroom?	When they are asked to differentiate between different 3D objects, like a cube or a cuboid	That's fine	No change
Work quickly and efficiently on routine tasks	a) What, to you, is "working quickly and efficiently"? b) What are some examples of "routine tasks" you can think of?	I think a teacher would know that.	Getting ready for a lesson, or a science experiment in secondary school… at home it would be things like getting ready for school or to go out	Work quickly and efficiently on routine tasks (e.g., getting ready for a lesson)
Scan and quickly determine important information on a page	a) Was this question easy to answer? b) How sure are you of your answer?	Easy to answer	No problems	No change

Appendix F

58-item Estimates of Children's Cognitive Abilities Measure

Survey link: https://monash.az1.qualtrics.com/jfe/form/SV_cIJY4hITyMkVfRb

Description

Thank you for participating in the Estimates of Children's Cognitive Abilities questionnaire, developed to obtain teacher-reports of children's cognitive abilities.

You have been invited to take part in this questionnaire because you are currently a teacher of a school-aged child aged between 5 to 18 years. This survey contains questions about your opinion on the ability of a particular child in your classroom to complete certain cognitive tasks. This is expected to take 10 to 15 minutes, and can be completed at a time and place of your own convenience.

Involvement in the study is voluntary and you are under no obligation to consent to participation. Participation in this research is completely anonymous. Under no circumstances will your responses be able to be identified as yours, nor will you be asked to divulge information of a personal or sensitive nature.

If you choose to provide your email address at the end of the questionnaire, you will enter a raffle to win a $50 Coles Group & Myer gift card. As your email address is not linked to your responses, your responses will remain anonymous.

If you have any questions, please contact me at:
Lydia Soh
lydia.s.soh@monash.edu
03 9902 4891

Please click 'Next' to begin.

Teacher demographics

Q1 What is your gender?

- ☐ Male
- ☐ Female
- ☐ Unspecified

Q2 What state do you work in?

- ☐ NSW
- ☐ VIC
- ☐ TAS
- ☐ QLD
- ☐ WA
- ☐ SA
- ☐ ACT
- ☐ NT

Q3 What is the postcode of the school that you work in?

____

Q4 Which of the following best describes where you work?

- ☐ Independent or private school
- ☐ Government school

- ☐ Catholic school
- ☐ Special education setting
- ☐ Other (please specify) ______________________________

Q5 Which of the following best describes the year level you teach?

- ☐ Pre-school
- ☐ Primary (please specify year level) ______________________________
- ☐ Secondary (please specify year level) ____________________________
- ☐ Primary & secondary (please specify year levels) ___________________

Q6 What best describes your role in the school?

- ☐ Classroom teacher
- ☐ Learning support teacher
- ☐ Other (please specify) ______________________________

Q7 How many years have you been teaching?

- ☐ 1 to 2 years
- ☐ 3 to 5 years
- ☐ 6 to 9 years
- ☐ 10 to 15 years
- ☐ 16 to 20 years
- ☐ More than 20 years

Q8 Have you ever attended professional development workshops on learning difficulties?

- ☐ Yes
- ☐ No

Q9 (if yes) How long ago was the most recent workshop you attended?

- ☐ Less than 6 months
- ☐ 7 to 12 months
- ☐ 1 to 2 years
- ☐ 3 to 5 years
- ☐ More than 5 years

Student demographics

When responding to the questions below, think of a student that you are currently teaching, or have taught within the last 12 months, whose performance in the classroom is familiar and well-known to you.

Q10 What is the student's gender?

- ☐ Male
- ☐ Female

Q11 How old is the student (in years)?

____

Q12 What grade is the student in?

____

Q13 How is this student performing in the following academic areas?

	Below the standard expected for his/her grade	At the standard expected for his/her grade	Above the standard expected for his/her grade
Reading	☐	☐	☐
Writing	☐	☐	☐
Speaking & Listening	☐	☐	☐
Mathematics	☐	☐	☐

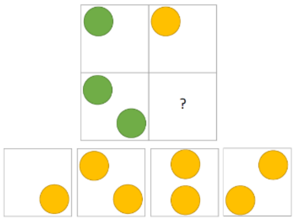
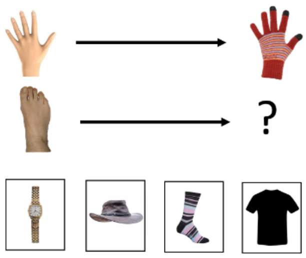
Continue having the same student in mind when answering the following questions.

The ECCA questionnaire is designed to measure a student's level of cognitive functioning in different cognitive ability areas. Cognitive abilities are the mental skills that people use to process information. There are several different areas of cognitive ability, and children tend to vary in their levels across the different areas (i.e., people tend to be high in some cognitive abilities and low in others, rather than high in all or low in all). One example of a cognitive ability area assessed in this survey is learning efficiency, which is the ability and efficiency with which an individual can learn, store, and consolidate new information in long-term memory.

The statements below will ask you to rate, from well below average to well above average, the student's ability when performing certain tasks. When responding to each item please compare the student to OTHER CHILDREN HIS/HER AGE. It is important that you answer as honestly as possible. There are no right or wrong answers, you just need to pick the response that you believe best describes the student.

If you believe that the student is capable but does not perform the task or behaviour any longer due to having grown out of it, award full credit for the item (i.e., "well above average").

Compared to other children of the same age, the ability of the student to complete these tasks is…

	Not sure	Well below average	Below average	Average	Above average	Well above average
Come up with a solution when confronted with a problem	☐	☐	☐	☐	☐	☐
Figure out the number that comes next in a number series (e.g., 1, 2, 4, 7, 11, ..?)	☐	☐	☐	☐	☐	☐
Think logically to solve a problem (e.g., could select the right piece in the puzzle below) 	☐	☐	☐	☐	☐	☐
Sort and classify objects according to a given category	☐	☐	☐	☐	☐	☐
Identify relationships between concepts (see example below) 	☐	☐	☐	☐	☐	☐
Use previously acquired knowledge to solve new problems	☐	☐	☐	☐	☐	☐

Compared to other children of the same age, the ability of the student to complete these tasks is…

	Not sure	Well below average	Below average	Average	Above average	Well above average
Provide synonyms and antonyms of a word (e.g., word "happy"; synonym "cheerful"; antonym "sad")	☐	☐	☐	☐	☐	☐
Provide oral definitions for words (e.g., able to answer "what does emotion mean?")	☐	☐	☐	☐	☐	☐
Demonstrate good general knowledge about the world around them (e.g., able to answer "what is used to make paper")	☐	☐	☐	☐	☐	☐
Correctly name pictures of familiar objects	☐	☐	☐	☐	☐	☐
Engage in conversations with others (peers and adults)	☐	☐	☐	☐	☐	☐
Display general knowledge of learnt social rules (e.g., able to answer "what should you do if you find an mobile phone on the ground?")	☐	☐	☐	☐	☐	☐
Communicate effectively what he/she is trying to say	☐	☐	☐	☐	☐	☐
Identify similarities between two common objects (e.g., how are dogs and cats alike?)	☐	☐	☐	☐	☐	☐

Compared to other children of the same age, the ability of the student to complete these tasks is…

	Not sure	Well below average	Below average	Average	Above average	Well above average
Remember the numbers given in a verbally spoken math problem for long enough to note it down on paper	☐	☐	☐	☐	☐	☐
Listen to and repeat a list of words in correct order	☐	☐	☐	☐	☐	☐
Remember two or more instructions in the correct order	☐	☐	☐	☐	☐	☐
Follows instructions to do two or more tasks (e.g., pick up your jacket and put away your lunch)	☐	☐	☐	☐	☐	☐
Write down dictated sentences accurately, ignoring spelling and punctuation errors	☐	☐	☐	☐	☐	☐
Accurately solve verbally presented math sums by using mental arithmetic (i.e., no pencil or paper)	☐	☐	☐	☐	☐	☐

Compared to other children of the same age, the ability of the student to complete these tasks is…

	Not sure	Well below average	Below average	Average	Above average	Well above average
Listen to and retell the important parts of an orally presented story	☐	☐	☐	☐	☐	☐
Remember the order of events that happened in a movie or TV show	☐	☐	☐	☐	☐	☐
Recall facts about what he/she has read in a book	☐	☐	☐	☐	☐	☐
Remember information presented visually (e.g., in a movie or a live demonstration)	☐	☐	☐	☐	☐	☐
Retain new information after only a few learning trials (e.g., learning a spelling list)	☐	☐	☐	☐	☐	☐
Learn and remember the relationship between unrelated objects (e.g., associating a face with a name; associating the visual symbol "a" with its corresponding sound "ah")	☐	☐	☐	☐	☐	☐
Apply effective strategies to learn new information (e.g., using mnemonics to remember difficult spellings)	☐	☐	☐	☐	☐	☐

Compared to other children of the same age, the ability of the student to complete these tasks is…

	Not sure	Well below average	Below average	Average	Above average	Well above average
Quickly name objects presented on a page	☐	☐	☐	☐	☐	☐
Quickly name letters presented on a page	☐	☐	☐	☐	☐	☐
Quickly name whole words presented on a page	☐	☐	☐	☐	☐	☐
Rapidly give numerous examples from the same category (e.g., fruits and vegetables)	☐	☐	☐	☐	☐	☐
Quickly name connected information when presented with a visual or verbal cue (e.g., seeing an image, hearing a word)	☐	☐	☐	☐	☐	☐
Rapidly give examples of words that share a semantic feature (e.g., fruits that grow on trees)	☐	☐	☐	☐	☐	☐
Quickly recall information related to a learnt topic	☐	☐	☐	☐	☐	☐

Compared to other children of the same age, the ability of the student to complete these tasks is…

	Not sure	Well below average	Below average	Average	Above average	Well above average
Build a model (e.g., with Lego or blocks) from a picture of the completed model	☐	☐	☐	☐	☐	☐
Put parts of an image together to form the whole image (e.g., jigsaw puzzle)	☐	☐	☐	☐	☐	☐
Differentiate between similarly shaped symbols (e.g., plus "+" and multiply "x" signs)	☐	☐	☐	☐	☐	☐
Pick out relevant information in the presence of distracting visual stimuli (e.g., playing "Where's Wally")	☐	☐	☐	☐	☐	☐
Extract visual information from graphs and handouts	☐	☐	☐	☐	☐	☐
Interpret information presented in a visual format (e.g., reading graphs, charts, maps, etc.)	☐	☐	☐	☐	☐	☐
Accurately judge distances between objects (e.g., not placing objects too close to an edge, not bumping into objects or people)	☐	☐	☐	☐	☐	☐
Find specific information on a printed page or computer screen (e.g., finding a specific word in a passage of text)	☐	☐	☐	☐	☐	☐

Compared to other children of the same age, the ability of the student to complete these tasks is…

	Not sure	Well below average	Below average	Average	Above average	Well above average
Sound out unfamiliar words (ability to decode)	☐	☐	☐	☐	☐	☐
Know which sounds go with which letters	☐	☐	☐	☐	☐	☐
Understand what is being said when background noise is present	☐	☐	☐	☐	☐	☐
Differentiate between speech sounds (e.g., difference between /ch/ and /sh/)	☐	☐	☐	☐	☐	☐
Hear the difference between similar sounding words (e.g., rhyming words)	☐	☐	☐	☐	☐	☐
Blend letter sounds fluently to form words	☐	☐	☐	☐	☐	☐
Filter out background noise to listen to my voice in the classroom	☐	☐	☐	☐	☐	☐
Identify the beginning, middle, and ending sounds in words (e.g., first sound in cat, last sound in run)	☐	☐	☐	☐	☐	☐
Manipulate sounds within words to form new words (e.g., what do you get when /t/ is added to the beginning of rip)	☐	☐	☐	☐	☐	☐

Compared to other children of the same age, the ability of the student to complete these tasks is…

	Not sure	Well below average	Below average	Average	Above average	Well above average
Quickly complete a series of simple maths sums presented on a worksheet (e.g., 3 + 2, 5 + 4, etc.)	☐	☐	☐	☐	☐	☐
Quickly and accurately copy information from the board	☐	☐	☐	☐	☐	☐
Respond to simple questions quickly (e.g., what did you have for breakfast?)	☐	☐	☐	☐	☐	☐
Quickly identify differences and similarities between two objects	☐	☐	☐	☐	☐	☐
Work quickly and efficiently on routine tasks (e.g., getting set up for a lesson; getting ready for school)	☐	☐	☐	☐	☐	☐
Scan and quickly determine important information on a page	☐	☐	☐	☐	☐	☐
Completes a task in the same amount of time as their peers	☐	☐	☐	☐	☐	☐

You have reached the end of the ECCA questionnaire. Thank you for taking the time to participate in this research.

If you have any questions, please contact me at:

Lydia Soh
lydia.s.soh@monash.edu
03 9902 4891

To thank you for your time, you can choose to enter the raffle draw to receive a $50 Coles/Myer gift voucher. Do you wish to enter the draw? If yes, you will be taken to a second survey so that you can enter your email address. Your survey responses will not be linked to the email address that you provide.

- ☐ Yes
- ☐ No

(if participant clicks yes, he/she will be directed to a new survey containing the following):

Please enter your email address below for a chance to win a $50 Coles/Myer gift voucher

____

(if participant clicks no, he/she will be directed to END OF SURVEY message)

Appendix G

Monash University Human Research Ethics Committee Approval Certificate

Monash University Human Research Ethics Committee

Approval Certificate

This is to certify that the project below was considered by the Monash University Human Research Ethics Committee. The Committee was satisfied that the proposal meets the requirements of the *National Statement on Ethical Conduct in Human Research* and has granted approval.

Project Number: 0228
Project Title: Parent- and teacher-report scale of cognitive abilities
Chief Investigator: Kate Jacobs
Expiry Date: 01/08/2021

Terms of approval - failure to comply with the terms below is in breach of your approval and the *Australian Code for the Responsible Conduct of Research.*

1. The Chief Investigator is responsible for ensuring that permission letters are obtained, if relevant, before any data can occur at the specified organisation.
2. Approval is only valid whilst your hold a position at Monash University.
3. It is responsibility of the Chief Investigator to ensure that all investigators are aware of the terms of approval and to ensure the project is conducted as approved by MUHREC.
4. You should notify MUHREC immediately of any serious or unexpected adverse effects on participants or unforeseen events affecting the ethical acceptability of the project.
5. The Explanatory Statement must be on Monash letterhead and the Monash University complaints clause must include your project number.
6. Amendments to approved projects including changes to personnel must not commence without written approval from MHUREC.
7. Annual Report - continued approval of this project is dependent on the submission of an Annual Report.
8. Final Report - should be provided at the conclusion of the project. MUHREC should be notified if the project is discontinued before the expected completion date.
9. Monitoring - project may be subject to an audit or any other form of monitoring by MUHREC at any time.
10. Retention and storage of data - The Chief Investigator is responsible fo the storage and retention of the original data pertaining to the project for a minimum period of five years.

Thank you for your assistance.

Professor Nip Thomson

Chair, MUHREC

Appendix H

Department of Education Letter of Approval to Conduct Research in Government Schools

Department of
Education & Training

2 Treasury Place
East Melbourne Victoria 3002
Telephone: 03 9637 2000
DX210083

2016_003154

Miss Lydia Soh
206/135 Roden Street
WEST MELBOURNE 3003

Dear Miss Soh

Thank you for your application of 8 August 2016 in which you request permission to conduct research in Victorian government schools titled *Parent- and teacher-report scale of cognitive abilities.*

I am pleased to advise that on the basis of the information you have provided your research proposal is approved in principle subject to the conditions detailed below.

1. The research is conducted in accordance with the final documentation you provided to the Department of Education and Training.

2. Separate approval for the research needs to be sought from school principals. This is to be supported by the Department of Education and Training approved documentation and, if applicable, the letter of approval from a relevant and formally constituted Human Research Ethics Committee.

3. The project is commenced within 12 months of this approval letter and any extensions or variations to your study, including those requested by an ethics committee must be submitted to the Department of Education and Training for its consideration before you proceed.

4. As a matter of courtesy, you advise the relevant Regional Director of the schools or governing body of the early childhood settings that you intend to approach. An outline of your research and a copy of this letter should be provided to the Regional Director or governing body.

5. You acknowledge the support of the Department of Education Training in any publications arising from the research.

6. The Research Agreement conditions, which include the reporting requirements at the conclusion of your study, are upheld. A reminder will be sent for reports not submitted by the study's indicative completion date.

VICTORIA
State
Government

I wish you well with your research. Should you have further questions on this matter, please contact Youla Michaels, Project Support Officer, Insights and Evidence Branch, by telephone on (03) 9637 2707 or by email at michaels.youla.y@edumail.vic.gov.au.

Yours sincerely

Joyce Cleary
Director
Insights and Evidence

02/02/2017

VICTORIA State Government

Appendix I

Catholic Education Melbourne Letter of Approval to conduct Research in Catholic Schools

Research Application 0500: Outcome (Approved)

1 message

research@cem.edu.au <research@cem.edu.au> 24 February 2017 at 16:03
To: lydia.s.soh@monash.edu, kate.jacobs@monash.edu
Cc: research@cem.edu.au

Dear Miss Soh

Congratulations, your research application, 0500 - 'Parent- and teacher-report scale of cognitive abilities', has been approved by Catholic Education Melbourne.

I am pleased to advise that your research application is approved in principle subject to the eight standard conditions outlined below.

1. The decision as to whether or not research can proceed in a school rests with the school's principal, so you will need to obtain their approval directly before commencing any research activity. You should provide the principal with an outline of your research proposal and indicate what will be asked of the school. A copy of this email of approval, and a copy of notification of approval from your organisation's/university's Ethics Committee, should also be provided.

2. A copy of the approval notification from your institution's Ethics Committee must be forwarded to this Office (if not already provided), together with any modifications to your research protocol requested by the Committee. You may not start any research in Catholic schools until this step has been completed.

3. A Working with Children (WWC) check – or registration with the Victorian Institute of Teaching (VIT) – is necessary for all researchers visiting schools. Appropriate documentation must be shown to the principal before starting the research in the school.

4. No student is to participate in the research study unless s/he is willing to do so and consent is given by a parent/guardian.

5. Any substantial modifications to the research proposal, or additional research involving use of the data collected, will require a further research application to be submitted to Catholic Education Melbourne.

6. Data relating to individuals or the school are to remain confidential and protected in line with the Privacy Act 1988 (Commonwealth).

7. Since participating schools have an interest in research findings, you should consider ways in which the results of the study could be made available for the benefit of the school community.

8. At the conclusion of the study, a copy or summary of the research findings should be forwarded to Catholic Education Melbourne. It would be appreciated if you could submit this via email to research@cem.edu.au.

I wish you well with your research study. The information provided in your proposal is now closed for further changes. If you have any queries concerning this matter or need to make amendments in the future, please contact Ms Shani Prendergast at research@cem.edu.au.

Yours sincerely

Mr Jim Miles
DIRECTOR ENTERPRISE SERVICES
Catholic Education Melbourne

Appendix J

Letter to School Principals

Lydia Soh, PhD Candidate
Faculty of Education
Monash University
57 Scenic Boulevard, Clayton 3800
Tel: 9902 4891
Email: lydia.s.soh@monash.edu.au

26 October 2017

To the Principal

I would like to invite teaching staff at your school to consider taking part in a research project being conducted by myself as part of my studies as a PhD student at Monash University. This research is conducted under the supervision of Dr Kate Jacobs, a Lecturer at the Faculty of Education at Monash University.

We are seeking schools to participate by recruiting current school teachers in Victoria for the study. The aim of this research is to assess the validity of the Estimates of Children's Cognitive Abilities, a newly developed teacher-report scale of children's abilities. We hope that the ECCA scale will provide an efficient method for teachers to gain a deeper understanding of a student's learning and cognitive profile, and how specific cognitive abilities are expressed in a child's academic life.

Involvement in the project would simply require you to forward an email with a link to the online survey and explanatory statement to teachers at your school. At no point would the research team have access to your staff member's email addresses. Further, participation in the project is anonymous and thus, no individual teacher or school will be identifiable when the results are reported.

If you have any questions or would like to discuss this further, please feel free to contact me by email or phone.

Thank you for considering this invitation.

Yours sincerely,

Lydia Soh | Dr Kate Jacobs

Appendix K

Explanatory Statement for Study 3

EXPLANATORY STATEMENT

For teachers

Project Title: Parent- and teacher-report scale of cognitive abilities

Project Number: 0228

Dr Kate Jacobs
Faculty of Education
Phone: 03 9902 4884
email: kate.jacobs@monash.edu

Lydia Soh
Faculty of Education
Phone: 03 9902 4891
email: lydia.s.soh@monash.edu

You are invited to take part in this study. Please read this Explanatory Statement in full before deciding whether or not to participate in this research. If you would like further information regarding any aspect of this project, you are encouraged to contact the researchers via the phone numbers or email addresses listed above.

The aim/purpose of the research
The aim of this study is to assess the validity of a developed parent- and teacher-report scale of children's cognitive abilities.

Why were you chosen for this research?
You have been chosen for inclusion in this research because you are a teacher of a school-aged child aged between 5 and 18 years.

What does the research involve?
If you choose to participate in this research you will be required to complete an online questionnaire which will contain questions regarding your opinion on the ability of a particular child in your classroom to complete certain cognitive tasks. Completion of the questionnaire is expected to take approximately 10-15 minutes, and can be completed at a time and place of your own convenience.

Can I withdraw from the research?
Involvement in this study is voluntary and you are under no obligation to consent to participation. However, if you do consent to participate, you may only withdraw prior to submitting the questionnaire due to participation in this research being anonymous. Once submitted, the identification of your questionnaire will not be possible.

Possible benefits and risks to participants
We hope that this study will provide insight into the validity of parent- and teacher-reports of children's cognitive abilities. The successful development of a valid parent- and teacher-report measure of cognitive abilities would provide an efficient method for gaining an understanding regarding a student's cognitive profile important for learning. We also hope that the construction of such a scale would help to increase parents' and teachers' understanding of the significance of specific cognitive abilities and how they manifest in a child's academic life.

Inconvenience/discomfort
This task is considered to be low risk; however, if any distress or other side-effect is experienced during participation, your involvement can be stopped immediately at any time. Being in this study is voluntary and you are under no obligation to consent to participation. Should participation in this research cause you any distress, it is recommended that you contact Lifeline's 24-hour crisis telephone line on 13 11 14.

Confidentiality
Participation in this research is completely anonymous. Under no circumstances will your responses be able to be identified as yours.

Payment
If you choose to provide your email address at the end of the questionnaire, you will enter a raffle to win a $50 Coles Group & Myer gift card. As your email address is not linked to your responses, your responses will remain anonymous.

Storage of data
Your responses to the questionnaire constitutes as data in the context of this project. Storage of this data will adhere to the University regulations and kept on the chief investigator's password-protected staff laptop for 5 years, after which all data will be destroyed. Only the researchers named in this explanatory statement will have access to the data. A report of the study will be submitted for publication, but individual participants will not be identifiable in such a report.

Results
If you would like to be informed of planned projects, and/or the aggregate results of any approved research that has been conducted, please contact the researcher(s) listed at the start of this explanatory statement.

Complaints
Should you have any concerns or complaints about the conduct of the project, you are welcome to contact the Executive Officer, Monash University Human Research Ethics (MUHREC):

Executive Officer
Monash University Human Research Ethics Committee (MUHREC)
Room 111, Building 3e
Research Office
Monash University VIC 3800

Tel: +61 3 9905 2052 Email: muhrec@monash.edu Fax: +61 3 9905 3831

Thank you,

Dr Kate Jacobs

Chief Investigating Officer

Lydia Soh

Student Researcher