THE EFFECT OF MODE OF DISCOURSE ON

EFL WRITING PERFORMANCE

by:

MAJID NEMATI

BA (English language and literature, Tehran University)

MA (Teaching English as a Foreign Language, Tarbiat Modarress University)

THESIS SUBMITTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

IN APPLIED LINGUISTICS AT THE SCHOOL OF EDUCATION

UNIVERSITY OF LEICESTER

September 1999

UMI Number: U601255

UMI

Dissertation Publishing

UMI U601255

ProQuest

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Dedicated to my wife **Azam,** who stood by me throughout a lifetime of

hardship. I owe her every single word of it.

Acknowledgements

ABSTRACT

THE EFFECT OF MODE OF DISCOURSE
ON EFL WRITING PERFORMANCE

It has been known for a long time that L1 writers perform differently in different modes of discourse. Despite its importance, there has been no conclusive evidence to shed light on the issue of ESL/EFL learners' writing performance across various discourse types especially in EAP environment. Therefore, this research study was designed to investigate differences resulting from the effect of four discourse modes (i.e. Narrative, Description, Explanation, and Argument) on EFL writing skill.

The research was conducted in three different phases mainly with university students in Iran. The hypothesised differences were examined in the three dimensions of production (through eliciting compositions), recognition (through cloze tests derived from compositions written in different discourse modes), and finally the learners' attitude towards these types of writing (through questionnaires).

The results of the production phase show a statistically significant difference between argument and description but not between narrative and exposition. For the recognition phase, significant differences were observed among all four types of discourse, ranking narration, exposition, description, and argument in order of their difficulty level as cloze tests. At the third stage of the study, an examination of learners' attitudes towards composition prompts was examined which showed their reluctance towards writing in argumentative mode. It can be concluded that argument proved to be the most difficult mode and narrative the easiest in all

three phases of this research study. Description stands somewhere in between.

Exposition did not follow any consistent pattern and seems to be too broad a category to be considered as a single mode and needs to be narrowed down into more manageable sub-modes. Overall, the findings of this study indicate that the L2 learners at tertiary level behave in the same pattern that has been established for the L1 young writers.

Recommendations are made to increase the construct validity of writing element of test batteries and to improve pedagogical insights into writing skill.

TABLE OF CONTENTS

CHAPTER ONE:
INTRODUCTION, BACKGROUND AND PURPOSE

CHAPTER TWO:

LITERATURE REVIEW

Section one: Some issues in writing

Section 2: Some issues in testing

CHAPTER FOUR:

METHODOLOGY

Main study

CHAPTER FIVE
RESULTS AND DISCUSSIONS

CHAPTER SIX
CONCLUSIONS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE
INTRODUCTION, BACKGROUND AND PURPOSE

Clearly, the development of a research agenda into prompt variables and prompt effects is a major task in itself, for there are so many unresolved questions that careful thought must be given to prioritizing these based on how much impact decisions about prompts will have on test takers' option and, therefore, on their performance.

Hamp-Lyons and Kroll (1997, p.25)

1.1 An Introduction to introduction

Within the genre of doctoral theses, if we accept such a genre, it is a norm to have a

short chapter to serve as an introduction. Such opening chapters, depending on their

scope, may vary from a couple up to a few scores of pages. Following the latter style

in the present work, this introduction is not merely an introduction but it goes further

to cover several issues related to previous studies. This should demonstrate the

necessity for a new study and underline its importance by reporting what has been

done by other researchers and by outlining the current situation.

1.2 Background and previous studies

Writing has always played a crucial role in students' academic careers and this role is

currently receiving more attention. Keller-Cohen and Wolfe (1987) estimated that 70%

of courses within the undergraduate curriculum require some sort of writing. Besides,

97% of faculty surveyed in the same study emphasised the importance of writing for

college courses. Therefore, writing seems an indispensable subject of every reasonable

academic syllabus.

Undoubtedly, formal institutionalised teaching should be accompanied by a proper method of evaluation. The importance of accuracy of evaluation, especially when it comes to writing, cannot be overestimated, because many important decisions made in higher education, from admission to graduation, are based exclusively on written tests and tasks.

Generally speaking, writing is evaluated in two ways, directly and indirectly. In direct methods of testing writing students are required to create a piece of writing , usually in the form of an essay, whereas indirect methods evaluate the writing skill through multiple-choice items, presuming that one's ability to write is assessed in assessing components of writing such as grammar and vocabulary.

Comparing these two, the indirect method of testing writing enjoys a relatively high reliability, while the direct method is widely considered valid. Although recommended by numerous scholars and educators such as Wiseman(1949), Wiseman and Wrigly(1958), and Britton et al. (1975), for evaluating writing in L1, the widespread direct assessment of writing skill had to be delayed in EFL circles until 1986, when the Test of Written English (TWE) was introduced by the TOEFL programme. Subsequently, the direct testing of writing became considered world-wide as the only valid method of evaluating writing.

While the direct evaluation of writing skill is generally accepted and practised, there are many controversial issues that, in Raimes' words (1990), are "causes for concern". Philosophically, these concerns, like any other concern, arise because of the lack of enough knowledge. It is admitted by scholars and pioneers of the field that there is still a lot left to be learned about direct assessment of writing proficiency (e.g. Hamp-Lyons, 1986a, White, 1985).

1.3 Components of direct writing assessment

There are four major components in a direct test of writing which may affect the test reliability and validity: *the writer, the task, the scoring procedure,* and *the reader* (Hamp-Lyons,1990). These components, each comprising several variables, can significantly change the test results and therefore it is justifiable to discuss them in more detail.

1.3.1 The writer

It is true that the judgement made on a learner's writing ability is mainly based on his or her written performance, yet it cannot be ignored that the person who writes and is supposed to fulfil the readers' expectations is an individual human being, and a rich collection of differences, such as gender, psychological personality orientation, social and economic status, exists between individuals. One can think of an even wider range of differences among different writers when it comes to second language writing as there is at least one more important difference which may cut across other specific groupings and that is the difference in the first language.

Although there is comparatively little knowledge about writer variables, it has been established that some L2 writers have severe cultural problems in understanding the required expectations and conforming to academic writing expectations (e.g. Ballard and Clanchy, 1991; Basham and Kwachka, 1991, Kaplan, 1987). Even if the writer conforms to the expectations of the test setter it is the reader, perhaps a teacher with a different set of expectations, who judges the written text.

Adopting a humanistic point of view such as that of the *expressive school*, one has to admit that a writer is an independent human being with a personality. But, the student writer is usually forced to fulfil a writing task which has been set by another person. Therefore, Weaver (1973) believes that in order for a task to be done successfully, the student writer should consider the topic as his or her own by understanding and appreciating the proposed task. Otherwise, the writer will write about something else. This phenomenon, dubbed by Hamp-Lyons (1988b) a "challenge", will cause considerable problems for raters especially when the writer produces a really good written text without covering the required task.

A new horizon has been opening up in the field of writing proficiency which views the writer as a thinker (Elbow, 1973, Holland, 1976, and Pollitt et al. 1985). This new dimension goes beyond the linguistic level by assessing the writing ability according to the writer's cognitive development. However, it needs more supporting evidence to be substantiated. It is not easy to draw conclusions about someone's cognitive development by simply analysing a piece of writing. As has been shown by Jacobs (1982), some tasks are cognitively more demanding, and therefore the observed cognitive development is limited to that individual task. This idea is supported by the findings of the present study.

Another issue regarding the writer variable in a writing test is the interaction between the writer and the reader. While writers keep the reader in mind, readers also make judgements about the writer while reading the essay, unless trained not to do so. This, as echoed by Wilkinson (1983), means that besides linguistics and cognitive scales in a model of writing development, there is a need for a moral scale, too. Lack of such a measure may result in misjudgements which can be traced into social factors rather

than writing ones. An instance in this case is the underestimated score that black students sometimes receive from their white readers (White and Thomas, 1981).

1.3.2 The reader

In any writing assessment, say of essay tasks, the final word belongs to the readers who judge the essays and rank them. Therefore, it is extremely important to minimise the inconsistencies which exist both within and between the readers; or in other words, increase the interrater and intrarater reliability measures.

There is no doubt that different people read and judge a piece of writing differently (Diederich et al. , 1961), but the question is why this happens, what kind of patterns may exist, and how it can be reduced by bringing the raters' views together. Hamp-Lyons (1989) points to cultural differences as a cause of such disagreements and attributes it to the readers' background experiences. Certain background factors such as sex, race, and discipline, have been named by researchers (Vann, Lorenz, and Meyer, 1991). Nemati (1993) suggests that Native Speaker (NS) raters are generally more lenient and generous about grammatical errors and, compared to Non-Native Speaker (NNS) raters, pay more attention to the content rather than the form of essays. Similar reports exists in the literature that "readers in related disciplines are more tolerant of grammatical errors than their colleagues in the English programmes" (El-Bacha, 1997; p. 65).

Although training the readers to rate more uniformly is "now an accepted part of any rigorous writing assessment program" (Hamp-Lyons, 1990, p.81), Robinson (1985) and Vaughn (1991) are doubtful about the effectiveness of these training sessions. Ironically, this can be partially because of the vast agreement that already exists among

all competent and qualified raters, even with a considerable background differences (Nemati, 1993). This relatively high degree of harmony among raters leaves little to be achieved by holding co-ordinating sessions to eliminate those deeply-rooted remaining differences which exists in the raters' points of view.

1.3.3 The scoring procedure

The major difference between direct tests of writing and indirect ones is their scoring procedure. Even further, the direct approach of writing assessment itself employs different methods and techniques. As these methods are fairly technical and the advantages and disadvantages associated with each one are a little complicated but essential to this thesis, they will be discussed in some detail in chapter two.

1.3.4 The task

Among different components of a direct test of writing- e.g., the task, the writer, the scoring procedure, and the reader-, the task has been the focal point of many researchers, as it is the final product and the basis of any judgements and decision making with regard to the writer's mastery over the writing skill. Task itself includes several variables such as use of pen and paper or alternatively using a typewriter or a word processor, and also length of time allowed for writing the essay, all of which may affect the writer's test score. But, there is one more factor, so controversial that even after the huge bulk of research done on it, it still remains challengingly mysterious. This problematic variable is the *topic* or *prompt* which refers to writing assignments on tests, defined by Rosen as "the complete verbalization of the writing assignment" (1969, p.36).

There is a general assumption among teachers as well as researchers that the nature of writing tasks depends on the given topic and therefore students' writing scores can change drastically from one topic to another (Hartog,1936; Hartog et al., 1941, Braddock, Lloyd-Jones, and Schoer, 1963; Britton et al., 1975; Poetker,1977, Hirsch and Harrington,1981; Applebee,1982; Freedman and Calfee,1983; and Pollitt et al., 1985). However, the counter argument has not been left untouched. Brossell and Hoetker Ash (1984), provided some evidence to show that topic difference is not that much significant. Carlson et al. (1985), found the same correlation coefficients for topics of different type and topics of similar type and Spaan (1989, 1993) observed similar performances on two supposedly different topics.

A similar field of investigation is the study of topic patterns favoured by learners. For instance, Chiste and O'Shea (1988) reported that both native and nonnative subjects, from a list of four topics, had chosen the first two topics which happened to contain fewer words. Because of this incidence they could not firmly conclude about learners' preference and conceded that "an explanation for this pattern is not possible" (ibid, p. 683). It can be either the size of questions or their location in the list of topics. Polio and Glew (1996) suggest that students perceive that they need a choice, though this choice may not necessarily improve their performance. Nevertheless, this choice can positively encourage writers by increasing the face validity of writing tests (Hayward, 1988). With regard to the form of topics, it has been found that native speakers prefer prompts in question form (Brossell and Hoetker Ash, 1964), especially direct questions (Hayward, 1988), while ESOL students seem to be unaware of this difference (Osburne, and Mulling, 1994).

The paradox of these contradictory findings can be accounted for by two reasons. First, the outcome of a study depends, to some extent, on the researcher's approach and expectations and as Hamp-Lyons (1990, p.74) put it " the solution one prefers will depend on one's statistics and on the expectations one started with". Secondly, it is vital to realise that topic or prompt is not a single entity (e.g. Ruth and Murphy, 1988) rather it is a complex of many other variables such as *purpose* (Witte et al.,1990), *audience* (Smith and Swan, 1977; Crowhurst and Piche,1979; Rubin and Piche,1979), *culture-related expectations* (Shuy and Fasold,1973; Hoover and Polizer,1980; White and Thomas,1980), *linguistic level of difficulty* (O'Donnell,1968; Harpin,1976; Brossell and HoetkerAsh,1984; Pollitt et al., 1985), *rhetorical specification* (Brossell,1983) and finally *mode of discourse*, The latter will be dealt with in more detail as the main focus of this study.

1.4 Mode of discourse and writing competence

It is still an open question whether or not some writers perform better in certain modes of discourse but not as well in the others. Modes of discourse have been classified differently in different sources and in some cases one might find them overlapping and confusing. It seems that the lack of a clear-cut definition of different modes has caused at least some of the confusion. Researchers tend to take the already-established distinction between the modes for granted without bothering to define them. Despite this confusing situation, there is a broad unanimity over the names and basic categorisation of the four primary modes of discourse- i.e., *narration, description, exposition* and *argumentation*. This traditional classification is still

overwhelmingly accepted and widely used in present literature (e.g. Scott,1996). She states the problem quite clearly:

> Does writing competence vary with mode of discourse? The four primary modes of discourse traditional to the discussion of writing are *narration, description, exposition,* and *argumentation.* Standards for a good narrative would differ somewhat from standards for a good description, just as good descriptive writing would differ from good expository or persuasive writing. However, the important question with regard to writing competence involves individual variation in different discourse modes. That is, are some writers competent when writing in one mode but less competent when writing in another? (p.13)

She further maintains that "writers may have competence in several, but not necessarily all, modes of discourse." (ibid)

1.5 Previous studies

As early as 1953, Kincaid drew attention to the fact that writers, especially better ones, perform differently in different modes of discourse i.e., narrative, descriptive, argumentative, and expository texts. However, it was Braddock et al. (1963) who first showed serious alarm at the neglect of mode specification in doing research on writing. Veal and Tillman (1971) made an observation on second, fourth and sixth grade children's writings of four different modes of description, argument, explanation and narration and found that the greatest difference between second and sixth grade children was in their expository compositions followed by narrative, description and argument. They also observed that for fourth grade students the scores for expository writing were higher than argumentative modes. Quite contradictory to most researchers, Quellmalz and Capell (1979) reported that ratings for narrative essays were lower than the expository ones and suggested three reasons for this phenomenon:

(1) There is a tendency among raters to regard expository as superior to narrative

(2) Different discourse modes are not emphasised equally in the curriculum

(3) Students suffer from a lack of knowledge to cope with narrative assignments

Similarly, in a frequently cited research project, Crowhurst and Piche(1979) observed a variation in syntactic complexity among the texts written by the same writer but in different modes of argument, description and narration. The subjects of this research were chosen from sixth grade and tenth grade children and both groups wrote argumentative essays which were syntactically more complex (i.e. longer T-units and clauses) than either descriptive or narrative essays. They also noticed that changes in mode resulted in changes in syntactic complexity while changes in audience had no effects. Previous studies done by Rosen (1969) and then by Perron (1977) had also shown similar results. The study done by Rosen is outstanding as it clearly identifies the difficulty order of discourse types. Investigating the essays written by O-level students, he realised that discussion/speculative writing was more difficult than narrative/descriptive writing. Relying on these results, among other reasons, narration and description have been classified as hypothetically easy modes in the present study.

There is an interesting point here that within the bulk of such research done in the L1 situation, two works, i.e. Kincaid (1953) and Rosen (1969), are much better substantiated and much better designed compared to others but less celebrated. Both of these works were PhD theses and it seems that, quite regrettably, this genre is not well circulated among the research discourse community. While many superficial journal articles are quoted everywhere, doctoral theses are left on shelves and my personal experience shows that many of them have not been consulted even for a

single time. Pearce (1974) refers to the important implication of Rosen's findings (ibid) for language evaluation:

> The kind of writing which the candidate is asked to do, or allowed to choose, may exert more influence on his performance, and hence on his result, than any other variable such as linguistic competence, intelligence, or age. (p.57)

In another influential study, Cooper and Watson (1980) asked 6 elementary students to write three compositions each in a different type of discourse: Explanation, Persuasion and Expression. They confirmed that "the writing situation" or topic can affect writers as young as nine years old. In this study argumentative writing was found to be the most syntactically complex. The study had two implications:

1. different discourse types require different thinking and planning strategies.

2. instruction and assessment should aim at higher-level discourse structure rather than sentence-level issues.

Crowhurst (1978) also studied the writings of six, ten and twelve grade students and reported that narration received higher scores than argument. It was also maintained that there is an association between syntactic complexity and the quality of argumentative writings but that such relation does not exist for narrative essays. She concluded that "narrative writing is generally less syntactically complex than argument"(ibid, p.8).

Later Freedman and Pringle (1981) demonstrated that 98% of 12 year children were able to realise the conventional schema for narrative structure most of the time but for argument structure this ability fell to 12.5%. These research findings have two implications:

1. students must be given more opportunity to internalise argument patterns by being exposed to them through reading and through direct instruction.

2. the ability to cope with argumentative writing requires cognitive maturity attainment.

These two points have fundamental bearings on the present research as this researcher, as a starting point, had hypothesised that the situation at university level would be different for adult students must have had enough opportunities to become familiar with argument patterns and, of course, they had attained relative cognitive maturity.

Quellmalz, Capell, and Chou (1982), choosing twelfth grade students as their subjects, conducted a research study concluding that the level of performance varies from narrative to expository discourse. Freedman and Calfee (1983) who compared compositions requiring quotations and those requiring opinions, found that the writers' scores were significantly different between the two types of composition. They concluded that even within a certain mode there are sub-modes which require different abilities of the writer. Similarly, Matsuhashi (1982) reported that tasks with different purposes may require different degrees of mental effort. Matsuhashi particularly points to *reporting* and *genaralising*, considering the former much easier than the latter for being linear and well-practised.

Reed et al. (1985), in an experiment involving English-speaking college students with different writing capabilities, found that for these students narration was the easiest, description second, and persuasion the most difficult mode of discourse to write in, because of the different degree of cognitive capacity on engagement each requires.

All of the foregoing studies focused on L1 subjects. There are only a handful of studies about the situation in L2 of which few are rigorous enough to be considered as

conclusive with regard to any aspect of topic variable , let alone the discourse mode. For example, the study of the effects caused by modification of formality level of language done by Hirokawa and Swales (1986), in which they did not find any difference between prompts written in formal language and those written in informal language, was too small-scale. A major example of important and related research study is that of Hamp-Lyons (1991) in which she investigated the effect level of difficulty of the essay prompt on ESL learner performance. The prompts were divided according to writing task type namely, expository/private; expository / public; argumentative / private; argumentative/public; and a combination of two or more of these types. It was hypothesised that the expository/private was the easiest and the argumentative/public the most difficult one to perform. Surprisingly, she found that the mean writing score increased as topic difficulty increased. In a more recent research study of this type, Koda (1993), focusing on American college students studying Japanese, realised that different linguistic (grammaticality and sentence structure) and rhetorical (organisation and coherence) competencies are required for the two different modes of description and narration.

And finally, in the most closely-related study to the present research (Hamp-Lyons, personal correspondence), Hamp-Lyons and Mathias (1994) investigated expert judgements of the task difficulty and surprisingly noticed that, although these experts shared the same ideas and criteria about what makes a prompt difficult, their predictions were the reverse of the pattern shown by the subjects' actual scores.

All in all, drawing upon the previous studies, Park (1987) makes the following conclusions:

(1) Discourse mode is significant with regard to L1 writing performance

(2) Among different modes, argument is the most difficult one and therefore it usually receives the lowest scores

(3) The effects arising from a change in type of discourse are stronger for elementary than for the high school level

It can be also concluded that "there has been . . . no real investigation of the effect of task variable on the measured writing quality of ESL writers on direct tests of writing" (Hamp-Lyons, 1990, p.75). Obviously, there is little solid and objective data to pinpoint the topic difference resulting from the mode of discourse and its effect on the ESL writer performance. There is even less, if any, evidence to account for such differences from the viewpoint of writers themselves and their attitudes and preferences towards different topics of different discourse types and finally the inherent cognitive and linguistic differences these texts bear.

1.6 The present study: Aims & Research questions

The present study is an attempt to investigate those aspects of essay topics which may affect the EFL learner's performance due to the differences in the discourse type of the elicited essay. As discussed in the early parts of this chapter, topic is not a single variable, rather it is a conglomerate of variables; each and every of them may affect the writer's performance. The present study is mainly concentrated on the four basic modes of written discourse i.e. *argument, description, explanation* and *narrative* and whether or not EFL writers perform differently while writing in these modes. From a psychometric point of view, this is of especial importance in writing assessment because task variables " must be manipulated and controlled to give every test-taker the opportunity to produce his or her best performance" (Hamp-Lyons, 1990, p.73),

therefore, it is important to know the mode of discourse in which, as we can expect, test takers are more likely to produce a better piece of writing .

 Besides, it is part of this study to investigate EFL learners' attitudes towards some proposed essay topics requiring different modes of discourse. This is to find out the degree to which their performance complies with their preferences.

The third and final major dimension of the study involves the text itself. It can be of great value to know whether these modes of discourse impose different mental and cognitive pressure on readers as well as writers. Accordingly the main three questions of this study can be stated as follows:

1.7 Main research questions

1. Is there any significant difference between the quality of essays written by EFL learners in the hypothesised difficult discourse modes (*argument/explanation*) with that of those written in relatively easier modes (*narrative/description*)?

2. Do the essays written in different discourse types bear different degrees of difficulty (cognitive and/or linguistic) to recognise and reconstruct?

3. Do EFL writers prefer certain mode(s) of discourse over other mode(s) when choosing a topic to write on?

1.8 Methodology

In order to answer these questions a relatively complicated methodology has been adopted, rather than just the traditional method of investigating students' writing

which largely consists of essay collection and evaluation. This study tries to view the impacts of discourse modes on EFL writing from the three different angles of production, recognition and attitude. To do so, a methodology with three stages was required, one stage for each angle.

For differences in production a traditional essay examination is proposed. Because of certain limitations which will be discussed later in the methodology chapter, it was not practical to elicit four compositions each in a different mode of discourse from all subjects. Unwillingly, instead, the researcher required each subject to write two compositions, one in a supposedly easy mode (narrative/description) and one in an assumed more difficult mode (argument/explanation) . These compositions were later scored and compared with each other.

The second stage of the study is somewhat innovative as it does not stop at the surface level of an essay's impression on raters. This stage of the study goes further, probing the linguistic and cognitive complexity involved in the recognition and reconstruction of the produced texts by turning them into cloze tests. Four essays written by the same EFL writer, each in a different mode of discourse, were converted into cloze tasks each with 15 blanks to see how difficult they look to student writers.

In the third and final stage of the study, the same topics were offered in a questionnaire to some EFL learners to rank them, in a Likert-type scale, according to their own taste, preference, and willingness to write on those topics.

The subjects for the present study are university students of English (EFL) in Iran.

1.9 Significance of the study

This study is an attempt to follow-up significant findings by other researchers which have contributed significantly to the field of second language writing. Nevertheless,

this study in its own right, includes some unique features and innovations both in methodology and findings.

It should be pointed out that none of the forgoing studies, except that of Hamp-Lyons', dealt with L2 and consequently she (1990, p.76) reminds us that "we need a number of rigorous studies that would identify topic-related performance differences, if such exist, and replicate results". Aiming to bridge the gap, this researcher intended to investigate as fully as possible the mode of discourse effects on EFL learners' writing performance.

This research is different from the research of others who have made similar attempts in L1 since it deals with EFL learners, and unlike the Hamp-Lyons' (1991) study, it is not merely based on a subjective judgement of topic difficulty. Above all, it will include narration and description as modes of discourse alongside argument and explanation. Besides, in this study two other dimensions of *recognition* and *attitude* are investigated, too. In other words, it is not only the quality of written essays in different modes which is taken into account but the cognitive and textual complexity of the essays are probed too. This is done by converting the essays into modified-cloze passages to see how different and how difficult they are for students to read and reproduce. In a third dimension, the subjects were given a chance to express their own attitudes and preferences towards writing in different modes of discourse. Remembering that Hamp-Lyons' study (1991) revealed that teachers' perception of difficulty might be wrong, one can easily appreciate the importance of this part of the present study. Therefore this study investigates the issue of mode of discourse and its effect on the ESL writing from both process and product points of view. This is quite unique.

Since the main fieldwork was carried out in Iran, the findings of this study may have some bearing on Iranian students, too. Perhaps one of the most important current educational investments in Iran is giving scholarships to research students to study abroad. This is being done on a regular annual basis and each year hundreds of the most talented students are selected from a large number of candidates who take the highly competitive exams in their related fields of study. These Ph.D. candidates have to get admission from highly-ranked and well-established universities, mainly from the English-speaking quarters of the world. To do so they have to demonstrate their language proficiency in two different stages. The first stage is taking a language proficiency test designed by the scholarship bureau of the Ministry of Culture and Higher Education (MCHE) and hence named after it: the MCHE test. This is actually an imitation of the TOEFL and in fact serves as a mock test or preparation for it. Those who get an acceptable score on this test and prove their high levels of skill in the English language will be paid for to take either TOEFL or IELTS or even both tests. Meanwhile, the scholarship bureau holds intensive preparation courses to get the candidates ready for their actual tests. It is likely that some of the most brilliant candidates with sophisticated knowledge of scientific disciplines fail to continue their studies abroad simply because they cannot pass the language tests. Some do not even apply as they realise that their English language proficiency is far below the requirements. One of the areas that causes major problems for these candidates is the writing element of IELTS and/or TWE. The findings of this study can theoretically provide the test batteries with test-takers' view of the topic differences and how they perform on different topics. Practically, the study can help these candidates and their language teachers to make the English writing courses more effective.

Concerning the EFL teaching in the Iranian educational system, this research has two major objectives. Firstly, there is little awareness among language teachers of teaching writing in general and the probable difficulties caused by writing in different discourse modes in particular. As will be discussed in the following sections, the present trend of teaching English in Iran has put the reading skill as the first priority. Reading and writing, however, are inseparable. After all, both are manifestations of the written language. Reading is in fact decoding the information encoded in writing, and reading, or critically reading one's own drafts of writing, is surely a significant element in improving students' writing – but this must be done on the basis of knowledge and awareness. Consciousness-raising of existing differences among different texts can help make the teaching of reading more efficient, which in turn should help students' writing development.

Cumming (1998, p.61) joins other scholars such as Grabe and Kaplan (1996), Muchiri et al. (1995), Silva, Leki, and Carson (1997) who "have lamented research's predominant focus on English in academic settings in North America and Britain". Since the subjects of this study are Iranian students, it can be a remedy to this imbalance by bringing some data from EAP situation other than that of the North America and Britain. This is, in a sense, a move from ESL to EFL.

Now, it is necessary to examine the teaching of English in Iran since the subjects of this study are Iranian students who, as explained, need high levels of English to study abroad. This necessary background is presented in two levels: first a historical perspective of the foreign language situation will be given and in the second part some significant educational institutes will be discussed with regard to their roles in EFL teaching in Iran.

1.10 The system of EFL education in Iran

1.10.1 A historical perspective

Teaching a foreign language in Iran, as in the rest of the world, has been an important part of the educational curriculum. Four or five decades ago, French was the prevailing foreign language among the intellectuals and, for obvious reasons, in Iranian educational institutes as well. After the second world war, it became the fashion of the time to be English literate. This was not only to enable the scholars and scientists survive in an academic environment in which English was the lingua franca, but there were also certain political reasons giving impetus to such shift. Whatever the reason, as English started to sweep around the world, it remained the unchallenged foreign language in Iran. In the late 60's and early 70's a new movement, pioneered by Shiraz University, to teach different subjects through the medium of English drew widespread attention towards the English language as a determining factor in academic achievement. Simultaneously, the introduction of intensive language courses into the armed forces in general and the Iranian Air Force in particular, on the one hand, and the mushrooming of English language institutes all over the country (at least 150 institutes just in Tehran) on the other, made the importance of teaching and learning English in Iran indisputable.

In the late 70's and early 80's, the necessity of learning and using the English language, or at least its monopolised role as the only foreign language taught, was severely questioned in the immediate post-revolution Iranian educational system. As a

consequence, English was replaced by Arabic in year one of secondary schools and other major languages like French and German were introduced into the secondary education to sit alongside English. However, the teaching of these new languages remained restricted to a limited number of schools and has not been put into practice nation-wide. Besides, at the tertiary level, English is still viewed as virtually the only foreign language. Obviously, those who had studied either French or German at secondary school would face daunting problems during their university studies. Therefore, the introduction of other foreign languages ironically resulted in a more respected position for English as the universal academic language of the time.

For certain reasons which are beyond the scope of this study, the status of the English language was overwhelmingly welcomed by the younger generation in post-war Iran and this increasing popularity still dominates both academic and non-academic environments equally.

The next section will briefly examine the present status of English language in Iranian educational institutes. Given the lack of published current research about ELT in Iran, this section is largely based on the writer's personal experience and an extensive discussion with other informed participants.

1.10.2 Nursery and Primary-level education

Primary education begins at the age of 6 and lasts for five years when a uniform exam is held throughout the country to evaluate the quality of the students' achievements.

There is no teaching of the English language in the primary education curriculum. However, in most private nurseries and some primary schools English is taught as an extra-curricular activity. The teachers are usually unqualified with a very limited knowledge of the language and the teaching is mainly focused on basic vocabulary. Generally speaking, this is merely to give a more prestigious image to the institute and to satisfy parental demand rather than being an educational activity, so it can be concluded that there is no considerable or widespread foreign language teaching at the primary level.

1.10.3 Secondary-level education

This period lasts for three years and is intended to prepare the students to transfer from primary school towards their high school education and above all to ascertain their interests and aptitudes to guide them to choose the best-fitting specialism or major to continue their future studies. Perhaps that is why these schools are known as the 'Guidance Schools'. Previously, English used to be taught in all three years of this level, but it has now been removed from the first year syllabus. The main reason for this change is that the pupils start learning Arabic in this year and teaching two foreign languages simultaneously proved unsuccessful. In years two and three, English is taught for two or three sessions a week totalling (200 minutes). The materials are compiled and distributed by a special bureau of the Ministry of Education. A typical teacher of English at this level would be a high school graduate with two years of training at a 'Teacher Education Centre'. English is usually taught in a crude version of Grammar-Translation method beginning with the introduction of the English alphabet.

1.10.4 High school education

The teaching of EFL in high school education in Iran, which usually takes four years, more or less suffers from the same shortcomings that makes learning English in secondary schools seem to be a complete failure. Again, the absence of a communicative approach to teaching and the widespread use of a raw version of grammar-translation method which keeps the learners as passive as possible, turns the English classes, normally totalling 180 minutes per week, into very boring sessions. The main focus of the textbooks is grammar, so that by the end of the period almost all major issues in the English grammar have been taught. The classroom medium of instruction and explanation for English grammar is, of course, Persian. It is necessary to point out that the syllabus designers intend to promote the reading skill, as the skill which students will mostly require for their future academic life where the ability to read the textbooks in many disciplines in the original English version would be a great advantage. However, little attention is paid to the fundamental components of the reading skill, i.e. little attention is given to developing a range of reading strategies, such as skimming, scanning, developing the recognition of vocabulary, intensive and extensive reading. Besides, writing as a skill which can have a crucial bearing on reading has been ignored. A good memory is all a student needs to pass the English course with flying colours. As a result "there is a halo of pessimism and disappointment cast over the teachers' assessment/judgement of what they are doing" (Tahririan, 1986).

1.10.5 Tertiary education

Universities are the few places among the formal educational institutes in Iran where teaching English is treated in a relatively proper manner and method. This is not, by any means, to say that the outcome is satisfactory as the period is too short and it is too late to correct the students' widespread fossilised bad habits, which have been instilled in school.

Teaching English at universities can be divided into two major categories: English-medium courses and non English-medium courses. There are about four closely-related majors at Iranian universities which are taught in English i.e. **English Language and Literature, Teaching English as a Foreign Language (TEFL), Linguistics,** and **Translation.** At the moment, there are even various English language Ph.D. programmes available at several Iranian universities. These courses each consist of about 135 credit hours[1] of instruction at BA and B.Sc. level and about 34 credit hours at MA level. Ph.D. programmes consist of some taught courses and a submitted thesis examined during a viva session. All other university majors fall within the second category as they are not taught in English. However, all students are required to take certain English courses. Basically there are four credit hours of language teaching in all bachelor's curricula which are taught in two separate terms, for two hours per week each. The first course is a general English course consisting of reading and grammar. The second course is mainly a semi-specialised reading course with an element of grammar and translation. Writing is comparatively neglected, almost left in limbo. There are specific assigned textbooks for the students of humanities, engineering, and medicine respectively. Prior to these courses, most universities make their students take a pre-sessional English course to compensate for their poor high-school language

---

[1] One credit hour means one hour of instruction per week for a term of seventeen weeks. There are two terms in an academic year.

classes. Having passed the language courses successfully, students have to take another two-credit-hour course of reading specialised English texts which are usually excerpts of some classic English language textbooks of that field of study.

The outcome of this educational system is two types of students. The first group are those students who major in one of the four courses using English as the means of instruction. The graduates typically have a considerable knowledge of the English language and can, to some extent, communicate with native speakers. The students of the second group are relatively incompetent at communicating in English, be it written or spoken. Bearing in mind that these students, in terms of quantity and quality, are the main body of the Iranian academic environment, one can clearly see the devastating language flaw among the university graduates which eventually cripples them in all international forums where English predominates. It seems that a move towards more communicative and integrative language teaching, with more emphasis on writing, can be a remedy for this disappointing situation.

1.11 Definition of terms

A number of key terms have been frequently used in this thesis which may need clarification as in the literature they are ambiguous and, in some cases, controversial and may denote meanings different from the intended ones. These definitions are simply working definitions.

second/foreign language writing : A distinction has been often made between second and foreign language. In British usage "a foreign language is a language which is taught as a school subject but which is not used as a medium of instruction in school nor as a language of communication within a country" (Richards et al., 1992, P.142).

A second language, by contrast, is not the native language of the community but is widely used for communication purposes and in some cases it is the official and formal language of the society. In north America, however, these two terms are usually considered as identical or interchangeable. Writing in a language other than the native language can be divided into the two mentioned categories, that is, writing in a second language or writing in a foreign language.

Grabe and Kaplan (1996) define EFL and ESL writing courses in the following way:

> EFL courses include those students who need to learn English (or in this case need to learn to *write* in English), who live in countries in which English is not regularly spoken or written as language of the community (e.g. the People's Republic of China, Indonesia, France). ESL courses include those students needing to learn English who live in countries where English is a language , or *the* language, of the community. (p.24)

As the subjects of this study were chosen from the Iranian students who are mainly exposed to English language during the class hours and had little chance to practise it in everyday communication outside class it can be concluded that the present study falls in the domain of *foreign language writing.*

modes of discourse: Richards (1992, P.338) defines *mode of discourse* as "what part the language is playing in this particular situation, for example, in what way the language is organized to convey the meaning" . As this vague definition implies this is a very loose and flexible term. The literature contains a wide range of applications to this term. Nevertheless, a four-type distinction among different discourse modes looks more prevalent (e.g. Scott, 1996) and therefore in this study such classification has been adopted. The four discourse types are: narrative, description, explanation, and argument. It seems that the distinction between these discourse modes have been taken

for granted and it has been assumed by researchers that simply by adding words like "Describe..." or "Argue..." to the prompt, an essay of totally different discourse type can be elicited. Park (1986) bemoans the situation:

> The fact is that the different discourse modes can easily be blended in a piece of writing because the single essay is usually composed of smaller units of different discourse modes. Therefore, the results of studies in discourse mode effect on writing performance should be interpreted with caution. (p.22)

Although different modes may overlap in many cases, it is arguably defensible that essays can vary with regard to their main mode of discourse which prevails all over the text and that is what matters. Similar patterns realised in several L1 studies confirm that different discourse modes are meaningfully distinguishable.

rater, judge, reader: All these terms have been used interchangeably to refer to the person who reads the essays and assigns marks based on the adopted scale.

topic, prompt, task, question, cf. rubric: Despite some minor differences one may find among these terms, they are considered the same here and refer to the assigned question or situation on which the writer is required to write. *Rubric*, is a more general term which besides including the prompt, gives other details such as the allowed time, and even how the essays will be scored. Sometimes parts of a rubric, for instance the rating scale, may not be available to test takers.

1.12 An overview of the present work

This chapter, so far, has tried to create a panoramic picture of the present study from the theoretical background to the surrounding related studies to the real life situation in which the fields study has been conducted. In the next chapter some basic and fundamental issues in the field of writing, testing and, of course, testing writing will be dealt with in three sections respectively. In the third chapter, a taxonomy of modes of discourse, which is the foundation of this study, will be reviewed. Other parallel classifications of modes of discourse will also be touched. Upon the fourth chapter reports the methodology in two sections: the pilot study and the main study. The main study itself comprises three sections each named after the aspect of difference they are to investigate among discourse modes i.e. *production, recognition,* and *learners' attitude*. The results of the main study are presented in the fifth chapter. In order to facilitate the comprehension of results and achieve a more coherent account of the investigations, related discussions immediately follow the statistical results for each phase of the study. Similar to the methodology chapter, a trilogy exists here, too.

The sixth and final chapter will be the conclusion in which an overview of this research project is presented with a consideration of its implications for language teaching and language learning, language testing, material development and its contribution to the field of applied linguistics is underlined.

CHAPTER TWO
LITERATURE REVIEW

"Writing under assessment conditions constitutes an unnatural act, measuring only what the writer can do with a specific task under severely constraining conditions. It is as if walking ability were to be determined by assessing the ability to walk on a slippery pavement with a broken toe and high-heeled shoes"

Ruth and Murphy (1988, p.83)

2.1 Introduction

This chapter looks at some fundamental issues in testing and writing. Obviously there are numerous topics within each of these areas and it is equally obvious that that it would be impractical, if not impossible, to address all these issues in depth. Therefore, the areas which seem outstanding and salient in relevance and relation to this research study have been selected for discussion here. Even with this focus, the issues to be covered remain vast, so the chapter is divided into three sections. The first one deals with writing in general. It aims to give a brief history of writing, a working definition of writing by comparing and contrasting it with its oral counterpart, speaking, and a review of theories concerning L1 and L2 writing learning and teaching and a few other related topics. The second section deals with some of the major characteristics of a good test. It serves as a basis to familiarise the readers, and of course the writer himself, with the world of language assessment, where a solid understanding of certain concepts like reliability, validity is the passport to enter. And finally in the third section of this chapter, the two previously discussed issues are related to the context of the main research questions and thus it deals with different aspects and methods of writing assessment and related topics.

Section one: Some issues in writing

2.2 Definition of writing

There is an irony in writing about writing and using writing to define writing. The more one delves into this skill and the more one knows about the delicacies involved, the more difficult it is to write on writing.

Finding a definition for writing and written language seems deceptively easy. This may be due to the fact that almost everybody who is literate does some sort of writing during their daily activities; though some of these activities are quite brief and may not involve more than a few words, such as writing a memo or making a shopping list. These two examples, however, do not belong to the type of writing which is the main concern of most researchers investigating the writing skill. The reason is that they hardly involve composing. Composing, which can be defined as "the combining of structural sentence units into a more-or-less unique, cohesive and coherent larger structure" (Grabe and Kaplan, 1996, p. 4), is the major issue in academic writing as it is the basis of almost all academic writing activities and much academic assessment. Composing can itself be divided into two further categories: telling (retelling) and transforming. The former refers to the type of writing in which the author has already got the knowledge to be transferred and is fairly aware of the final product, for instance narrative and description. Unlike telling, in transforming the writer tries to find his way towards an unknown destination so the final product is not definitely clear

at the early stages of writing. Argumentative writing is the outstanding type of writing with this characteristic.

Figure 2.1 may give a better view of different types of writing activities, academic and non-academic. Most academic writing falls in the "*writing with composing*" category while some academic tasks such as "questionnaire" is regarded "writing without composing".

Figure 2.1 Patterns of composing with differing audiences (Grabe and Kaplan, 1996, p.4)

Audience	Writing without composing	Writing with composing	
		For knowledge telling	For knowledge transfer
Self	shopping list	Personal Diary	'Journal' notes
One known other	note to milkman	Personal letter	
One unknown other		Business letter	
Small group known		Lesson plan sermon	
Small group unknown	Questionnaire	Newsletter item	Proposal
Large group	Tax form Driver's licence application form		Poem, drama novel Short story

2.3 Speaking vs. writing

The relationship between writing and speaking has been viewed in various ways by different linguists and researchers. One might view writing as "merely a way of

recording language by means of visible marks" (Bloomfield, 1933, p. 31) or as visual

and graphic representation of speech on paper, as Halliday (1989) does, where he says

that whatever spoken can be, to some extent, written down, which is in a sense

correct. Accepting a similar analogy made by Robert and Lunzer(1968, p.220), stating

that reading is a "kind of listening with a visual rather than aural input", we can then

define writing as a kind of speaking with visual output of spelling and punctuation

replacing pronunciation. After all, both are so-called productive skills trying to make

communication possible by means of encoding a message. On the other hand, many

scholars refuse to believe in a simple close relationship and direct link between written

and spoken language (e.g., Raimes,1983c, Peacock,1986). The now widely accepted

understanding is that this difference is not that of a dichotomy but rather a spectrum or

continuum (Lewis, 1993). Biber (1988), using a multidimensional model of textual

variation, compared some 23 different genres of written and spoken texts and came to

the conclusion that there was no single dimension that could distinguish all types of

spoken texts from all types of written ones. In fact, writing does resemble speaking

because the former was undoubtedly derived historically from the latter. Yet, of

course, there are many differences. It goes without saying that speech is the prior form

of language phylogenetically for all known cultures and ontogenetically for all children

except in special cases. Historically speaking, linguists unanimously agree that speech

was used first. This is evident because even today there are unwritten languages which

prove that speech can stand without a written representation but not the opposite way

around. This priority also exists in children's language development. Speech is virtually

universal. That is, everybody learns a mother tongue the basis of which is achieved, as

early as the age of 3, whereas writing development is usually delayed until the age of 5

or 6, when formal education begins. Even then there is no guarantee that children will master writing skills and indeed at least one-fifth of the world population are illiterate. The problem is enhanced when writing is taught in EFL/ESL context. Therefore, writing is usually referred to as the fourth or last skill.

Now it seems warranted to examine the elements which make writing so difficult, demanding and, as Smith (1982) put it, "painfully slow". These differences are categorised and analysed in Figure 2.2, with the understanding that particular communication contexts may change the emphasis:

Fig.2.2: Differences between written and spoken language (from Bowen et al., 1985, p.68)

Visual (Written) Language	*Aural (Oral) Language*
1. Unidimensional, Unidirectional	Bidimensional, Multidirectional
2. Normal listener feedback	Delay or absence of feedback
3. Paced	Tempo ad libitum
4. Accuracy depends on monitor that must anticipate to act	Can be pondered, strengthened, corrected, edited, revised
5. Physical facility: speech tract and hearing mechanism	Physical facility: eyes and hand
6. Simple to arrange and retrieval	Complex equipment necessary to arrange for storage and retrieval

----

Basically, these differences can cause four types of difficulty in writing which make it more demanding than speaking. These four types or domains of difficulties existing in writing are:

a. linguistic

b. psychological

c. cognitive

d. social

These are each discussed below.

2.3.1 Linguistic differences

The most outstanding linguistic difference between writing and speaking is the absence of prosodic and paralinguistic elements in writing. Prosodic features are systematic and part of the linguistic system of the language. They include such delicate systems as intonation, contour, stress and rhythm. Unlike these prosodic features, paralinguistic features are not systematic. Among them gesture, body movement, tone of voice and facial expressions are of special importance. In turn, writing compensates for this shortcoming by using punctuation symbols (e.g., exclamation marks, block capitals). Yet, punctuation by itself cannot substitute for these finely-tuned phonological or paralinguistic features, leaving written communication still difficult and demanding.

Despite the fact that written language does not enjoy the advantage of some prosodic and paralinguistic features, it is more compact and brief. If we compare a conversation and its equivalent written text, their volume and size would be different (White, 1980). Kress (1982, pp. 29-30) offers brilliant examples of such difference.

Three reasons can be cited here to explain the differences. Firstly, most informal speech is repetitive and full of redundancy, digression and fillers like "you see", "I mean" and so on. Secondly, in speech the speaker mainly uses simple sentences stitched together by lots of "buts" and "ands". Writing, however instead of co-ordination uses more complex and embedded sentences employing connecting words such as "however", "who", and "in addition". Thirdly, the richness of vocabulary in written texts enables the writer to convey the intended message as concisely and precisely as possible. Again it should be recalled that there are some cases of spoken language, such as very formal lectures, which are lexically very dense and may be written to be spoken, while some written texts employ a simple language resembling everyday conversations and, like drama scripts, may also be written to be spoken.

Another linguistic factor which makes writing different from speaking is the occurrence of unfamiliar and rare grammatical structures in written texts. This complexity sometimes trespasses on the boundaries of sentences and results in bizarre overall patterns of discourse organisation. Although the burden here is mostly put upon the shoulders of the reader, it is a distinctive characteristic of written language. A further important feature is the uniformity of written language. It can be argued that there is only one standard, formal, grammatical, bookish form in any language. Almost all languages yield themselves to a far wider range of different forms spoken by inhabitants of different regions. Such dialects can, of course be written but in formal contexts, certainly in academic settings, this is rarely acceptable except in scripted dialogue and regional/local (oral) literature. Written language, on the other hand, seems to be less flexible towards these diversities. For example relatively trivial

differences of spelling between British English and American English irritates many users of the language more than the gross differences of pronunciation do.

2.3.2 Psychological differences

Besides the linguistic factors discussed before, there are some psychological elements which affect the mind of writers, making the process of writing more intricate than that of speaking. First of all, the nature of speaking is here and now intricately tied to the immediate context. In other words, there is usually a shared physical setting between the speaker and the listener. Consequently, there is always some kind of interaction between the speaker and the interlocutor who is present. This mutual interaction manifests itself as feedback returned to the speaker. Such feedback in the form of nodding, verbal consents, eye contact and gestures, functioning like radar, helps the speaker to determine the next step in the course of communication. Whenever the interlocutor feels lost s/he can immediately ask a question or elicit agreement from the other to probe the current status and success of communication and if necessary correct it. It is crystal clear that the writer, in most cases, is deprived of such feedback or at least it takes a fairly long time to receive it. Even in language classrooms where, in a sense, teachers have a duty to provide feedback on writing, it is commonly delayed, often for days at a time. Therefore, the language of writing is widely held to be context independent, and uninterrupted by interactive feedback.

2.3.3 Social differences

There is also a social expectation which makes writing psychologically frightening to some people. As was said before, there is no dialect of written language in most languages. This means that a language such as English has got its well-established, formal and standard version. This prestigious version is socially valued more highly than colloquial ways of speaking. When somebody writes, s/he has to conform to the norms of the written language he or she is using. People tolerate little violation of these rules and tend to stigmatise the writer of any infringements as illiterate. Bacon's well-known epigram that writing makes "an exact man" is still valid and widely accepted by people to imply that writing should be exact. Even student writers are well aware of these prejudices. They are also aware of the relatively permanent nature of writing (compared to transient nature of speech) which turns any piece of writing into a semi-permanent document that can be potentially circulated, read, re-read and judged by an infinite number of readers. Therefore, it is no surprise to find people reluctant to write because of this psychological pressure. When students or others write they are jeopardising their prestige; even one ungrammatically odd sentence or misspelled word can ruin the assessment of otherwise good work. Fortunately or unfortunately, these mistakes are more easily accepted and often overlooked in most contexts of speech. However, there has been a recent movement to relax some strict conventions of writing, at least at discourse level, and give writers some degree of individual authorship (e.g. Dudley-Evans, 1997).

2.3.4 Cognitive differences

Both the linguistic and psychological nature of writing ultimately lead into cognitive pressure imposed upon the writer. As discussed, the writer is deprived of some prosodic and paralinguistic facilities as well as interlocutor's feedback. This often means that a written text should be as explicit as possible. It should contain all the information required to convey the message to the readers and convince them without any further argument. Writing has to create and constitute its own context. No text can achieve these merits unless it is logically well organised. The only way to organise a text is planning and thinking it over and over which requires and develops key cognitive capabilities. Writing with less thinking is characteristic of children's writing but not of adequate academic writing. Students can keep reading for hours and even enjoy it, but the same people may find writing cumbersome and boring after much less time.

2.4 Summing up

The aforementioned differences between speaking and writing were based on stereotypes of differences between these two skills commonly listed as clear-cut binary distinctions. The reality is that there is no distinct boundary between the two. The apparent extremes can get close together, as they do in many cases and whenever it happens writing shows the properties of the other skill and vice versa. For instance with regard to linguistic differences, writing can foster some of the prosodic and paralinguistic features. An example of this case is the stage script or director's notes for a radio play or film in which it is quite common to describe the intonation, pitch of voice and body movement of the actors. On the other hand, the written language is not

always complex. Nowadays, there is a tendency towards using the colloquial speech in modern novels or in e-mail communication and even finding some ungrammatical sentences does not necessarily make readers frown especially if they are held to represent speech. Writing may enjoy feedback, too. Slipping a note to a partner may be accompanied by feedback almost at the same time. Some types of writing are not cognitively demanding tasks, either. Sometimes writing and reading are inextricably bound to each other, such as in reading/speaking lectures or T.V. news.

All in all, speaking and writing are similar in many ways, as they are different in many others. They are, after all, both fundamentally representations of language for communication. It is the intended function that determines which properties are emphasised and to what extent they should be incorporated in a specific task of writing or act of speaking. In short, the function of speaking tells the speaker what the common and shared knowledge is to start communication, whereas the function of writing usually dictates what the most important issue is to write about. It can be concluded that any analysis of speaking and writing without taking their functions into account will be imaginative, but unreal and therefore misleading.

2.5 Writing in history

No wonder writing is occasionally referred to as the fourth skill. This is because it comes last in the traditional sequence of learning the four skills (i.e. listening, speaking, reading and writing) of a language. Arguably, it is the most difficult of these skills. Interestingly, this sequence remained intact in the process of human language development. In other words, writing, compared to other language skills, is a fairly recent achievement. The oldest writing found by the archaeologists dates back to 2500

years BC (Crystal, 1987) and specialists speculate that writing might have been invented some 5000 (McArthur, 1996, p.1036) or 6000years ago (Grabe & Kaplan, 1996, p.5). The word writing originally means *engraving*, *drawing*, and *cutting* (McArthur, ibid) which obviously refers to the early stages of inventing writing people motivated by the need of a relatively more permanent record. The meaningful and influential role of writing in the West is, however, much more recent and for western languages is mainly rooted in the Greek civilisation. Aristotle can be named as the pioneer of studying writing and rhetoric, firstly as a means of effective and logical argument and secondly for its aesthetic aspects. Since then writing remained an important educational subject everywhere. Until the Renaissance the focus in writing classes had been on sentences as the building blocks of texts. It was in the light of this period that writing began to be viewed beyond the boundaries of the sentence, as a result of which writing gradually turned to be synonymous to *composing*. Throughout the nineteenth century composing received proper attention. This was mainly due to the huge leap in science which in turn made literacy an immediate concern for societies. Realising that "knowledge is power" and in order to become a literate member of the society, ordinary people who were the target of literacy work became willing to participate in such classes. To meet the needs of the era, composition classes were mainly aimed at sentence accuracy and to some extent paragraph development.

"There is no doubt that developments in ESL composition have been influenced by and, to a certain extent, are parallel to developments in the teaching of writing to native speakers of English" (Silva, 1990, p.11) and that is why a brief history of teaching L1 writing has been given. However, in the twentieth century, in general, and in the second half of the century, in particular, the distinction between L1 writing and

ESL writing due to rapid changes in trends and research-based theories became so vast that two different accounts should be given to sketch the historical developments of this period.

Within the realm of L1, writing in the twentieth century has been viewed mainly as a means of communication which replaces oral argumentation in order to demonstrate the learner's knowledge. This change, to some extent, can be attributed to its convenience in large classes and over-populated educational institutes. Consequently, writing has lost its position as a subject *per se* and has become a vehicle for other subject matters. This shift in the role of writing from a goal to a tool in the academic environments undermined the importance attached to writing for itself. With the introduction of so-called objective tests , there seemed less room for writing in the curriculum. However, in the final quarter of the century, as the multiple-choice tests began to give place to more integrative productive written examinations, the lost importance has been restored. This is particularly true in those courses in which assessment is largely by coursework or based on cumulative portfolios.

2.6 Methods of composition instruction

2.6.1 Controlled composition

This method, also known as guided composition, is the logical consequence of an Audio-lingual view of language, dominant during 40's and 50's. Following the same principles that language is made of a set of patterns and learning is just habit formation, this method, then, tries to teach writing by utilising strictly controlled grammatical patterns and does not permit any inaccuracy during activities, which include

substitution, expansion, transformation, completion, etc. (Pincas, 1962; Rivers, 1968; Paulston and Bruder, 1976). As learners build up on their writing skills at sentence level, they will be transferred to more realistic discourse in passages. Again here, passages are controlled with regard to their grammatical and vocabulary difficulty level. These passages are merely a collection of sentence patterns which are to be manipulated correctly by the learner and there is little attention to the ideas expressed in the text. Obviously, there is no targeted audience other than the teacher who is supposed to proof-read the compositions and point out all the mistakes. The teaching of writing in this regime is designed to minimise errors.

Pioneers of this method like Fries (1945) hold the idea that writing comes secondary to other language skills and is less important; in fact, learning writing is an instrumental aim and it is not an end in itself.

Although some proponents of this method like Briere (1966) called for a more moderate version by allowing learners to create their own texts, others like Pincas (1962) maintained that such uncontrolled practices are in violation of so-called scientific principles of learning through conditioning, as bad habits (i.e. writing grammatically inaccurate sentences) might take place.

Silva (1990, p.13) bitterly points to the reality that this method is "still alive and well in many ESL composition classrooms and textbooks".

2.6.2 Current-Traditional Rhetoric

Controlled writing, like other methods, has some inevitable shortcomings. The major problem with this method is that it focuses on grammatically correct sentence patterns and hardly considers the product as a text. As a reaction towards this problem, in the

1960s a new movement in teaching writing became fashionable. The underlying

principle of this new approach was that ESL writers write differently from native

speakers, not because they cannot produce grammatical sentences, but because they

are used to a different text arrangement in their mother tongue. The classroom

application of this idea of contrastive rhetoric, initially proposed by Kaplan (1966),

tries to familiarise the language learners with the patterns that native speakers follow in

their thoughts and hence in their writing. These patterns based on contrastive rhetoric

were primarily taught at paragraph level and later on stretched to the whole essay.

Silva (1990, p.14) concisely explains the tenets of this method:

> Here attention was given not only to its elements (topic sentences, support
> sentences, concluding sentences, and transitions), but also to various options
> to for its development (illustration, exemplification, comparison contrast,
> partition, classification, definition, causal analysis, and so on). The other
> important focus was essay development, actually an extrapolation of
> paragraph principles to larger stretches of discourse. Addressed here were
> larger structural entities (introduction, body and conclusion) and
> organizational patterns or modes (normally narration, description, exposition,
> and argumentation), with exposition typically seen as the pattern most
> appropriate for use by university-level second language writers.

The pedagogical procedures prescribed by this method were mainly those of working

with suggested topics and pre-written paragraphs. Students are usually required to

identify and arrange topic sentences, support sentences, and so on, from a list of

scrambled sentences and then form an outline which will eventually lead into a

composition almost always with a nature of academic writing. The reader and the final

judge is the teacher.

As Silva (1990, p.14-15) points out, despite all the criticism, the current-traditional approach is still strongly reflected "in many of the most well-known and popular contemporary ESL composition textbooks."

2.6.3 The process approach

Once again with the failure of the current-traditional approach in enabling L2 writers to produce native-like texts, a new method began to claim teachers' allegiance. This new method, which had its roots in the expressive school, blamed its predecessor for viewing writing in a too simplistic, linear, and prescriptive way. Well-known researchers such as Raimes (1987 and 1983), Zamel (1976 and 1983, 1987), Spack (1984), Hamp-Lyons (1986a) and Kraples (1990) came to the conclusion that writing is a complicated and recursive exploring process through which the writer thinks, learns and learns to think. Therefore, what learners need is guidance not control. The emphasis, here, is put on the communication rather than the form (Widdowson, 1978; Reid, 1993). So, writers are provided with ample time, planning, writing and revision and little interference from the teacher (Taylor, 1980; Hedge, 1988; White, 1988). In the process approach, the teacher relinquishes her traditional omniscient, omnipotent role. Peers, instead, are encouraged to read and revise their classmates' writings. Therefore, collaborative writing is of paramount importance in this method (Peacock, 1986; Witte and Cherry 1986; Robinson, 1988).

The process approach has its own critics. It has been claimed (Horowitz, 1986a) that this method cannot prepare students for academic writing since in this method it is typically the writer that identifies and fulfils the task, whereas in real life students hardly get a free choice of topic. Accordingly, the final product may not conform to

the established norms of academic writing. Besides, this approach seems to be too *lenient* as process-oriented teachers are supposed to be more *humanistic* and respond to the learner instead of only to their writing. This can create a false impression that there is nothing wrong with the final product. Another disadvantage of the process method is that all the writing and re-writing stages are often found to be boring by learners with little interest in writing.

The process approach, altogether, can be motivating for students, but it is time-consuming and cannot work in academic settings unless it is combined with the current-traditional approach (Robinson, 1988).

2.6.4 English for academic purposes (EAP)

As the majority of writing research and instruction takes place in academic environments and for academic purposes, the success or failure of a method is usually judged based on its performance in such situations. The process method has been frowned upon by many academics (Reid, 1984; Horowitz, 1986a; McDonough, 1986; Barra, 1993, among others) who highlighted several shortcomings with its outcome in preparing students for their future writing requirements. They accused the process approach of neglecting to consider "variations in writing processes due to differences in individuals, writing tasks, and situations; the development of schemata for academic discourse; language proficiency; level of cognitive development; and insights from the study of contrastive rhetoric" (Silva, 1990, p.16). Since it was also accused of creating classroom situations that bear little resemblance to real ones and not addressing certain types of writing tasks which may be academically important, the process approach had to be replaced with (or supplemented by) another alternative that presumably could

cater more appropriately for the students' needs. As the name suggests, this alternative, stressed *English for Academic Purposes (EAP)*. Therefore, this approach intends to provide the students with those language skills which they will need to survive in an academic environment. In order to do so, students must learn to become members of certain discourse communities (Swales, 1984b, 1990). The major debate is to what extent language teachers should teach their students' related academic and scientific content (Horowitz, 1986b; Braine, 1988; Johns, 1988) and what modes of writing should be taught in academic writing classes. In this regard, Kroll (1990b, p.141) reiterates that "there are competing interpretations of what a teacher's role should be and even of what modes or types of prose should be produced in the composition classroom".

The present study can suggest a tentative answer to these controversies by analysing all four basic modes of discourse and pin-pointing any probable differences among them so that writing teachers can have a better understanding of writing tasks.

Section 2: Some issues in testing

2.7 Introduction

As this thesis is meant to be in the domain of applied linguistics, one might ask why language testing is significant and how it can cater for language teachers' needs. The question is quite relevant and there are logical and well-grounded answers, too. Historically, evaluation has always been an inevitable part of education for it serves educators in different ways directly or indirectly. The educational purpose for which a language test is designed is known as the function of a language test.

Generally speaking, there are two major test functions: *prognostic* and *evaluation of attainment*. The former, unlike the latter, is not based on any particular taught subject matter or textbook context or specified course of instruction. Instead it is meant to enable the test-users to make sound decisions about the likely future of the test-takers, for example to ascertain a suitable occupation for them to pursue or the most appropriate foreign language to study. *Placement, selection* and *aptitude* tests are instances of prognostic tests whereas *achievement, proficiency* and *knowledge* tests fall in the evaluation of attainment category. Achievement tests can be divided into those which are *general* and those which are used for more specific *diagnostic* purposes. All these functions of tests can therefore be categorised as shown in Fig.2.3:

Fig.2.3: Functions of language tests (from Farhady et al., 1995)

47
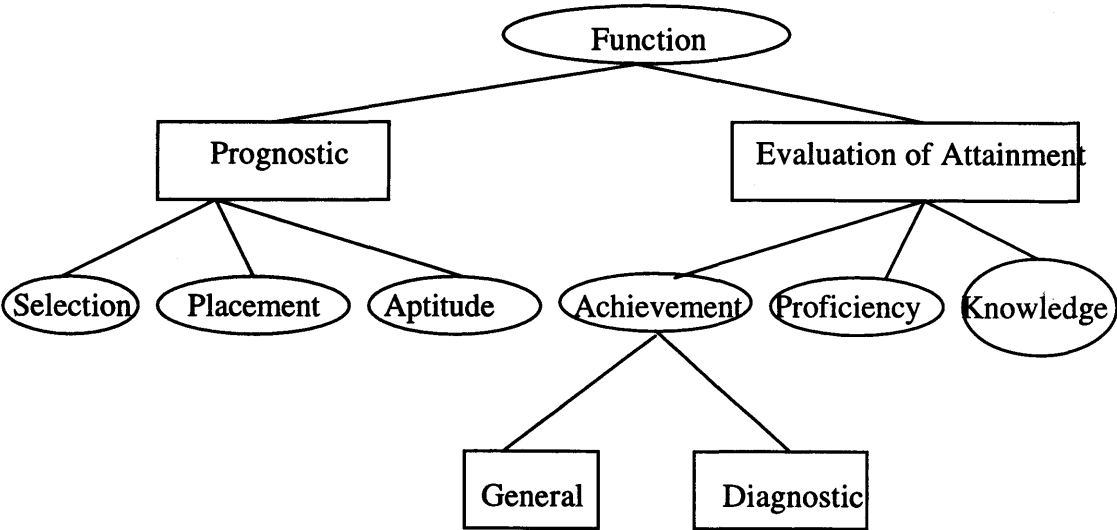
2.8 Characteristics of a good test

Whichever function a test is supposed to serve, it has to show certain characteristics to establish itself as an accurate device to assess the testees. These characteristics are *reliability, validity,* and *practicality,* which will be discussed in some detail in this chapter.

2.8.1 Reliability

Reliability refers to the degree to which a test yields the same results in different administrations. It means that the results of a reliable test are consistent over various administrations. However it is very unlikely that virtually the same results on a second observation will be obtained. Any change can be due to *"systematic variation"*, such as learning, or unpredictable factors such as the test-taker's mood during the assessment session, which is known as unsystematic or error variation. Unlike error variation, systematic variation increases the reliability of a test.

It is obvious that in any measurement there is some degree of error, and the obtained score is not the subject's real potential score. Therefore, in order to calculate the true score we need a certain index to account for this inaccuracy. In statistical terms this index is called the Standard Error of Measurement (SEM) which can be calculated using the following formula in which S_x refers to the standard deviation of observed scores and r stands for reliability:

$$SEM = s_x \sqrt{1-r}$$

In other words, the observed score is not necessarily the true indicator of the testee's performance. Instead a range of scores based on the "observed score $\pm$ SEM" would be a more careful interpretation.

It has been argued, so far, that in any assessment one should take the SEM into account. However it seems that reliability is of greater priority as it is impossible to calculate the SEM without calculating the reliability beforehand. In fact, reliability is simply estimated, not calculated. The reason is that the true score can only be approximately estimated.

Reliability can be estimated in two ways either through measurement of stability or through measurement of equivalence. Four major methods of estimating the reliability of a test are discussed below.

2.8.1.1 Test-Retest Method

Reliability, by definition, means the consistency of the test results in repeated administrations. It implies, therefore, that a way of estimating the reliability would be giving the same test to the same subjects twice. The correlation coefficient between the two sets of scores, then, would be the reliability of the test. It should be noted that there must be a reasonable time lapse between the two administrations and some care should be taken to prevent any significant learning or other kinds of interference taking place between the two test administrations. It is evident that meeting these requirements, i.e. administering the same test to the same subjects under exactly the same conditions without intervening changes, is practically impossible in usual classroom circumstances and that is why a second method seems more convenient.

2.8.1.2 Parallel-Form Method

In order to avoid the trouble of gathering the same examinees under the same conditions to take the test for a second time, this method prefers to administer a parallel version of the same test. The second parallel version will then be administered on a second occasion. It is true that the parallel-form method solves some of the problems associated with the test-retest method, yet it creates another set of problems. Writing two sets of parallel tests is not an easy task. In fact, it is very awkward to write two tests in two different forms but aiming to evaluate exactly the same skills and subskills. In addition, there remains the problem of intervening changes affecting the testees or the testing. Consequently, it seems that problem remains unsolved and there is a need for a third solution.

2.8.1.3 Split-half Method

In an attempt to alleviate the burden of writing two similar tests or administering the same test twice, both of which are unrealistic and difficult to administer (Feldt & Brennan, 1989), it was suggested that halves of a single test could be used as two parallel tests to be administered on a single occasion. This method, compared to the other two, seems more convenient as long as the process of splitting the test into halves is done carefully. To ensure that the two parts are parallel it is advisable to arrange test items in an easy-to-difficult order and then divide the test into odd and even items. The odd and even items are considered as two groups, or levels, of the test. In the next step, the correlation between the two halves should be calculated. It has to be pointed out that this correlation is NOT the reliability of the whole test but of one half of it which is always smaller than the reliability of the total test. To calculate

the total reliability of the test the following formula, known as the Spearman-Brown prophecy formula, should be used:

$$r_{(total)} = \frac{2(rhalf)}{1 + (rhalf)}$$

The major shortcoming of this method is the difficulty of developing a homogeneous test which can be divided into two parallel halves. Achieving such homogeneity while avoiding any dependency among the items, especially when a test comprises a series of subtests aiming at measuring different language skills and subskills, would be extremely difficult. Therefore, there was still a need to develop an even easier method of estimating reliability.

2.8.1.4 KR-21 and KR-20 Methods

One major disadvantage of the split-half method is that a test should be divided into two parallel halves, which is sometimes very difficult to do. An alternative would be splitting a test into all possible ways, computing the reliability coefficients for all these splits and finally the average of these coefficients would be the reliability of the test. As Wood (1990, p.136) said we can "imagine a set of items or questions or tasks or markers or raters being formed into all possible permutations and the aggregate scores on all these forms correlated in all possible pairs." In practice this procedure with fairly large tests would be really difficult, if not impossible as there would be so many coefficients to calculate.

Instead, there are fortunately two formulae which are fairly simple and easy to handle.

These formulae, which are named after Kudar and Richardson, the two statisticians who developed the formulae, are quite simple. The first one KR-21 requires only the mean and the variance of the test. Besides there is no need to design a homogeneous test nor does it require a parallel version or a second administration of the test. In fact this method measures the relationship between a test and another hypothetical form of it.

This non-correlational formula is given below.

$$(KR\text{-}21)\ r = \left(\frac{k}{k-1}\right)\left(1 - \frac{\bar{x}(k-\bar{x})}{kv}\right)$$

where:

K= the number of the items in a test

$\bar{x}$ = the mean score

v= the variance

However, when a reasonable degree of item homogeneity does not exist and a more precise measure of variance is required, another formula known as KR-20 should be used.

Still, there is another formula known as *'Cronbach's alpha'* or *'coefficient alpha'* which can be regarded as the basis of all formulae based on the analysis of variance components.

Although all the aforementioned types of reliability are widely used in the field of language testing, they are not applicable for essay tests. The reason is that in essay tests all the hindering obstacles for test-retest reliability estimates are present, while there are no items to divide into two to run split-half tests. For the same reason no internal-consistency-based reliability test can be used for such tests. Instead there are two more types of reliability which are central to direct tests of writing: *inter-rater reliability* and *intra-rater reliability*.

As these types of reliability are directly related to this research project, they will be addressed in more detail.

2.8.1.5 Inter-rater and intra-rater reliability

Not all tests include a number of items. Rather, in cases like evaluating compositions or oral fluency, scores are usually presented as an average of a number of scores given by two or more judges. In such cases it is necessary to estimate the inter-rater and intra-rater reliabilities.

Inter-rater reliability shows the extent to which there is agreement among the judges in ranking the subjects. *Intra-rater reliability* shows how consistent raters are in their judgements. Although some scholars (Wood, 1990) believe that the latter is theoretically more important than the former, in practice it is the inter-rater reliability which stands at the focal point of attention for most research practitioners. Some

scholars (Hatch & Lazaraton, 1990) do not even mention intra-rater reliability at all. Brown (1996, p.206) referring to problems associated with intra-rater reliability, such as the results being "confounded by the raters' remembering, on the second occasion, their ratings from the first occasion", states that this form of reliability is not usually reported as often as the inter-rater reliability estimates,

In order to compute the intra-rater reliability, we need at least two or more markings for each sample. These scores are obtained through making the rater evaluate the sample in intervals. These scores can be, then, correlated with each other using, say, Spearman rank-order coefficient. Alternatively the rater consistency can be examined by adding these independent scores and applying the coefficient alpha formula. For inter-rater reliability Brown (1996) recommends that whenever there are two raters their ratings can be correlated as the estimate of the reliability, and then it would be advisable to adjust the result using the Spearman-Brown prophecy formula but for more than two raters it is better to sum the ratings and put them in the coefficient alpha formula.

So far we have examined the advantages and disadvantages of different methods of estimating reliability. It is true that some of them are more easily compared to other ones and KR-21 is the most practical one, but each method has a certain nature which makes it more suitable for certain situations and some require certain assumptions. For example, while the split-half method emphasises the internal consistency of the test items, the test-retest method reveals the consistency of a test over time. The latter method is also the only one to account for the sources of error outside the test itself.

2.8.2 Factors influencing reliability

There is not unanimous agreement about the acceptable level of reliability for a test. Some writers (e.g. Ingram, 1977) put it as high as 0.95 but, to be realistic even a marginal reliability of 0.70 can be acceptable (Harris, 1969). It all depends on the importance of the decisions to be taken based on the test scores. This relativity has been emphasised by Thorndike & Hagen (1969), Guilford & Furchter (1978) and Farhady et al. (1995).

Within the last two decades, as a result of many rigorous research studies and vigorous training programmes, the score reliability has improved considerably. Most major test batteries boast a reliability of 0.80 or more which is "commonly regarded as a satisfactory level for decision-making purposes" (Hamp-Lyons, 1990, p.69).

No matter how we set the acceptable standard, there are certain factors which can affect the reliability of a test. While Heaton (1975) enumerates five such factors, Farhady et al. (1995, pp. 140-8) classify them as follows:

1) **Testees**: as testees are human beings, their performances fluctuate as their moods change. These unpredictable changes can increase or decrease the estimated reliability. As a general rule, according to Harris (1969, p. 14), "the more samples of students' performance we take, the more reliable will be our assessment of their knowledge and ability." This assumes that any fluctuations will be averaged out, but samples will probably need to be taken over time and the larger the time span for sampling the more likelihood that intervening variables (including learning) may affect scores. In addition, the homogeneity of subjects can cause an under-estimation of the reliability of

the test, whereas a large variance among the subjects' abilities can falsely increase the reliability estimate.

2) **Test factors**: There are a few test-related factors which may influence the reliability to a larger degree compared to the changes in the learners' conditions. These factors can be further divided into three sub-categories:

a) *Structure of the test*

A test with homogeneous items aimed at a certain skill or trait can yield more precise and stable results. This stability can be enhanced by increasing the length of the test and incorporating at least 35 items in it. Speed is another important element in evaluation. Basically there are two types of test, namely a speeded test and a power test. In speeded tests, the learner is given a very limited time to answer relatively easy questions. It is not possible to estimate the reliability through split-half or rational equivalence technique and therefore other techniques such as test-retest or parallel forms should be applied. The reason is that in these tests items are easy and testees commit fewer errors, though they may leave a few items not attempted. Consequently, the correlation between the odd and even items would be extremely high.

Direct tests of writing are a combination of power tests (there is no limitation to the writer's performance) and speed tests (these tests are usually timed); however it seems that there is a tendency, as implied by the nature of process approach and portfolio evaluation, to decrease the speed dimension and add to the power dimension.

b) *Administration factors*

There is a close relationship between the test reliability and test administration because a reliable test is supposed to give almost the same results when administered again, and hence it can be concluded that the test administration elements should remain unchanged. To achieve an acceptable administration enough care should be taken to stick to the time allowed, to make instructions as clear as possible, and to prevent any irregularity which can affect the subjects' performance, such as distracting noises, and to consider the physical conditions of the test environment. In an essay test, all these factors, especially the perceived clarity of the given topic and rubric, can drastically change the results.

c) *Scoring procedures*

Besides the structure of the test itself and the administrational procedures, the way a test is scored can affect its reliability. There is no doubt that objective tests, despite their validity shortcomings, are more reliable and more convenient when it comes to scoring. Multiple-choice item tests are a perfect, commonly used example. Even completely subjective tests such as compositions can be improved by setting a scoring profile and training the raters.

Intra-rater reliability may be affected by inconsistencies either in rating criteria or the way they are applied, or both. Raters usually shift their stress from one criterion to another as they proceed with their scoring task. For example, in a writing evaluation scheme little weight should be given to the handwriting, but in reality a student's text with tidy and calligraphic handwriting can, and usually does, attract the rater's attention resulting in a score which is artificially higher than it should otherwise have

been. This can cause inconsistency in raters' judgements. Besides, even if they stick to their criteria and their weightings, there is no way to make certain that these criteria are applied correctly and objectively in relation to each script . Raters inevitably compare each candidate against the previous one and it is, to some extent, the performance of the previous candidate that determines a given score rather than that writer's own performance. Similarly, inter-rater reliability is subject to same variations. Different raters involved in an evaluation programme may interpret the criteria in different ways or even, in unlikely cases, where they share the same view of the criteria, they may still apply them in different ways.

Research shows that a number of other testee variables such as gender, social background, mother tongue, or expertise in certain disciplines, can influence composition scores and hence the reliability of the test (Takashima,1987; Siegel, 1982; Brown, 1991; Hamp-Lyons, 1991).

It is a well-known maxim in the field of language assessment that reliability is necessary but not sufficient, which means another important area must be investigated – validity

2.8.3 Validity

Validity, by definition, means the extent to which a test measures what it is supposed to measure. Compared to reliability, validity is a more subjective concept and unlike reliability it cannot be independently statistically calculated. Probably that is why it has not been generally regarded as important as reliability. However, in recent years validity has been the main concern of test writers and the academic establishments that use the test results (Weir, 1983; Ruth and Murphy, 1988;Hamp-Lyons, 1990, 1991).

Nowadays, validity is a growing issue, consideration of which is bringing about major changes in language evaluation.

A good illustration is the Test of Written English (TWE). Before its introduction, TOEFL used to measure the candidates' writing skill by testing vocabulary, grammar, and other components of writing. Although the results were highly reliable, the results could not satisfy the educational organisations which were supposed to rely on the test results, typically for university admissions. They had the experience of admitting students with scores indicating a good command of English syntax and/or a wide range of vocabulary who were not, as it became apparent, good writers when they came to write assignments or other work in their new institution. In other words, the multiple-choice items used to assess the learners' writing skill were measuring something else and were therefore lacking a crucial requirement of a good test: validity.

As mentioned, validity is not a single, isolated, independent measure. It is rather a complicated, test-dependant issue for which different aspects of a language test should be taken into account and for this reason there is not one type of validity but several types which are discussed below.

2.8.3.1 Content validity

Content validity, as the name reveals, concerns the content of the materials included in a test. It means that for a test to achieve content validity, or appropriateness of the test as some experts call it, should incorporate an acceptable amount of the content to be tested in the test. This content must also be appropriate for the degree of learning that the testees are required to achieve. The major drawback of content validity lies in the subjectivity of judgements made about the degree of appropriacy. Like other

subjective judgements, one remedy is to ask for more than a single person's opinions. Alternatively it is possible to define the details of the content to be tested and then transfer these details onto a Table of Specifications. Another problem with content validity is that it can not be expressed in statistical values. However, other types of validity have this advantage, if it is an advantage.

2.8.3.2 Face validity

Face validity means whether or not a test looks valid. So a test, to be face valid, has to look acceptable, at least to the test-takers. For example, a test of reading is supposed to involve some sort of reading activity or it might look irrelevant to non-experts. Integrative tests, such as cloze, usually do not enjoy a high level of face validity. It is evident that this is not a well-grounded criterion and for this reason face validity is not seriously regarded as a viable type of validity. However it is advisable to bear this concept in mind while designing a test, as it can sometimes either motivate or frustrate the examinees.

Fortunately, direct tests of writing have an outstanding face validity and for this reason both educational institutes and individual learners have a positive attitude towards them (e.g. Carlson, 1983; Keller-Cohen and Wolfe, 1987; Criper and Davies, 1988).

2.8.3.3 Criterion-related validity

A test of writing cannot be convincing unless using it proves it yields results similar to a previously established test. So by criterion-related validity we mean the correspondence between one test and another acceptable one intended to evaluate the

same thing. Although there is no limit in choosing a criterion, it should be noted that the more widely realised the criterion the more convincing would be the result.

There are two types of criterion-related validity which according to the time they are administered: *concurrent* and *predictive*.

Concurrent validity can be achieved by administering the new test and a criterion test simultaneously and then calculating the correlation coefficient existing between the two sets of scores. A high correlation proves that the newly developed test measures the targeted trait in almost the same way that the criterion does.

Predictive validity works in almost the same way as concurrent validity. The difference is that the two tests are administered at two different times. This may, of course, introduce other variables, as mentioned earlier.

It has been strongly established and generally accepted that direct tests of writing correlate highly with more objective tests (e.g. Nemati, 1993).

2.8.3.4 Construct validity

While other types of validity establish how well a test measures a particular trait, construct validity deals with the trait itself and tries to ascertain whether or not that trait, also known as a construct, is psychologically real.

For instance, a test of writing skill is supposed to measure a supposedly existing skill called writing and it would not achieve construct invalidity if the test measures one or more other skills or subskills.

Among different types of validity, construct validity can be named as the most important and the most complicated one. This is important because without this type of validity a test is not only of little value but also can be misleading as it tries to

measure something which does not exist. This is so important that in Hamp-Lyons'

words (1990, p.72) it is "the overarching validity which subsumes and, in some sense,

consumes all the others." . Construct validity is a complicated issue because it is

highly subjective and tremendously difficult to establish. It is usually established

through a delicate statistical procedure known as factor analysis.

According to Hamp-Lyons (ibid, p.73), there are four components in any direct test of

writing which affect the validity of the test: the task, the writer, the scoring procedure,

and the reader(s). The present study, in fact, questions the task validity of the current

practice.

2.8.4 Factors influencing validity

It has been argued that validity is a vital, yet delicate requirement for any good test.

This delicacy makes validity highly vulnerable. Different types of validity have been

discussed previously and the violation of any of them can be a serious threat to the

total validity of the test being used. Fortunately most of these mistakes can be avoided

relatively easily by a careful test-user; however the following factors must always be

regarded as probable pitfalls:

1) Directions: Directions should be as clear as possible and leave no ambiguity to the

test-takers with regard to the task they are supposed to perform, the time allowed,

penalties and all other relevant Do's and Don'ts.

2) Difficulty level: a test should not be too easy nor too difficult for the testees.

3) Structure of item arrangements: test items must be constructed in such a way that the testees' inability to answer could be attributed only to their lack of knowledge in that particular point but not to other reasons, say misunderstanding caused by an ambiguous question.

4) Arrangement of the items and correct responses: it is recommended that questions in a test are recommended to be arranged in order of difficulty with easiest questions first, to encourage the test taker to go ahead. Besides, in a multiple-choice test item responses should be arranged randomly to avoid any predictable pattern, which can invalidate the test.

2.8.5 Reliability versus validity

Although reliability and validity seem to be two separate concepts, they are closely related to each other. Reliability, as discussed before, is simply an independent statistical index of consistency in a test result. So, even a very high degree of reliability does not guarantee the validity of the test results as validity is a subjective and relative concept which depends directly on the purpose for which the test is used. Nevertheless, a valid test does show at least a moderate degree of reliability (at least as the square of the validity) which can be statistically proven but further details would be beyond the scope of this thesis. It is worth recalling that validity is the correlation between two tests and that if the tests are not reliable (yielding consistent results), they cannot correlate. That is why validity is more important than reliability.

Considering the issue of testing writing, it seems that usually there is a trade-off between validity and reliability. It sounds an obvious argument that offering a choice of essay topic will make the process of marking more difficult as raters are supposed to

compare different compositions. This means that a choice of different topics means lower inter-rater reliability. On the other hand, as Polio and Glew (1994, p. 45) argue "it is possible that forcing students to write on a particular topic renders the test less valid." Therefore, Tedick (1993) and Weigle (1994) warn that reliability is usually achieved by sacrificing validity and, in short, validity and reliability are sometimes incompatible (Henning, 1987). The dilemma of offering a choice of topics or not is central to the present study.

2.8.6 Practicality

Besides validity and reliability there is a third criterion that a good test should meet: this is *practicality*. Tests are meant to be used in real life and in ordinary educational situations and if a test is not practical, however reliable and valid it might be, it would be of little use. Generally speaking, a practical test is easy to administer, easy to score, and easy to interpret and apply.

For a test to be practical it has to be easy to administer. In other words, it has to be clear, with directions given in plain language. Besides, a test with too many subtests can be quite confusing for both teachers and students. And above all, the allotted time for a test is another important factor which must be calculated with caution. Too long a time needed for a test can be a disadvantage. All in all , a test has to be convenient to both administrators and test-takers.

Probably the most important factor in making a test practical or impractical is its scoring procedure. If a test is intended to be frequently used in the classroom by teachers or utilised in a large scale evaluation it needs to be easy to score. Maybe this is why multiple-choice tests became so popular in the 60's and 70's. Although today

composition tests are becoming fashionable again, researchers are still trying to find more objective and easier methods for scoring this type of tests, too.

Finally, the results obtained from a test should ideally be easy to interpret. This is not as simple as it might seem when it comes to norm-referenced tests as there is a need to transfer the raw scores and provide a table of norms, too.

Section three: Some issues in writing assessment

2.9 Introduction

Writing, compared to other skills and components of language, has always suffered for its alleged lack of an established reliable and valid assessment method. This Achilles' Heel has crippled both teachers and researchers working on writing ability and, as Hirsch (1977) put it, it is the "single most important snag to practical progress in composition teaching and research". After all, teachers and researchers have to evaluate their students' writing ability to realise any progress or weakness in a methodology, teaching materials, or even in the test itself.

Generally speaking, methods of assessing writing ability can be divided into two opposing major groups which are entitled differently in different books and articles. Labels such as 'direct and indirect', 'objective and subjective', 'quantitative and qualitative', 'suppletion and recognition' 'multiple-choice and composition or essay-type', have been used to brand these dichotomous approaches. Whatever the name, the former method tries to assess writing ability through writing related components

whereas the latter tackles the issue by dealing with the actual writing samples. The following is a brief but relatively thorough account of the two.

2.10 Indirect methods of assessing writing

Carroll and Hall(1985) define objective evaluation as:

> A process best done by getting testees to select the correct form from a number of given options. These should, therefore, no reason for variation in scoring as between different markers. (p.3)

This method usually involves a set of multiple-choice items in which the testee is required to choose either the (most) correct or the only wrong option (error recognition). The former is mainly used to examine such components of writing as grammar, spelling, or punctuation while the latter is normally used to recognise the problem with a defective sentence. The Michigan English Assessment Battery (MELAB), the Test of Standard Written English (TSWE), and section 2 of the Test of English as a Foreign Language (TOEFL) are all well-known examples of objective tests.

Although they are currently somewhat out of favour, objective tests of writing have survived mainly due to two merits. Firstly, they boast of a high degree of statistically calculated reliability, and hence may be considered scientific. This characteristic was regarded as a highly-valued asset especially in the heyday of the so-called psychometric era. Secondly, this type of test is cheap to administer and easy to score. It is not surprising, then, to see that the major testing batteries employ this type of test to run their world-wide business. Besides, advocates of direct writing tests refer to a few

studies (e.g. Godshalk et al., 1965; College Entrance Examination Board, 1966; Breland and Gaynor, 1979; Culpepper and Ramsden, 1982; Nemati, 1993) that demonstrate a reasonable but not high correlation between indirect tests of writing and essay tests. Ackerman and Smith (1988) refer to this point, asking why moderate correlation coefficients between direct and indirect tests of writing are not higher. They conclude that these two methods assess dissimilar skills and therefore are not of equal value. In other words, objective tests do evaluate only certain abilities required for writing skill. Harris (1969) names the major areas touched by this type of test as the following:

a)-Formal grammar and style (including subject-verb agreement, structural parallelism, case of pronouns, comparison of adjectives, etc.)
b)-The ability to organise materials
c)-The mechanics of writing i.e. punctuation, capitalisation and spelling

With regard to test validity, the most one can say in favour of direct tests of writing, is that they "do possess a certain degree of predictive validity" but nothing more (Munro, 1991, p.17).

Despite the fact that objective tests gained an amazing popularity in North America, they were never regarded as a serious method of language testing on this side of the Atlantic. From the very beginning, objective tests of writing have always been under bitter criticism from the side of teachers, researchers, parents and, above all, the educational institutes which were supposed to receive the test results and make decisions accordingly. Wiseman (1949), Wiseman and Wrigley (1958), and later Britton (1975) were among the first researchers who stood against direct writing tests, warning that these tests can undermine the importance of the writing skill. Now, after

decades of challenges, objective tests of writing seem to be utterly defeated. In fact, the last stronghold to collapse was TOEFL. The introduction of Test of Written English (TWE) in 1986 was, in a sense, the white flag of surrender raised by the camp of indirect assessment proponents.

Objective tests of writing had to be abandoned since they suffer a number of fundamental flaws. First of all, it should be pointed out that, as a common misconception, teachers think that multiple-choice items are as easy to construct as they are easy to administer. This is far from reality. Writing a good test stem and selecting acceptable distracters is a tedious undertaking even for professionals, let alone less-experienced teachers. Objective tests are called objective because of the objectivity of the way the test items are scored. Put another way, objectively scored tests can be constructed entirely subjectively and vice versa. Secondly, despite what is frequently mentioned, objective tests do not have an acceptable predictive validity because this prediction is more efficient about poor performance on objective tests since "good performance on the tests does not necessarily predict good performance in writing" (Pilliner, 1977, p.18). Multiple-choice examinations even lack superficial face validity. Although this is the simplest type of validity, it has an important effect on people involved in the test. It seems awkward and unrealistic to take a writing test without writing a word. Ackerman and Smith (1988) point out that the objective method of writing assessment makes little or no attempt to measure unity, content or organisation because the examinee is not involved in actual writing. More importantly, the objective test model lacks other types of validity. Briefly speaking, this model fails to take the testees' future academic writing needs into account (content validity), to reflect the psychological reality of writing behaviour in the test (construct validity), to

highly correlate with other measures (criterion validity), with other test scores collected at or about the same time (concurrent validity) or at a significantly later time (predictive validity).

It is still true, as mentioned above, that objective tests do show a correlational relationship with essay tests; however, a correlation does not prove a cause and effect relationship. Besides it should not be forgotten that correlation = r; variance = r^2; therefore, even in a desirable case of achieving a correlation of 0.80, 36% of the variance in test score will be left unaccounted for (Hamp-Lyons, 1990). As a result of all these shortcomings, even well-constructed objective tests can maintain only a minimum degree of validity which makes the results shaky for both the test taker and the users of the test results (Perkins, 1983; Weir, 1990). So, as Spolsky (1995) put it, the dichotomy between direct tests of writing and essay tests became a choice between unreliable testing of writing or reliable testing of something else. Today, it is evident that educational experts have unanimously chosen the former and as a result objective tests "have not only been defeated but also chased from the battlefield" (Hamp-Lyons, 1990).

2.11 Direct assessment of writing

In an attempt to define direct assessment of writing Hamp-Lyons (1991) gives five characteristics of a direct test of writing. First, the test-taker must write a piece of text with a reasonable length of at least 100 words. Second, the writer is required to respond to a prompt. Third, every written text is read and rated by at least one reader-judge. Fourth, the rater's judgement is based on some sort of scale. And finally, the

reader expresses his or her judgement as a number. These characteristics, however, seem to be more applicable to holistic evaluation as they fail to take frequency-count methods into account.

Historically speaking, the direct assessment of writing was the only method practised prior to the domination of the psychometric era in the 1950s and 1960s. In these two decades multiple choice tests replaced essay tests, but again in the 1970s and afterwards essay tests began to receive more attention due to the present tendency towards humanistic and communicative task-based learning. This approach, compared to indirect assessment of writing, has received more approval from researchers and this approval has steadily been increasing as this approach is adjusted in line with the research results. Heaton (1990) believes that the only possible way to evaluate a student's writing ability is by means of a composition test. Weir (1990) recommends writing tasks because of their greater construct, content, and face validity and washback effect. He enumerates four basic advantages for essay tests:

1. The essay has traditionally been accorded high prestige as a testing technique...

2. the topics are extremely easy to set and it is a familiar testing technique to both the candidates and the users of the results

3. It is a suitable vehicle for testing skills, such as the ability to develop an extended argument in a logical manner, which cannot be tested in other ways.

4. The big advantage it shares with other tests of extended writing is that a sample of writing is produced which can provide a tangible point of reference for comparison in the future. (p.60)

Composition tests, however, have not gone uncriticised with regard to their reliability, validity and practicality. The major flaw in such tests is the difficulty of quantifying performances and representing them as scores (Oller, 1979). Second, in composition

the issue of avoidance is a problematic controversy. This means that subjects can use whatever structure they wish and thus skip any structure, vocabulary, or even ideas which they have not mastered and so mask their weaknesses (Harris, 1969). Another drawback with essay tests is the imagination factor involved in such tests. In fact, these tests depend on the testee's imagination, creativity, background or cultural knowledge. These factors may or may not be taken into account by scorers -clearly, in some cases they could be confusing variables.

The main weak point of the essay test is its scoring method. It was pointed out earlier that objective tests can obtain a very high level of reliability which seems too far a target for direct tests of writing to achieve. In Ackerman and Smith's (1988) more extreme negative view, the scoring procedures of essays is so poor as to outweigh their advantages. Ironically, composition tests, nowadays, easily obtain a reliability of 0.80 which is, by any standard, fairly high and acceptable. Now it seems warranted to examine different scoring procedures to appreciate the cause of these controversies. Although methods of scoring essays have been categorised differently in different sources four or five major marking methods will be discussed here.

2.12 Objective vs. holistic assessment of essay tests

It is very important to notice that the applications of terms *holistic* and *objective* may be confusing, because each denotes two different concepts. The dual meanings of holistic will be discussed later in this section but two different usages of the term "objective" will be clarified here. By *objective tests* most authors mean indirect evaluation of writing skill mainly by means of multiple-choice items. In this sense

objective tests contrast with essay tests in which the test taker is required to produce a real piece of discourse. The term *objective scoring* is used to refer to those methods that assess elicited essays objectively and quantifiably by various kinds of error-count method or other objectifiable features. Although these methods make the subjects write compositions, these compositions are rarely read for meaning or their communicative efficiency. Instead, occurrences of certain categorised errors are counted. To avoid further complicating an issue, already misleadingly confused in the literature, Figure 2.4 gives a general view of classifications in second language writing assessment. The following sections discuss the objective and subjective sub-categories in more detail.

Figure2.4: Methods of Evaluating Writing Skill

```
                    ┌─────────────────────────────┐
                    │ Methods of Evaluating Writing│
                    └─────────────────────────────┘
                         /                    \
                   ┌─────────┐           ┌──────────┐
                  (  Direct  )          ( Indirect  )
                   └─────────┘           └──────────┘
                        │                     │
              ┌──────────────┐      ┌────────────────────┐
              │ Essay Tests  │      │ Multiple-Choice Tests│
              └──────────────┘      └────────────────────┘
                   /      \                  │
         ┌──────────────┐ ┌──────────────┐ ┌──────────────────┐
        (Objective Marking)(Subjective Ranking)(Objective Scoring)
         └──────────────┘ └──────────────┘ └──────────────────┘
                │              │                /        \
                │              │      ┌──────────────┐ ┌──────────────────┐
                │              │      │ Clerical Staff│ │Electronic Machines│
                │              │      └──────────────┘ └──────────────────┘
                │              │
     ┌──────────────────────┐ ┌────────────────────┐
     │Frequency-Count Method│ │ Holistic Scoring    │
     └──────────────────────┘ │ Primary-Trait Scoring│
                              │ Multiple-Trait Scoring│
                              │ Analytic Scoring    │
                              └────────────────────┘
```
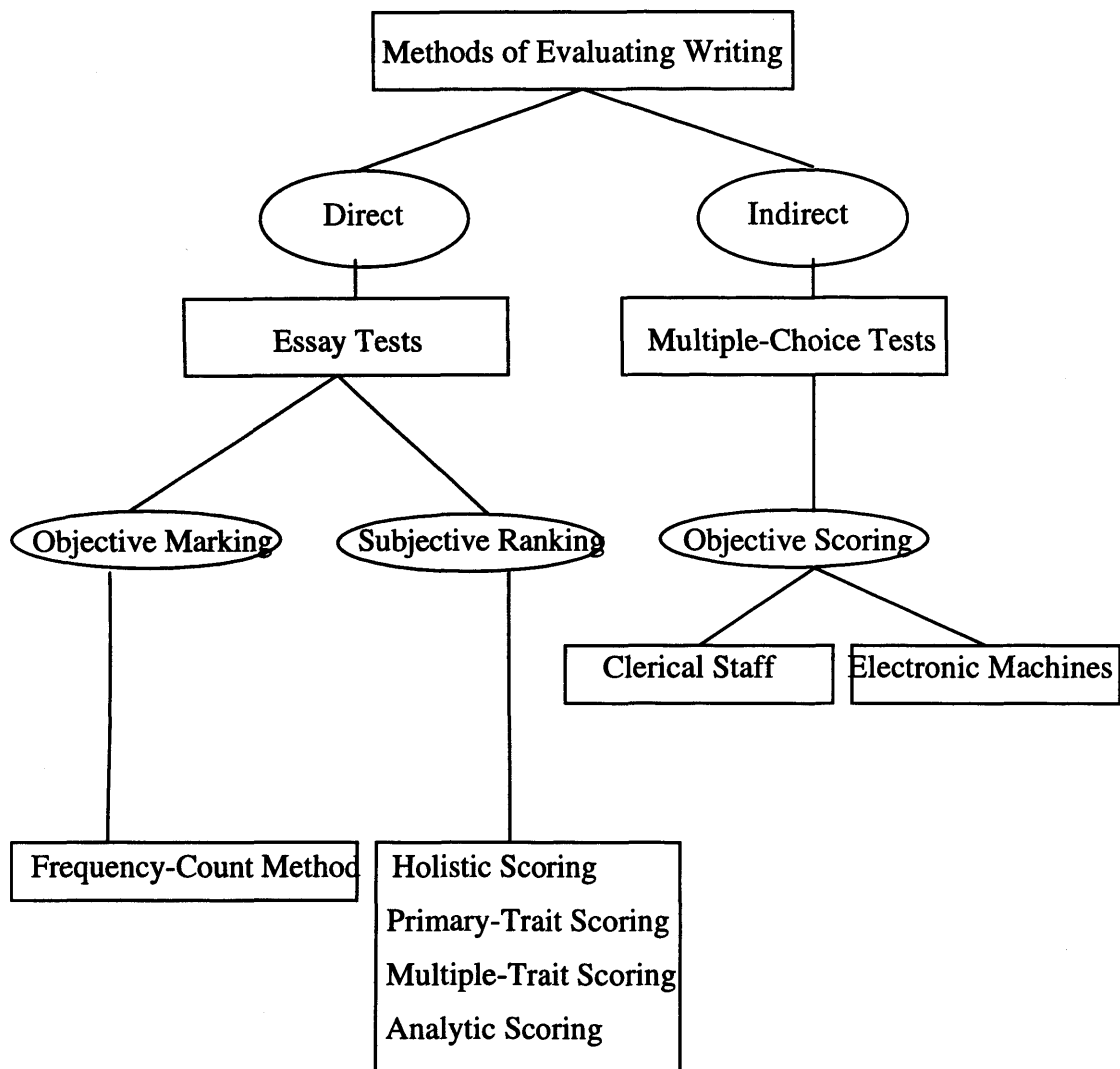
2.13 Frequency-count marking

As Figure 2.4 shows, essay tests, presumably, can be scored objectively by counting

the frequency of specific features, or applying certain formulas and indices. In these

methods scorers 'tally or enumerate certain elements in the composition, such as: the

number or type of words, clauses, the T-units, cohesive devices, misspelled words, misplaced commas, or sentence errors' (Jacobs et al., 1981, p. 29).

La brant (1933) can be considered the pioneer of this approach. She focused on some objective measures such as sentence length, clause length, and subordination ratio. Among these, the subordination ratio proved to be a significant indicator of the student's maturity in writing. Similarly, Hunt (1965) developed five objective measures to evaluate writing skill. Among these five factors T-unit (terminable unit) proved to be the best index of writing maturity which was later confirmed by Cameen's (1979) findings. Other researchers (e.g. Crowhurst, 1980; Odell, 1989; Robb et al., 1986, Wilkinson, 1989), however, refuted Hunt's measures, arguing that syntactic complexity depends on the writing task. To render a bad situation worse, Larsen-Freeman (1978) reported that neither the average number of words per T-unit, the average number of T-unit per sentence, nor the average number of words per error-free T-units were significant.

Many research studies have been carried out having the concept of error as their focal point. Perkins (1980) asserts that objective measures without taking the role of errors into account 'are of no use in discriminating among holistic evaluation at one advanced level of proficiency' (p. 64). In this connection Arthur (1979) found that there is a relationship between the teacher's ranking of the essay and the frequency of spelling and grammar errors. Supporting Hunt, two other investigators, Flahive and Snow (1980) developed two new measures: an index of complexity, loosely based on the work of Endicot (1973) and the second one, the error per T-unit ratio which revealed that Hunt's measures were quite accurate. Larsen-Freeman (1978) found that the error-free T-unit scores significantly differentiate between ESL writing levels. The

element of committed errors is so appealing that a number of scholars (e.g. Humburg, 1984; Magnan, 1985; Oller 1979) each tried to formulate the role of the error element in evaluating writing. Madsen(1983) labels this objective technique as a 'points off' method. In his terms, "Students begin with 100 points or an A grade. Then, they lose points or fractions of a grade for errors." (p. 120)

The relationship of cohesive devices and the maturity of writing skill is another widely-used measure of students' writing ability. Some teachers, like Donley (1978), believe that superior writers can handle cohesive devices more appropriately and, therefore, achieve a more coherent text. The most outstanding experiment done in this area is that of Evola, Mamer, and Lantz (1980). They investigated the relationship between cohesive devices used in a piece of writing and the objective score as well as the subjective rating given to it. It was found that there is a significant but weak correlation between the correct usage of conjunctions, pronouns, and articles and the level of writing proficiency. They admit that cohesive devices are not highly reliable measures because they provide only "minimal indicators of overall language proficiency...(and) cannot be expected to reflect ... communicative ability..." (Evola, Mamer, & Lantz, 1980, p. 191). Farzanehnezhad (1992) in a more recent study propounded two new measures called Measure of Cohesion (MC) and Maturity Index (MI). The MC can be computed by adding the number of cohesive devices and dividing the sum by the number of the T-units used in a composition. The MI is calculated through the following formula: MI = mean T-unit length + clause/ T-unit ratio + MC.

Examining the type of words selected by a writer as the building blocks of a written text, in order to find an objective measure of EFL/ESL writing proficiency has not

been ruled out. Nemati (1993) found a significant negative correlation between the composition scores and the ratio of monosyllabic words in those compositions. It was further shown that a negative correlation exists between the proportion of uses of the auxiliary verb 'TO BE' and the writer's mastery over the writing skill. Unlike "to be", an increase in the ratio of other verbs to the total number of words indicated an increase in the writers language proficiency.

Although Jacobs et al. (1981) admire the objectivity and high reliability of the Frequency-Count method, they sharply criticise its dubious validity because in this method of assessment the communicative effect of the composition is replaced by the number or kind of elements which de-emphasise communicative aspects. Once again, it should be pointed out that these methods, or different techniques within a method, are considered holistic assessment as they require the test takers to create a continuous written discourse (cf. objective multiple-choice tests), yet they are totally different from holistic methods which rely on the subjective scores, mainly based on the rater's impression of the composition. Accepting Hamp-Lyons' characteristics listed earlier, it would be hardly possible to put this method in the pigeon-hole of direct evaluation of writing since there is no subjective rating process involved.

2.14 Holistic methods

There are a few methods of evaluating writing, generically called holistic methods, which are in fact different generations of one single approach. All these versions have two common features. First, they all test writing ability directly by making the test

takers physically write, which is totally different from objective multiple-choice tests. Second, in all these methods the written essays are marked by employing one or more raters relying on their subjective judgements, which differs from the objective and mechanically quantifiable procedure used in frequency-count methods. Confusingly, in some sources holistic methods are considered synonymous to holistic scoring, but here holistic methods is a general term which is used to refer to a bunch of scoring approaches including: *'Holistic scoring'*, *'Primary-trait Scoring'*, *'Multiple-trait scoring'*, and , *'Analytic scoring'*. Each of these is examined in more detail below.

2.14.1 Holistic scoring/Impression marking

The traditional writing assessment method, also known as Integrative, Global, Rated, and Multiple Marking, is the marker's evaluation of the composition as a whole discourse without taking the individual aspects of the writing into account. Cooper (1977) describes a three-phase procedure for holistic scoring in which the reader either matches a piece of writing with another one, which serves as a benchmark, or scores it for the prominence of certain features or assigns it a letter or number grade. Weir (1990) maintains that separating the discrete features of a composition is a violation of the notion of impression marking . In order to avoid such a violation sometimes the term *focused holistic scoring* is used when the judgement is based on a given guideline or rubric, which is nowadays the common practice of most educational institutes.

Holistic scoring has been practiced as long as the history of education and can be regarded as the precursor of other scoring methods. Madsen (1983) praises it as "one

of the best ways to evaluate the complex communicative act of writing" (p.121). It is popular because of its inherent merits. Munro (1991) lists four of them:

1. It is reasonably fast, and therefore inexpensive.

2. It produces a reliable ranking of the writing products.

3. It meets the technical requirements of norm referenced tests.

4. Both discourse-producing and text-producing skills are included on a single scale.

For example, two popular and renowned language tests use this method for their writing sections: **TOEFL** and **IELTS** . The **Test of Written English (TWE)** is the writing section of TOEFL, a language test widely used in the United States and many other countries for the admission of students to colleges and universities. It is marked with a holistic approach that ranks the applicants on a 6-band scale ranging from 1 signifying a complete incompetence to 6 which means a complete rhetorical and syntactic competence. After a series of studies intended to monitor the reliability of scores (Carlson et al. 1985), this test was introduced in 1986. These studies continued after its introduction and are published in several volumes of TOEFL reports. As a result of such investigations the test was revised in 1990 and is going to be changed in the revised version of TOEFL known as TOEFL 2000. Probably the most revolutionary change in this test would be the use of word processing equipment.

IELTS, another holistic test mainly used in the UK and Commonwealth, uses a 9-point scale to rank the applicants' writing ability.

The major difference between IELTS and TOEFL is that IELTS, uses different modules to suit the testee's area of speciality. As a result it may be claimed that it

probably has a higher content validity. However research (Geranpayeh, 1994) shows that there is a correlation between the scores obtained by the two tests.

Holistic scoring, like any other method, has been subject to all types of criticism. The problem of low inter-rater and intra-rater reliability is in fact the target of most criticisms. Nonetheless, holistic scoring will be discussed here because these are common issues and other scoring methods in the domain of holistic assessment that make use of human judgements suffer as well. The topics of reliability and validity will be dealt with thoroughly in a separate section.

The major shortcoming of this method lies in its nature. Papers are judged against each other in a ranking system without any reference to an independent criterion, as a result the given scores are not fully meaningful. Even the raters themselves do not perfectly share the notions behind each score. A composition receiving a high grade is just a relatively good one and an essay marked as poor is a relatively poor one and can be regarded as average or even excellent when compared to another bunch of essays, which is the inherent feature of norm-referenced tests. When the scorers do not take any specific feature into account, then, the writer's weaknesses will remain unknown and no diagnostic information can be provided for the people interested in the test results, including the students. Holistic scoring, then, is of little value with regard to its educational washback on teaching and such a method, in Hamp-Lyons' words (1991) "permits a disjunction between teaching and assessment, a disjunction we have suffered under for all too long".

Viewing these objections, it becomes evident that they are all aimed at the traditional version of holistic scoring: general impression marking. In order to eradicate the shortcomings and improve the quality of assessment new alternatives were introduced

The basic difference is that these new versions of holistic scoring try to use certain criteria to assign marks to essays. In other words, this is a shift from norm-referenced tests to criterion-referenced tests.

2.14.2 Analytic scoring

In order to alleviate the major problematic area of impression marking, the lack of established criteria to mark the compositions against them, analytical scoring was introduced into the field of evaluating writing. It was adopted meteorically and practised almost everywhere during the late 60s and 70s and even the early 80s. According to Weir (1990) 'analytical marking refers to a method whereby each separate criterion in the mark scheme is awarded a separate mark and the final mark is a composition of these individual estimates' (p.63). So, a teacher, following this analytical approach, identifies several components of writing such as 'organisation, wording or ideas' (Chastain, 1988, p.261), 'style, grammar, and mechanics, etc' (Brown and Bailey, 1984, p.22). These components are not necessarily of the same weightings.

One of the earliest studies concerning the effectiveness of analytical scoring was conducted by Hartog et al. (1936). They realised that variation between markers was, to some extent, reduced by the analytic method. Following them, Cast (1939) found this method slightly superior to the holistic method particularly with a single marker. Diederich, French and Carlton (1961) conducted a widely-cited research study and came up with five factors: ideas, form, flavour, mechanics and wording. They had applied the rating scale devised in three high schools for a year (Diederich, 1964),

before the results of factor analysis reduced the five factors down to two: factors of 'general merit' and 'mechanics'. Incorporating the comments given by teachers they granted ideas and organisation double weighting and added a new category for handwriting. Mullen (1977) proposed four facets contributing to overall writing proficiency, namely (1) control over English structure, (2) organisation, (3) appropriateness of the vocabulary, (4) quantity of writing. Similarly, Jacobs et al. (1981) in their highly acclaimed study developed a weighted composition profile. This profile includes five categories: content (30%), Organisation (20%),vocabulary(20%), language use (25%), and mechanics(5%). Brown and Bailey (1984, p.26) are convinced that "they adequately demonstrated the face, content, concurrent, predictive, and construct validity of their profile" and Hamp-Lyons (1990, P.78) considered it the best-known scoring procedure of the time.

Reid and O'Brien (1984) introduced a similar analytic scoring guideline based on two major categories: (1) organisation, length and content (2)grammar, vocabulary and spelling. Madsen (1983) introduces a simple profile which is almost the same as other profiles but with slightly different weightings.

Analytic marking was originally initiated to reduce the subjectivity of holistic scoring and did so by providing descriptive information about the students' performance. It can give the teacher an image of the student's areas of strength and weakness and thus it is of more diagnostic and pedagogical value. Besides, the immediate outcome of reducing the element of subjectivity in a test is achieving a relatively higher reliability level and that is one of the advantages associated with this method.

Although Brooks (1986) claims that all the qualities incorporated in the past analytical marking schemes were extremely elusive, nebulous, and inaccessible to assessment, it

seems that this allegation is hyperbolic. To be realistic, one can diagnose three basic

flaws in analytic scoring. Firstly, it is difficult, if not impossible, to weigh any single

feature being analysed, particularly vaguer and more abstract ones such as 'flavour'.

Since there is no well-grounded rationale, different scholars propose different

weighting scales. Secondly, analytic marking 'may overlook certain general qualities

that characterise the essay as a whole' (Cast, 1939, p.264). Cast, however, admits the

superiority of this method over others, but did not advocate the exclusive use of it.

Last but not least, is the test administration. Compared with general impression

marking, analytic scoring is up to four times more time-consuming (Munro, 1991), and

therefore, especially when the evaluation of large numbers of test-takers is involved, it

is more costly to administer.

2.14.3 Primary Trait Scoring (PTS)

In an attempt to design a perfect version of holistic scoring without the above

shortcomings resulting from being norm-referenced and being solely impressionistically

scored, the National Assessment of Educational Progress (NAEP) developed the

primary-trait scoring (PTS) method. This method, unlike holistic scoring, instead of

imposing a normal distribution on the scores adopts a percent acceptable system. The

underlying assumption of PTS is that writing does not take place in a vacuum. A writer

writes to a certain audience to fulfil a certain need and any evaluation should be based

on the accomplishment of such requirements. In Brown and Bailey's words (1984) this

method is "rhetorically and situationally specific". It, therefore, assumes a scoring

criterion in terms of tightly specified text features in a given context. This includes all

salient aspects of the required task, its expected quality and primary characteristics, the

explanation of a scoring scale and even some sample-papers. According to Chastain (1988), a teacher following a primary-trait approach evaluates characteristics unique to the particular audience and purpose of the writing such as persuasiveness, clarity, and so on. Two complete examples of such tasks can be found in Lloyd-Jones (1977), the founder of this approach..

Basically, two major problems are associated with this method of scoring compositions. Firstly, it is, in practice, impossible to make the reader ignore other aspects of the written essay except those regarded as primary traits of that particular task. In fact, reading comprehension is a complicated phenomenon in which the reader is exposed to virtually unlimited interwoven and interacting facets present in the passage being read (Hamp-Lyons and Henning, 1991). To be realistic, all this means going back to the original starting point, impression marking. The simple reason for this is that the PTS reader is, in spite of given guidelines, behaving in a way similar to what scorers do in general impression marking, in general, and focused holistic marking, in particular. Secondly, each task, even each topic, should be carefully designed and separately developed through a demanding process and all these elaborate guidelines should be rewritten again in case of another test administration. Thus, it is expensive, time-consuming and, therefore, unsuitable for large-scale research purposes and experimental studies (Grabe and Kaplan, 1996).

2.14.4 Multiple Trait Scoring

Like other innovations, the shortcomings of PTS gave impetus to the development of a supposedly better approach. Hamp-Lyons (1991) labels this method Multiple-Trait

Scoring. Unlike PTS, markers using multiple-trait scoring do not limit themselves to a primary trait within a dimension, rather they consider multiple dimensions defined in terms of certain text-producing skills and subskills and measure them separately. This allows 'the rater to consider a number of aspects of the essay (Grabe and Kaplan, 1996, p.405). In this method, again, the writing tasks as well as subskills are defined in advance. The rubrics may differ from one to another. One may be highly specific while another may be designed so generally that it might require a holistic score. The fulfilment of these component subskills will be rated against a refined scale of 1-4. It is obvious that this type of test is a criterion referenced one.

Multiple trait scoring has its own advantages. First of all, it is much less costly than PTS for this scoring rubric works with a number of essay prompts as long as they do not violate the assumptions and parameters. Another positive characteristic of multiple trait scoring is its superiority over primary-trait scoring in its diagnostic capabilities (for essay writers rely on multiple traits instead of focusing on a single, salient primary trait) and therefore it is suitable for student writing assessment.

The multiple-trait scores, ironically, deprive the readers from their valuable holistic views. Mutilating a composition into subskills involved in the creation of a certain piece of discourse is, in a sense, a degradation towards the old analytical scoring, if not the traditional multiple-choice tests. Once again it seems that, as Gestalt theory would hold, a whole is more than the sum of its parts. A subject might perform brilliantly on scoring rubrics without writing a readable, effective essay. Besides, the combination of different performances under different rubrics is either not clear or is tackled simplistically. For example, it is assumed that doing three rubrics will result in a better piece of writing than doing two which is not, of course, always true, because facets of

writing weave together differently and with different weightings. Above all, defining and designing refined rubrics and using them appropriately is no easy task which can otherwise cause serious validity and reliability problems. Studies by Canale et al. (1988), Cumming (1990a), Hamp-Lyons and Henning (1991) are all labelled as examples of such dubious experiments by Grabe and Kaplan(1996).

It should be pointed out that in the literature a method known as performance or performative assessment is described as a method which employs a carefully selected topic to elicit the desired type of writing which requires certain text-producing skills (e.g. classification) and then evaluates it focusing on the fulfilment of this task. However, it seems that performance assessment, a term coined by Faigley et al. (1985), is, in fact, another version of multiple-trait scoring or at least 'most likely to be carried out using a carefully constructed multi-trait scoring rubric' (Grabe and Kaplan, 1996, p. 406). This conclusion is also confirmed by Allaei and Connor (1991), who assert that 'performative assessment tasks separate dimensions required for certain text-producing abilities and often specify an audience' (p. 228). This definition fits multi-trait scoring as well.

2.14.5 The future: Portfolios

There is a growing body of evidence in the literature indicating that portfolio assessment is gaining enormous popularity in academic situations. A portfolio is an album of pre-defined written texts produced by a student during a certain time period. A good example of this method is being practised in the State University of New York where students are required to assemble a portfolio of their course writing. Each

portfolio consists of three revised essays (a narrative, descriptive, or expressive essay; an academic essay; and an essay criticising another essay; one in-class essay, and the cover sheet describing the purpose and process of writing of each essay (Grabe and Kaplan, 1996).

Although there are an insufficient number of reported research studies evaluating portfolio approaches, briefly speaking, a portfolio programme is likely to be considered successful in the sense that it makes students write more and, hopefully, write better when marked attention is paid to the re-drafting and revising processes after feedback. They have to write for a specified audience and therefore, produce a more authentic piece of writing. As the students are assessed through a collection of writings produced over an extended period they will no longer overestimate the final examination, rather they will take all writing tasks more seriously. With regard to validity, the portfolio system meets the assumption that writing is not a one-shot process; instead, almost all writers have the opportunity to revise their writings. Besides, in this method a writer is viewed in different situations and for this reason better assessment will be carried out.

There are, however, a few problems attributed to portfolio assessment. Teachers may complain about being asked to work more. Some think that students are given wider opportunities to collaborate or cheat. Therefore this approach seems too lenient for academic evaluation purposes since assessors of writing need to know the precise conditions under which written tasks are produced and that all students have equal opportunities in comparable conditions. There is also the dilemma of combining the separate grades to assign a single score to a student which may involve problems of

weighting. And among the serious questions which still remain to be answered, there are unsolved problems with regard to reliability of portfolio assessment.

Portfolio approaches for evaluating writing have already been introduced into several Iranian universities and are gaining more popularity. It seems that the problems associated with this approach (i.e. demanding teachers to work more, and the issue of plagiarism) will be exacerbated in Iran because of its particular cultural and economical situation.

In this chapter, several fundamental issues in the field of writing and testing and testing writing which are somehow related to this research study have been reviewed. There is one more issue which is central to this study and that is the mode of discourse. As it is a relatively different issue, the next chapter, chapter three, is dedicated to its historical perspective, a review of its different classifications and other related discussions.

CHAPTER THREE

THE TAXONOMY OF MODES OF DISCOURSE

"There is no uniformity in the approaches to the classification of mode in the research studies to be discussed, so we will use the traditional categories—narration, description, exposition, and argument ... "

Ruth and Murphy (1988, p.78)

3.1 Definitions and classifications

The term *"modes"* of discourse is a relatively recent substitute for the more traditional term *"forms"* as was used by Cairns (1902) and Giovannini (1943) many years ago. In literature, however, there has been a tendency to use *"genre"* and *"type"*, and *"kind"*, instead. Ruth and Murphy (1988, p. 88) also call it *"method of development"*. Yet, mode is conceptually distinct from genre. Defining genre, Swales (1990, p. 58) writes:

> A genre comprises a class of communicative events, the members of which share some set of communicative purposes. These purposes are recognized by the expert members of the parent discourse community, and thereby constitute the rationale for the genre. This rationale shapes the schematic structure of the discourse and influences and constrains choice of content and style . . . in addition to purpose, exemplars of a genre exhibit various patterns of similarity in terms of structure, style, content and intended audience.

In other words, they are "texts with common linguistic and structural configurations" (Harris, 1993, p.127). Culture and social purpose are also crucial in evolving a category of genre (Bhatia, 1993; Cope and Kalantzis, 1993; Littlefair, 1991). Examples of genres are: recounts, sermons, conversations, reports, speeches, poems, letters, and novels. It is theoretically possible to have several modes in one given genre or one certain genre, like poetry, written in two different modes of say narration and

argumentation. There can also be numerous kinds of genres but there are very limited classes of modes i.e. between three to six according to different classifications. Therefore, as Kinneavy (1971, p.36) maintains, modes are "more manageable". In other words, theoretically all modes of discourse can be found in a given type of genre. Richards, Platt and Platt (1992, p.338) defines *mode of discourse* as "what part the language is playing in this particular situation, for example, in what way the language is organized to convey the meaning". D'Angelo (1976, p.115) goes beyond the organisational differences and believes that: "each form is assumed to have its own function, its own subject matter, its own organizational patterns and its own language...".

Distinctions between different categories of writing have been made for a long time. The first and the most famous classification which is cited in almost all writing books, and which has been the basis of the present research, is a four-category one, although there are a couple of other alternatives. Before the 1970s, there had been some sort of orthodoxy among language teachers and textbook writers to teach writing using only the four major rhetorical categories: *narration, description, exposition* and *argument*. Brooks and Warren (1952) write:

> We can see with only a moment of reflection that these four types of intention [to inform the reader, to change him, to convey to him the quality of experience, to tell him about an event] correspond to the four basic kinds of discourse: EXPOSITION, ARGUMENT, DESCRIPTION, and NARRATION. Exposition embodies the wish to inform the reader, argument the wish to make the reader change his mind or attitude, description the wish to make the reader perceive something, narration the wish to make the reader grasp the movement of an event. (p.30)

They further point out that these divisions of discourse are arbitrary but in fact they correspond to different needs that the writer intends to fulfil. Harris (1993, p.16), too, admits that this fourfold division is "the most-long established" attempt to classify writing. Ruth and Murphy (1988, p.88) also agree that these traditional forms of discourse "have dominated the organization of writing curricula and testing even to this day."

Britton et al. (1975), somewhat sarcastically stress that these four modes of discourse have become unassailable and have been accepted by almost all researchers and authors and have become part of any discussion about written composition. Kinneavy (1971) also admits that the classification of discourse modes into **narration, exposition, argumentation,** and **description** still prevails today.

In an attempt to ascertain the origin of this type of classification Britton et al. (1975) account for its history in the following way:

> The four categories have evolved from the period when rhetoric broadened its realm of inquiry from persuasive oratory to all forms of written discourse; their point of origin in this form is to be found in Campbell, who wrote in 1776, "All the ends of speaking are reducible to four", and went to identify these ends as: to enlighten the understanding, to please the imagination, to move the passions, to influence the will. By Bain's time (1866) the categories have a more familiar ring. Without supporting rationale he announced "five leading kinds of composition, namely, Description, Narration, Exposition, Oratory and Poetry." In our own day Grierson, in the direct line of descent of this predominantly Scottish tradition of rhetoric, refers to the familiar four and asserts that they are based on "a division which arises out of the fundamental division of experience" – a bold but eccentric psychological criterion. (p. 4)

It can be concluded, as D'Angelo (1976) did conclude, that the first acceptable classification of the modes was actually done by Bain, (1866). Kinneavy (1971) believes that this classification is grounded in the nature of the reality they deal with but not the nature of the language as there is always a relationship between a written text and the outside realities. This relationship, according to Kinneavy (1971, p.35) can be investigated by a question like "What is this text about?" which may receive such responses as: "It is a story about the wife of Napoleon's general;" or , "It's a description of the topography of Northern New Mexico." Considering this sort of question-answer investigation of text, one would come up with four categories namely: a narrative, a series of classifications, a criticism or evaluation, and a description. To Kinneavy, each of the modes has its own *peculiar logic, organisational patterns* and to some extent, *stylistic characteristics.*

Despite these peculiarities, Kinneavy (1971) maintains that these modes do overlap and that in fact it is impossible to have a discourse written purely in one single mode. Park (1986) bemoans the situation:

> The fact is that the different discourse modes can easily be blended in a piece of writing because the single essay is usually composed of smaller units of different discourse modes. Therefore, the results of studies in discourse mode effect on writing performance should be interpreted with caution. (p.22)

This tricky point has been touched upon by many authors. Andrews (1995), in a realistic evaluation of the situation, believes that despite the overlapping among the modes these terms are commonly used so instead of denying them it is best to define

them. Brooks and Warren (1952, p.30) agree with the idea that "none of these kinds of discourse often appears in a pure form", however, they assert, "there will always be a *main intention*, a fundamental wish. Borrowing the term from Morris (1946), Kinneavy (1971. p.37) also reiterates that by a mode we mean the "dominant" mode not a pure mode.

3.2 The fall of modes

Although it is widely accepted and commonly practised, the classification of discourse mode into the major groups of narration, explanation, argumentation and description has not been left without criticism especially in the 1970s and afterwards. For instance, Britton et al. (1975), criticising this type of classification, refer to some of the drawbacks of the classification:

■ This is a product-oriented distinction and little attention is paid towards the process of writing and therefore of little help for the process of emerging mature writers from young writers.

■ this taxonomy is supposedly based on the intention of a given piece of writing and when it comes to narrative, unlike exposition or persuasion, different intentions can be attributed to narratives such as a fictional story or a factual report of events.

■ These four categories are not of equal status. Narration and description are different from exposition and argument in that the former two types can be part of the latter two. In other words, two different types of discourse can be found in one piece of writing while a broad type of discourse, like exposition, may contain a variety of functions.

Andrews (1995) also reiterates that a lack of distinction between narrative and argument as modes and functions is misleading.

Harris (1993, p.17) finds three problems with this traditional classification alongside "a general sense of dissatisfaction". First, this traditional practice, he argues, belongs to the time when education was there as a luxury largely for an elite minority and now that a wider range of pupils receive education then writing curriculum should be revised to cater for their needs which is different from that of the older times. The second reason is that the nature of the writing task has a bearing on development and achievement. Therefore in order to describe development in writing, there is a need for a more sophisticated measure which, he regrets, has not been applied to "an extensive study of writing development" (ibid, p.17). And finally, like achievement, assessment, especially when based on a criterion-referenced assessment, depends on the kind of task, too. For this link between the task undertaken and the test score, it is necessary to describe the demands of different types of writing being used as the criteria to assess the learners' writing abilities.

Connors (1981) summarises the rise and fall of the modes in this way:

> "The history of the modes is an instructive one; from the time of their popularization in American rhetoric textbooks during the late nineteenth century, through the absolute dominance they had in writing classrooms during the period 1895-1930, and into the 1950's when they were finally superseded by other systems, the modes of discourse both influenced and reflected many of the important changes our discipline has seen in the century." (p.444)

3.3 Categorical vs. relational systems

This traditional type of classification of modes looks at writing as a *finished product* rather than a *process* and this is the focus of most criticisms and the reason for attempts to propose other systems. Ruth and Murphy (1988, p.88) call the

classifications that accept the former point of view, **categorical systems** and the those who prefer the latter, **relational systems**. Besides the classic taxonomy of modes of discourse, there are other categorical systems that view the issue from a product approach, which is the approach chosen for the present research, too.

3.3.1 Other categorical classification

Another categorical classification of modes, which is not as famous as that of Bain's (1866) or Britton's (1975), was given by Rockas (1964). In an attempt to develop a thorough taxonomy of kinds of discourse, he combined literary modes to Bain's classical system. In this system modes are divided into concrete and abstract. Description, narration, drama, and reverie are concrete and definition, process, dialogue, and persuasion are abstract modes. He further classifies these modes by means of procedure. Description and definition are static, narration and process are temporal, drama and dialogue are mimetic, and finally reverie and persuasion are considered to be mental modes. This amalgamated classification has not been received warmly by researchers.

D'Angelo (1980), in his classification, refers to aims and purposes of writing. He distinguishes four types of discourse: informative discourse, persuasive discourse, literary discourse, and expressive discourse. Using the term mode as synonymous with the term topic, he distinguishes more than one topic in most discourse. Generally there are two types of topics in D'Angelo's classification: static and progressive. Identification, description, analysis are examples of static topics and narration, process and cause and effect are instances of progressive topics.

Undoubtedly, the most influential recent development in classifying modes of

discourse in north America, has been suggested by Kinneavy (1971) in form of four

elements of a triangle. He explains that these four elements determine the *aim* of

discourse. For example, when the focus is on the reader-audience and the aim is to

persuade, the final product is persuasive discourse. Kineavy does not accept argument

and exposition as modes since they are, in his view, aims of discourse. These two

modes are replaced by *evaluation* and *classification* as new modes to stand along

with *narration* and *description*.

Fig. 3.1 Kinneavy's communication triangle

```
Encoder     \‾‾‾‾‾‾‾‾‾‾‾‾‾/     Decoder
Writer       \   Signal   /      Reader
Speaker       \  Message  /       Audience
               \  Test   /
                \       /
                 \     /
                  \   /
                   \ /
                    V
                 Reality
                 Subject
```
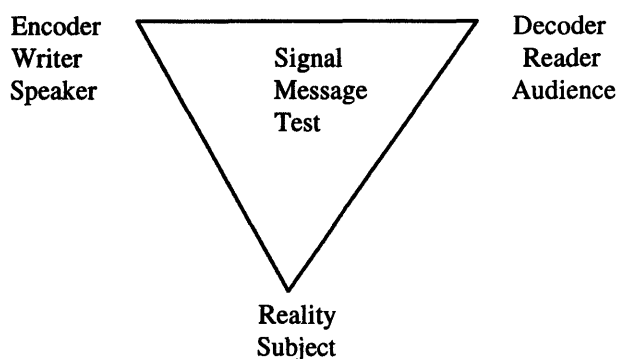
(From Ruth and Murphy, 1988)

3.3.2 Relational Systems

To find a remedy for the problems found in other taxonomies, Britton et al (1975) ,

suggest a multi-dimensional model which could characterise all mature written

utterances and the process of their development. They believe that the notion of genres

was "both too broadly diffuse and too exclusive to form a starting point" and instead

they try to use the findings of psychology and linguistics in child development.

They, finally, distinguish two sets of categories: those for the sense of audience and

those for the function.

The sense of audience "is revealed by the manner in which the *writer* expresses a

relationship with the reader in respect to his (the writer's) *undertaking*" (ibid, p. 65).

The main divisions include:

1- self,

2- teacher,

3- wider audience (known),

4- unknown audience,

5- additional categories

The schematic categorisation of the audience is given below:

Figure 3.2 : Audience categories adapted from Britton et al. (1975)

```
   1              2              3                    4                  5
 ┌──────┐   ┌─────────┐   ┌───────────────┐   ┌─────────────┐   ┌──────────────┐
 │ SELF │   │ TEACHER │   │ WIDER AUDIENCE│   │  UNKNOWN    │   │ ADDITIONAL   │
 └──────┘   └─────────┘   │   (KNOWN)     │   │  AUDIENCE   │   │ CATEGORIES   │
                          └───────────────┘   └─────────────┘   └──────────────┘

              ┌────────────────┐          ┌──────────────┐
              │   Expert to    │          │ writer to his│
              │  known laymen  │          │ readers (or his│
              └────────────────┘          │ public)      │
 ┌────────────┐                           └──────────────┘
 │ Child      │  ┌────────────────┐
 │ or adolescent│ │ Child (adolescent)│
 │ to self    │  │ to peer group  │
 └────────────┘  └────────────────┘

              ┌────────────────┐          ┌──────────────┐
              │ Group member   │          │ Virtual named│
              │ to working group│         │ audience     │
              └────────────────┘          └──────────────┘

                                          ┌──────────────┐
                                          │ No discernible│
                                          │ audience     │
                                          └──────────────┘

 ┌────────────────┐  ┌──────────────┐  ┌────────────────────┐  ┌────────────┐
 │ Child (adolescent)│ │ Pupil to teacher,│ │ Pupil to teacher,    │ │ Pupil to   │
 │ to trusted adult │ │ general       │ │ particular relationship│ │ examiner   │
 └────────────────┘  └──────────────┘  └────────────────────┘  └────────────┘
```

There is no doubt that this classification and its focus on the concept of audience is innovative and remarkable; however, it seems that there are a couple of problems with this classification. Britton et al. (1975) do not make clear the differences between two different types of writings, or genres, written by the same person for the same audience. In other words, other factors such as the nature of the task can influence the appropriateness of register in a piece of writing, too. Besides, in this approach an audience is supposed to be set for the learner to write to, rather than writing to teacher as was practised in traditional approaches. This is not a real but just a "pseudo-audience" which can be interpreted as replacing "one artificiality with another" (Harris, 1993, p.22).

With regard to the principal functions of the written utterances Britton et al. (1975) view the degree of the reader's involvement as a key point which leads into the distinction between participant and spectator. As a result they distinguish three major functions (p.81):

Figure 3.3 Functions of writing (From Britton et al. 1975)

Participant role------------------------------------- .-------------------------Spectator role

TRANSACTIONAL_____EXPRESSIVE_____POETIC
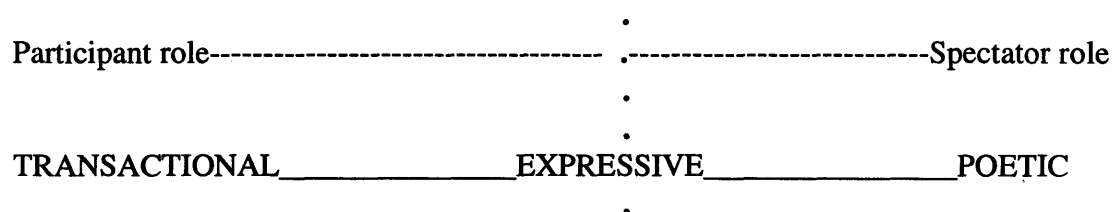
The transactional category is broken down into **informative** and **conative** sub-categories. Conative itself is further divided into *regulative* (where compliance is assumed) and *persuasive*.

This seven-category classification is in fact a development of James Moffett's (1968) four-category classification in which he suggests that writers, depending on the degree of abstraction, perform four functions: *recording*, *reporting*, *generalising* and *theorising*. In this classification (which has its roots in the Piagetian theory of abstraction), recording corresponds to drama, reporting to narrative, generalising to exposition and theorising to logical argumentation. In temporal terms, it is a movement from past to present and from present to future. Drama says what happened in the past, narrative tells what is happening, exposition explains what happens, and finally, argument predicts what may happen in the future. Andrews (1995, p.67) prefers Moffett's own work as he finds it "the most extensive account of relationships between narrative and argument as modes of discourse". In other words, recording resembles closely the structure of external reality while theorising tends to be the manifestation of man's mind. Their classification of modes of discourse based on the function they fulfil can be schematically represented as below:

Figure 3.4: Function categories adapted from Britton et al. (1975)

FUNCTION CATEGORIES

```
                        ┌─────────────────┬──────────┬──────────────┐
                        │                 │          │              │
            ┌───────────────┐  ┌──────────────┐  ┌─────────┐  ┌──────────────┐
            │ TRANSACTIONAL │  │ EXPRESSIVE   │  │ POETIC  │  │ ADDITIONAL   │
            │               │  │              │  │         │  │ CATEGORIES   │
            └───────────────┘  └──────────────┘  └─────────┘  └──────────────┘
                 │                                    │              │
         ┌───────┴────────┐              ┌────────────┴─────────┐    │
  ┌─────────────┐  ┌────────────┐  ┌───────────────────┐  ┌──────────────────┐
  │ INFORMATIVE │  │ CONATIVE   │  │ IMMATURECATEGORIES │  │ SPECIAL CATEGORIES│
  └─────────────┘  └────────────┘  └───────────────────┘  └──────────────────┘
       │             │                                            │
       │       ┌─────┴──────────┐
       │  ┌──────────────┐  ┌──────────────┐
       │  │ REGULATIVE   │  │ PERSUASIVE   │
       │  └──────────────┘  └──────────────┘
       │                              
       │         ┌──────────────────┬──────────────────┐
       │  ┌───────────────────┐  ┌───────────────┐  ┌─────────────┐
       │  │ PSUEDO-INFORMATIVE│  │ PSUEDO-CONATIVE│  │ DUMMY RUN   │
       │  └───────────────────┘  └───────────────┘  └─────────────┘
       │
  ┌────┴────┬──────────┬──────────────┬─────────────┐
┌────────┐ ┌────────┐        ┌────────────┐   ┌──────────────┐
│ RECORD │ │ REPORT │        │ ANALOGIC   │   │ TAUTOLOGIC   │
└────────┘ └────────┘        └────────────┘   └──────────────┘
```

| GENERALIZED NARRATIVE OR DESCRIPTIVE INFORMATION | ANALOGIC, LOW LEVEL OF GENERALIZATION | ANALOGIC-TAUTOLOGIC (SPECULATIVE) |
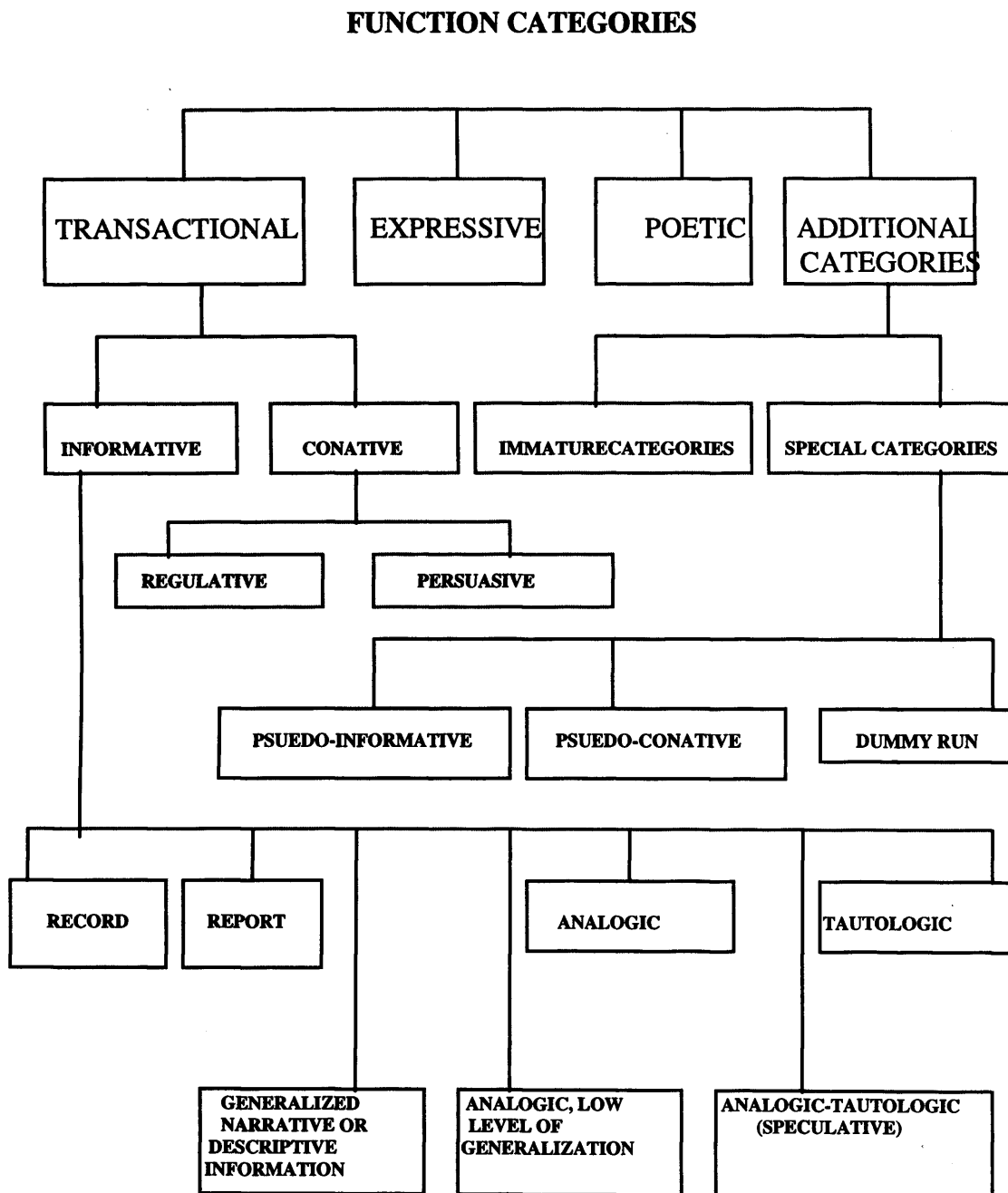
In short the model proposed by Britton is based on two notions of **participant** and **spectator.** The former, in a writing activity, is manifested as self-expression and the latter refers to the social effectiveness of the writing. Criticising this two-part model of discourse classification, Lloyd-Jones (1977) held that:

> "It is an excellent model for directing observations of the gradual socialization of children, but it tends to take for granted the demands of the subject, of information processing, which is important to responsible adults and thus to the schools.... by limiting the observations about the writing to the participants in communication, the encoder and the decoder, the two-part division diminishes our sense of how the external reality influences our reasons for writing and how the code itself works."
>
> (p. 38)

Considering all these contributions, Harris (1993) summarises the prevailing present belief that the final form of a text is the outcome of interactions between two parameters of *purpose* and *readership*. In other words, depending on the intentions of the writer or the expectations of the reader and the relationship between the writer and the reader, a type of text is shaped. He also makes a distinction between narrative and non-narrative texts as two major genres.

3.4 The significance of modes of discourse

Categorising language into the four major modes in EAP is significant for it enables us to deal with few types of language instead of dealing with a long list of academic disciplines which use the language and hence it is quite convenient. Regarding the application of these modes in academic writing, Kinneavy (1971) maintains that narration turns into history, description becomes analyses or description, argument

becomes criticism or evaluation, and finally exposition becomes theory or classification. He, therefore, renames the four modes with *narration, classification, description,* and *evaluation.* He further adds that this quartet exists in all scientific disciplines (ibid, p.37), for example, in linguistics there are historical linguistics, descriptive linguistics, theoretical linguistics, and prescriptive linguistics and in literature there are literary history, literary analysis, literary theory and literary criticism.

All these different models proposed by Moffett (1961), Kinneavy (1971), Britton (1975), D'Angelo (1975), and Lloyd-Jones (1977) have been influential on writing assessment in one way or another. For instance, Kinneavy's discourse theory was the underlying theory in Texas assessment, while the National Assessment of Educational Progress (NAEP) was based on that of Lloyd-Jones.

It is quite important to realise that most of the criticism attacking the traditional classification is about the inadequacy of this approach towards certain explanatory aims for instance giving a discourse theory of writing or giving enough attention to the process of writing. Otherwise, this classification, no matter how long-standing it is, is still useful to assess the quality of the final products and the differences which exists among models of writing. And that is why even in the 1990s this classification still survives and can be traced in many textbooks and research papers (e.g. Scott, 1996; Hale et al. 1996, among others).

3.5 The four modes of discourse

So far, the different classifications of discourse modes have been reviewed from a historical perspective. It is better now to examine the four modes which are the basis of the present study in more detail:

3.5.1 Exposition

Hale et al. (1996, p. 13) define exposition as a mode of discourse that is "intended to explain or clarify a subject" which entails " the expression of ideas, opinions, and explanations".

Exposition is the most common kind of discourse in which the main intention of the writer is to inform the reader. This information can be clarifying an idea, analysing a character or a situation, defining a term and giving directions and in general is "intended to inform or promote understanding of a particular piece of knowledge or fact" (Hale, et al, 1996, p.13). It is called exposition for, unlike narration and description, it sets forth a subject directly rather than giving the chronology or qualities of it. Exposition may take the form of expository narration without being ordinary narration or it may take the form of expository or technical description without being real or suggestive description. In both cases the primary purpose of the writer is giving information about a subject. There are different methods of exposition the most common of which are reviewed below:

3.5.1.1 Definition/Identification

This is a simple method of exposition where the writer intends to tell the reader what something is. This "thing" can be a word, phrase or a term which can be either

concrete or abstract. Ironically, the past two sentences are examples of identification themselves.

The intention behind any definition is to attempt to say what something is. It is actually impossible to define terms without a clear knowledge of them and, therefore, the process of definition is the process of giving and gaining knowledge; giving knowledge to the reader and /or clarifying the writer's knowledge of the term being defined.

Any definition comprises of two parts: the *(to-be-defined)* and the *(definer)*. In a good definition the two elements should be both positive or negative (to show some sort of deficiency) and interchangeably equivalent. To do so, the writer should use the words and terms which are already known to the reader (of course, not the *to-be-defined* itself) and refer to the information that the reader has already got. In other words, there must be a common ground between the giver of the definition and the receiver of it. Therefore, the audience has a determining effect on definitions and the writer's choice of words and technical jargons.

The terms being defined are not always simple and straightforward. Instead, it is more likely in academic writing to come up with complicated terms which may require several paragraphs and even a long essay to be defined leading to what is called **extended definition** by Brooks and Warren (1951, p. 91).

There are three parts in an effective definition: first, defining the given term; second, listing the object or concept to which the term belongs and third, deciding on the distinguishing characteristics of that term which makes it different from the other members of its class (Hale et al. p. 14).

3.5.1.2 Illustration

Illustration, as the name suggests, tries to explain a group or class by giving examples. This method is widely used in almost all academic textbooks to provide learners with something more familiar, usually after an abstract discussion or general statement.

Hale et al. (1996) points out that while writers aim to a type, class, or group, they do not usually do so by presenting examples, they, instead, explain the general by presenting the particular.

3.5.1.3 Process

In this expository pattern, the writer writes about how something is or should be done in detailed chronological order. This instruction may also include negative directions in which the writer describes what should not be done. The reason for the necessary steps to be taken may also be given. This sub-mode listed in Hale et al. (1996) seems to be identical with what Brooks and Warren call *technical description* which is, of course, an exposition not a description as the name may misleadingly suggest.

3.5.1.4 Cause and effect/ problem-solution

This mode of discourse requires writers to present a causal relationship or define or describe a problem along with its solution. If the writer is presenting a causal relationship, it can be either from cause to effect or from effect to cause. This category has been distinguished as an independent sub-mode of exposition by Hale et al. (1996) but Brooks and Warren (1952) include this in the analysis mode and call that *causal analysis*.

3.5.1.5 Comparison and contrast

Comparison and contrast is another technique of exposition in which similarities (*comparison*) or differences (*contrast*) between two or more similar but not identical situations or subjects are pointed out. There can be three types of intentions behind any comparison and contrast:

1) The writer wants to inform the reader about one thing by relating it to another thing or other things that the reader already knows properly.

2) The reader does not know either sides involved in the comparison and contrast, therefore, the writer tries to inform the reader about the things being compared and contrasted by using general rules and principles familiar to the reader.

3) The reader knows both sides of a comparison and contrast so the writer tries to draw the reader's attention towards some general rule or idea.

The items being compared and contrasted are usually dealt with one by one but, if the items are highly complicated and there are too many details attached to them then it is better to compare and contrast the related details rather than explaining one item thoroughly and starting the next one.

3.5.1.6 Classification and Division/Enumeration

These terms obviously are ways of dealing with groups of items which bear some common features and can be regarded as classes. However, if the focus is on putting individuals together because of a certain common quality or feature this is classifying but in division a class is broken down into sub-categories according to the *differentia*, the characteristics that can separate the members of a class, since it is not shared by them. It must be noticed that only one criterion should be applied at each stage of

classification and division; besides, as Brooks and Warren (1952, p.71) put it "the subclass under any class must exhaust that class." In other words, all the members of a class should be potentially accounted for in the subclass.

Generally, there are two types of classification and division schemes: **simple**, in which there are two classes based on a certain characteristic that distinguishes between those who have that characteristic and those who do not. and **complex**, in which there is no limitation and one can recognise as many classes as available.

During the process of complex classification and division it is natural to come up with larger classes at the top of a taxonomy and smaller classes at the bottom. The further the process is carried out the smaller the classes become. Each larger, more general and superior class is referred to as a **genus** (superordinate in semantics) and each immediate smaller, more specific and inferior class is considered as its **species** (hyponym in semantics).

Hale et al. (1996, p.14) in their outstanding work consider any situation a case of classification/enumeration if the writer is expected to "(a) break or clusters, objects, events, or people according to their common elements, factors or characteristics; (b) devise a system for categorization of objects; and/or (c) list a number of items.

3.5.1.7 Analysis

A very important way of writing exposition is by way of analysis in which a subject is divided into its component either spatially (*technical analysis*) or logically (*conceptual analysis*). It is obvious that the former is more suitable for objects with some sort of physical entity while the latter is mainly used for abstract ideas. In an analysis the relationship between these parts and the way they contribute towards the main

characteristic function or purpose of the subject they constitute may also be explained (*functional analysis*). Besides, analysis may take the form of analysing different stages of a process which, then, will be regarded as *chronological analysis*. If this linear sequence of events is exceeded in order to try to establish the reason why and the conditions in which this is happening then the analysis will be of a *cause and effect* type.

The major difference between analysis and classification is that in analysis the relationship among the parts should be identified.

According to Hale et al. (1996, p.15) one may do one or several of the following while writing an analysis:

a) apply some theoretical framework to an objective in question

b) apply the interpretive methods of a specific school of thought

c) distinguish facts from hypotheses

d) evaluate assumptions, interrelationships, or causal relations

e) detect logical fallacies in arguments

3.5.2 Argument

Another mode of discourse is argumentative discourse in which the writer tries to convince and/or persuade the audience to accept his or her viewpoint and abandon their own alternative choice. According to Frowe (1989, p.55) "central to arguing is the giving of reasons in order to support a particular point of view" and therefore unsupported assertions are not considered argument (Hale, et al., 1996). The major difference between exposition and argument is that exposition is mainly for informing the audience but, argument tries to resolve a conflict between two opposing options or

opinions and if it tries to transfer any information it is mainly to tell the audience that there has been a misunderstanding which should be cleared up.

As Freedman and Pringle (1989, p.75) point out "writing argument is not the same as being argumentative in writing". They propose a simple working criteria for argumentative structure and that is "a written argument needed only a clear thesis (either explicit or implicit) from the beginning and a substantiating set of logically developed points and/or illustrations proving the thesis and forming the body of the essay" (ibd. pp. 75-76).

In any argument, the conflicting sides hold an idea called a **proposition**. There are two types of propositions. First, when the holder of the proposition believes that something is true. This is called the *proposition of fact*. Alternatively, the sides involved in an argument may hold that a certain policy should be adopted or abolished. For example one of the argumentative topics in the present research was to ask the subjects whether our resources should be spent on space projects rather than relieving the poor of the world. This type of proposition is called the *proposition of policy*.

There are certain characteristics for good propositions. A good proposition should be clearly stated. To do so it is recommended to avoid vague and subjective words. It has to express only one argument, too. Besides, the proposition should be unprejudiced. In other words, the conclusion should not indicated in the proposition otherwise, there would be no need to argue. Afterwards, all the pros and cons of the proposition should be taken into account and the available evidence should be cited to support the writer's argument. The evidence can be already established facts or opinions stated by authorities acceptable to the targeted audience.

Having enough evidence in hand does not guarantee a sound and successful reasoning. There are certain methods to help the argument be logical. One of these methods is generalising by induction in which some particular examples are given as evidence and then a general rule is given as logical conclusion. Another type of induction by analogy in which two instances are shown to be similar with regard to some important points and then it can be concluded that the two instances are similar or behave in a similar way when it comes to the point to be proven.

Another method of reasoning is deduction. Deduction, unlike generalisation and analogy, is not based on some degree of probability but it is intended to prove axioms and induce certitude. In one pure form of deduction, called a **syllogism,** a *major premise* and a *minor premise* are brought together to result in a logical *conclusion*. There is another form of deduction called a hypothetical syllogism in which the reasoning takes place on a *if-then* basis. Although it is very rare to find a syllogism in its pure form in argumentative writings, the implied form of syllogism technically known as an **enthymeme** is frequently used by writers when they draw conclusions even though the major and minor premises are not apparently stated.

Argument is not always to prove things but it can be used for refutation, too. One of the common techniques of refutation is called *reductio ad absurdum* argument. In this type of argument a particular premise is initially accepted to draw out its "unacceptable (absurd?) . . .unpalatable implications" (Frowe, 1989, p.61). There is another type of argument mentioned by Frowe (ibid) which is explained below as a type of fallacy.

All types of argument are subject to fallacies. These fallacies which are usually difficult to detect are quite common and make the argument yield an erroneous result. *Equivocation* is a type of fallacy in which, as the name implies, one single term is used

in different meanings. Another type of fallacy known as *begging the question* happens when the argument is based on assumptions which are not proved. A very common type of fallacy is ignoring the question. Here the arguer jumps the question for different reasons including his lack of knowledge about the issue or for the dislike of it. This avoidance is similar to what happens in essay tests, known as *challenge,* when the examinee writes about something different from the assigned topic. An even more common type of fallacy is the non sequitur in which an association between the absence or presence of two events is mistakenly considered as a cause and effect relationship. There is one more type of argument listed by (Frowe, 1989, p.61) called *ad hominem* arguments which seems to be a fallacy rather than a sound argument. She maintains (ibid) that "An *ad hominem* argument challenges the consistency of the speaker's beliefs and practices" and then admits that it may not settle the point at issue, but rather forces scrutiny of the proposer's position. All these fallacies will result in defective reasoning.

Many writing textbooks put an extra emphasis on argument and the writer's preparation in demands before getting into the actual writing. For instance Brooks and Warren (1952, p.172-3) recommend:

> The composition of extended argument calls for very careful planning. One point must lead to another, effect must be traced to cause, premise must give conclusion. Random thoughts, no matter how important in themselves, will not carry conviction. Therefore it is a good idea to think through an argument before beginning the actual writing. To prepare a systematic outline of the argument is the best way to be sure that the subject is covered and the relationship among the parts is clear. (p.172-3)

Following this piece of advice they recommend that a good argument should include an *introduction*, an *argument* or *discussion* and a *conclusion*. The introduction should give all the necessary information about the present situation. All the evidence which contributes to the writer's desired conclusion and all the evidence that helps to refute the opposing alternative should be given in the discussion. And finally in the conclusion the writer points out the important points in the discussion and relates them to the question and the final resolution. Throughout the process of writing an argument the writer should bear in mind that the argument cannot be a successful one if it fails to persuade the audience to come to agreement with the writer. To achieve this crucial persuasion the writer should keep the style of presentation as vivid as possible to catch the interest of the audience. Besides, to help to overcome hostility or indifference, a common ground should be established between the author and the audience and a respect for the audience should always be present in arguments. After all, as Andrews (1995) put it, argument is a dance rather than a battle.

3.5.3 Description

Description, or according to Brooks and Warren (1952, p.195) suggestive description, is "the kind of discourse concerned with the appearance of the world. It tells what qualities a primarily concrete object has, and what impressions it makes on our senses. It aims to suggest to the imagination the thing as it appears immediately before an observer." The word suggestive is used to distinguish the real description from other types of descriptive modes like technical description which is really a type of exposition and has already been discussed. Besides objective and tangible things and places, more subjective issues like feelings and states of mind can also be described

which, of course, compared to ordinary description, requires more skill and imagination on the side of the writer and a little more figurative language. Description is one of the most versatile modes and is usually used in conjunction with other modes, especially narration.

In description the writer tries to impress the reader by the use of a single sense e.g. smell or vision. To appeal to a sense the writer has to give the details of the subject he or she is describing. In real life when the observer uses the senses all the details are present in the first encounter but in a written description the details must be given one by one in a sequence. This requires the writer to choose a suitable pattern for the details to be described. This pattern depends on the view angle from which the writer is observing. The point of view can be *fixed*, *moving* or what Brooks and Warren (1952, P.203) call *frame image*. Obviously, in a fixed pattern the writer gives the details in the same order that he observes the object to be described. In a moving pattern the writer does not stand in a certain place but moves around and therefore the point of view changes constantly. If an object cannot be described in either way because of its size or other reasons then the writer may compare it with a smaller object and use it as a frame to help the reader to visualise the details.

The patterns explained here for a piece of description are based on the objective physical position and viewpoint of the writer. However, this viewpoint can be a subjective one depending on the author's attitudes towards the issue which is being dealt with. There is no doubt that the arrangement of details in two pieces of writing about the same thing but written with two different *moods*, one positive and one negative, would be different. Besides, the details written in a piece of descriptive writing depends totally on the *interest* of the writer. Each writer includes just those

details which are of interest to him or her and the purpose the final written text is supposed to serve. In yet another pattern of writing a description, known as *impressionistic pattern*, the writer enlists several details of the thing being described in no obvious order but to leave a desired impression on the reader with regard to a certain mood or interest. The last distinctive pattern of achieving unity in description is what Brooks and Warren (1952, p.207) call *absorbed description*. In this pattern the descriptive details are related to each other through a line of action or explanation or argument.

So far, seven patterns of arranging a description have been distinguished and discussed in this section. However, it is easily imagined that, in most cases these patterns combine to leave a more effective impression on the audience. But, successful use of different patterns requires some degree of experience.

What has been said with regard to patterns concerns the ordering of details given in a description but it does not explain how these details should be selected and presented. Selecting and presenting are parameters determining the texture of a written description.

The details which are included in a description should be both *vivid* and *significant*. Vividness is defined by Brooks and Warren (1952, p.212) as "striking" and a descriptive detail is vivid "if it can set the imagination to work so that the reader calls up the object in his mind's eye". Such details can be either obvious characteristics of the thing being described or it can be a minute and delicate feature which can only be emphasised and kept in the spotlight with the help of masterly subtle language. The latter, if exaggerated, can create a caricature rather than a plain description. Details also should be significant. In other words, only those details that contribute to the

creation of the dominant impression and the mood desired to communicate should be included and any distracting detail should be avoided.

In academic writing description generally concerns with "the qualities of a type, class, or group" Hale, et al. (1996. P. 13).

3.5.4 Narration

Narration may be defined, as maintained by Scholes (1981, p.205), as "the symbolic presentation of a sequence of events connected by subject matter and related by time". Narration, as a mode of discourse, is not merely a fictional story telling. It has a much broader meaning to refer to a reported sequence of actions and movements happening over a duration of time. Narration tells the reader what has happened and how it happened. It does not explain about an event, its cause or results, nor does it intend to make us see certain aspects of it. Narration, instead, presents the event itself. A narrative is the final product of that mode of discourse known as narration. Not all and every single narration necessarily ends in producing a satisfactory narrative. To qualify as a narrative, according to Freedman and Pringle (1989, p.75) "a piece of writing would have to include some information about the setting as well as at least one complete episode, a complete episode being one in which a protagonist responds to an initiating event, primarily through goal-oriented behaviour".

A more comprehensive view of what a narrative is, has been given by Cortazzi (1994, p. 158). He enumerates at least "three necessary conditions for narrative":

-*temporality* ; there must be a sequence of events from a beginning rising to a middle action which is the peak of tension and a finishing in a resolution.

-*causation*; the final state is the inevitable outcome of the middle actions

-human interest; a narrative should "allow for the projection of values and motives to characters in a story" (ibid, p.158)

Time in narration is a key issue. In narration the time cannot be segmented out as it is in description. Narration requires a unit of time (episode) in which an action takes place. Of course, in a written narrative, unlike Labov's restrictions (1972), it is not necessary to tell the events in their natural order. The writer, instead, can change the sequence of events, by using flash backs and flash forwards, to catch the reader's attention.

Once again it should be emphasised that narration, like other modes of discourse, is usually found mixed with other discourse types. As narration tells us about an action and actions are done by people therefore a cause and effect relationship usually exists in narratives. So there is a close relationship between narration and exposition. Not only does narration use exposition, other modes of discourse, especially argument and exposition, in return, use narration. All the jokes, anecdotes, and stories with which a writer tries to illustrate his explanation or convince the audience are good examples of the involvement of narration in other types of discourse. However, sometimes narration becomes completely intermingled with exposition as in accounts of actions carried out by a researcher in the laboratory during an experiment. As discussed in the exposition section, this type of discourse is in fact an *expository narration* rather than a type of narration.

3.5.4.1 Patterns in narration

One of the simplest patterns is proposed by Hoey (1983). It is based on a problem-solution pattern which identifies the following elements:

Situation

Problem

Solution

Evaluation

However, it seems that the simple working pattern given by Brooks and Warren (1952) is more convenient. They distinguish three major parts in any narratives. Firstly, any narration needs a background, some necessary information about the situation to begin with. This background, technically called *exposition*, can be very short and simple, like that of jokes or quite lengthy and complicated. After this background comes the middle or main part of narratives in which a series of events are given which will finally lead to the pinnacle of the *complication* in the story called the *climax*. If a narrative fails to create such climax it will fail in drawing the reader's attention, too. As a narrative reaches its end or *denouement*, the climax tends to be resolved and is replaced with the conclusion. It is here that the reader usually gets the full meaning of the actions being narrated and achieves some sort of awareness. These three sections exist, or at least are implied, in almost all narratives, though in many cases it is hard, if not impossible, to draw a line between them and separate them completely. In a widely-cited model, Labov (1972), Labov et al. (1967) and Labov et al. (1977) distinguish six parts in oral narratives: *abstract, orientation, complication, evaluation, result* and *coda*. Among these elements, abstract and coda are most particular to oral narrative and storytelling. This model was later trimmed by Martin and Rothery (1980) who used only three elements: orientation, complication and resolution. Cortazzi (1994) emphasises that the sequence of these elements is not a fixed one , for example,

evaluation can occur almost anywhere. He (ibid) prefers a more elaborate model proposed by Longacre (1976) which comprises the following six elements:

aperture an optional formulaic opening

stage information about time, place, participants

episode(s) an *inciting moment* which gets something going, a *developing conflict* which intensifies the situation, and a *climax* or resolution

denouement a crucial final event after a series of episodes

conclusion optional narrator's comments or interpretation

finish a formulaic closing

(From Cortazzi, 1994, p.159)

All these patterns can be summarised as follows:

Figure3.5 : Patterns of narration

Brooks & Warren	Hoey	Labov	Martin &Rothery	Longacre
		Abstract		Aperture
Background	Situation	Orientation	Orientation	Exposition
Exposition (Climax)	Problem	Complicating action	Complication	Inciting moment
		Evaluation		Developing conflict
				Climax
Denouement	Solution	Resolution	Resolution	Denouement
				Conclusion
	Evaluation	Coda		Finish

Adapted from Harris (1993)

In different narratives, the size of these sections differ from each other depending on the situation. Sometimes it is necessary to give the reader a lengthy and detailed introduction or *exposition* while in other situations a brief introduction may suffice. Similarly it is possible to give a brief account of several years in a paragraph as a scaffolding to the main story which may happen in a couple of hours but occupying several hundred pages. The important factor in determining the proportion of each section is the judgement of the writer to decide what details should be included in a narrative. Once again, like description the issue of selection seems to be of vital importance.

Another important issue in writing narratives is the *point of view*. By point of view, the relation between the narrator and the action is meant. Unlike description this relation is not a physical one. Here this term signifies the position of the narrator in the events being narrated whether the narrator is involved and therefore the first person point of view is used in the narrative or the narrator remains simply an outsider and thus it is just a third person point of view. Even in each of the two major types of point of view several variations can be distinguished. In the former point of view the narrator can be either the main character or merely an observer and in the latter i.e. the third person point of view the narrator may acquire different approaches towards the events and the characters according to the way he or she sees them. For instance if the narrator is an omniscient one who knows every thing the reader finds it possible to be aware of even the intentions and thoughts of characters and can be aware of the events which happen in other places where the character is absent. This is of course more possible in more imaginary narratives. At the other extreme, the narrator may focus on a certain character and tell the story in the way that character sees it. And these two extremes can mix and create a variety of points of view. While point of view is widely considered of major importance in narrative, and it is clearly potentially influential in description, this aspect of mode has rarely been discussed in relation to exposition or argument. This seems odd, since academic exposition, or argument necessarily involves (one or more) points of view.

As academic tasks, Hale et al. (1996, p.12) observed that, narration deals with "particularizaing concrete events rather than generalizing about abstract notions".

3.6 Summary

This chapter has tried to review the issue of modes of discourse. Different systems of classifying the modes have been discussed and an attempt has been made to find their roots in history. Besides, the four basic modes have been viewed in more detail. The next chapter, chapter four, is the beginning of the present research study which comprises the methodology used for both the pilot study and the main study.

CHAPTER FOUR
METHODOLOGY

"We know surprisingly little about the differences in performance that are caused by different kinds of writing topics, but we do know that such differences exist."

(White, 1985, p.116)

4.1 Introduction

Perhaps the methodology chapter is the most interesting, and possibly the most important part of any doctoral thesis. The obvious reason for suggesting this is that any significant results and outstanding findings are based on the method with which the research has been conducted. Minor flaws in the design of the study, which may be and often are overlooked, can affect the outcome drastically. When it comes to statistics the situation can be even worse as a competent statistician might find ways to support even a false hypothesis or otherwise disguise actual findings (Huff, 1954). Bearing these facts in mind, this chapter attempts to give readers an open and thorough account of the procedures taken to carry out this research study.

This chapter comprises two sections: section one which explains the pilot phase of the study, while the second section reports the main study. The main study has a threefold division. This is because the main study investigates how EFL writing varies across different modes of discourse using the three dimensions of production, recognition and learners' attitudes, as discussed in chapter one.

4.2 The pilot study[*]

Before embarking on the main study it was necessary to run a small-scale pilot-study. As discussed in chapter one, research findings suggest that narration and description,

---

The pilot study of this research investigation has already been published in Nemati (1996).

compared to argumentation and explanation, are considered easier modes to write in for young L1 learners (Kincaid,1953; Braddock et al., 1963; Moffett, 1968; Rosen,1969; Veal and Tillman, 1971; Perron,1977; Crowhurst and Piche,1979; Crowhurst, 1980). However, at the beginning of the study, it seemed to this researcher that the evidence for this was based largely on studies conducted with younger writers or first language writers; for adult, academic L2 learners the situation is reversed. It seemed logical to expect students with a heavy background of logical reasoning and critical thinking to be more fluent in writing, let's say, argumentative paragraphs compared to writing a piece of narration. Therefore, the pilot study was actually a first step in the process of finding a way through the previous blurred and even contradictory research findings.

4.2.1 Research questions and hypothesis

The research question for the pilot study was whether or not it is more demanding to write in certain types of discourse compared to others and how non-native learners' performances differ across different discourse modes. The independent variable in the pilot study was the mode of discourse which was divided into three subgroups: *Argumentation*, *Description*, and *Explanation*. These discourse types were manifested in three suggested prompts (see the **Materials and Instrumentation** section). The dependant variable, then, was the scores each subject received for the essays written in these three topics. The range of the scores, following the IELTS was from 1-9.

Consequently, the following null hypothesis was proposed:

H₀: There is no statistically significant difference between the scores of essays written in three different discourse modes (i.e. *argumentation*, *explanation* and *description*) obtained by the same NNS subjects.

It should be pointed out only three discourse types were included in the pilot study and that narration was left out. This is because teachers at the Leicester University Language Centre (where the pilot study was conducted) preferred not to give their students narrative tasks arguing that there is no place for narrative writing in a language class intended for academic purposes. This refusal, while disappointing, seems to lend indirect support to the researcher's perception mentioned above: if narrative writing is considered "non-academic" it may be easier to write but usually it may well be less familiar and much less practised as a mode compared to those other modes to which attention is clearly given in such EAP courses. As a result of this refusal, the design of the pilot study was not as complete as the researcher wished it to be. However for this stage of the research it was acceptable enough to work with three discourse types as this would give a preliminary view of the way that the topics work.

It can however be argued that this reluctance on the side of most language teachers to prepare their students for narrative writing is not justifiable. On the importance of Narrative discourse type Cortazzi (1994) writes:

> "Narrative is one of the most frequently occurring and ubiquitous forms of discourse. ... Narrative is, of course, an important genre in its own right. It is probably the first to be acquired at home and the most exploited in the early stages of learning in school. Certainly it has been the most studied. ... Narrative is now seen by many scholars as being of fundamental importance to our mental and social life. (p.157)

All this is enough to establish narration as an indispensable part of language. Besides, it is also unjustifiable to argue, as it is usually assumed to be the case, that narrative is not that important in academic life, either. Referring to Bruner (1986, 1987, 1990)

Cortazzi (1994, p.157) believes in the educational value of this type of discourse and considers it as a complement for logico-scientific thinking which "deals with observation, analysis and proof while the second [narrative] handles issues of belief, doubt, emotion, intentions, and accommodates ambiguity and dilemma."

4.2.2 Method

In order for the researcher to get a rough idea about the topics, scoring procedures and so on, this pilot has been designed. The Leicester University Language Centre (LULC) was selected for the experiment. LULC is a supporting centre for new students to brush up their academic English and improve the language skills required in an academic environment, including writing.

4.2.3 Subjects

The subjects of this study were chosen from the foreign students taking pre-sessional courses at Leicester University Language Centre. Twelve students, male and female, with different nationalities, were randomly selected from two different writing classes. The composition of subjects is given in Table 3.1:

Table 4.1: The profile of students participating in the pilot study

Under/Post graduate		Sex		Mother tongue					
under	post	female	male	German	Mandarin	Spanish	Greek	Japanese	Swedish
5	7	9	3	1	1	3	2	4	1

125

As these students had already been admitted by the university and it requires a language proficiency score of 550 in TOEFL or band 6 in IELTS, it could be roughly concluded that the subjects were either at upper-intermediate or advanced level.

4.2.4 Materials and Instrumentation

The subjects were required, as part of their assessed work, to write an essay in each pilot session. The time allowed for each essay was 30 minutes in order to keep the situation similar to the timed essay evaluation practised by major language test batteries.

In order to gain the agreement of LULC to run the experiment there, it was necessary to compromise with the teachers in general, and the writing course co-ordinator in particular, over choosing the topics. After some discussion narrative writing was omitted, as mentioned earlier. The descriptive prompt was also modified in way to provide the LULC with the students' perceptions of their needs. Consequently, the students in the pilot study wrote three compositions in three consecutive sessions each in a different mode of discourse - i.e. *Argumentation*, *Explanation*, and *Description*. The essay prompts used in the pilot study are given below:

Argumentative prompt: **"Argue the advantages and disadvantages of T.V. Take a clear position."**

Descriptive prompt : **" Describe your background in learning English language and your expectations of these language classes."**

Explanatory prompt : " Explain how a foreign language should be learned"

4.2.5 Scoring Procedure

The compositions were scored by two raters, separately. The raters were both experienced teachers of English with masters degrees in TESOL. In order to avoid imposing any criteria on the raters and risk affecting their judgements, the use of analytic profiles for scoring ESL compositions was rejected in favour of a pure impressionistic method. Therefore, the raters were given a chance to judge the essays solely based on their professional judgement and intuition and the impression made by the writings. It was vital to leave the raters on their own, without feeling the burden of focusing on particular features. It is necessary to reiterate that this research project aims, at least in one phase, at the complex variations which probably exist between different types of written texts and the way they affect the reader's mind. So using any pre-fabricated profile could hinder the natural interaction between the text and the reader. This advantage of impression marking has been echoed by Madsen (1983, p.121) who states that "it is one of the best ways to evaluate the complex communicative act of writing".

The raters were not told to concentrate on certain aspects of the writings either since the researcher, with Weir (1990), believes that the separation of discrete features of compositions is the violation of the principles of impression marking. However, before embarking on rating the essays, the raters met each other in a co-ordination session, not only to alleviate the lack of a scoring profile, but also to achieve a unified scoring

scale, which finally was agreed to be a nine-level one, similar to that accepted and practised by IELTS (see appendix G).

4.2.6 Scoring reliability

Similarly, the interrater reliability for the two raters of the pilot study, computed through the *Pearson Product Moment*, was as high as 0.68 ($p<0.000$) which is an acceptable, if not good, reliability estimate for raters without specific training who judged the compositions merely according to their impressions. This, once again, confirms the idea stressed by Munro (1991) and demonstrated by Nemati (1993) that despite what it may seem at the first glance, impression marking is quite reliable. However, the effect of having co-ordinating sessions cannot be ignored. Co-ordination seems to be more important than training in its technical terms. Shohamy et al (1992, p.30) reports that "overall, no difference existed between the professional and lay raters." This means that to be a rater it is not necessary to be an outstanding expert with long experience in evaluating writing. Especially for research and small-scale testing any experienced language teacher can be considered as a potentially acceptable candidate. On the other hand, co-ordination is very important to make sure that the raters have a clear understanding of the scoring scale and typical essays they are going to evaluate. Weigle (1994, p.214) stresses the point that the major benefit of training sessions is that they help the raters "to understand and apply the intended rating criteria" and modifies "the rater's expectations in terms of the characteristics of the writers and the demands of the writing tasks." She also adds that the familiarity between the raters may increase the conformity among them and hence increase the inter-rater reliability.

4.2.7 Analysing the Data

Due to the design of the pilot study, the final outcome was three sets of scores each belonging to the subjects' essays written in three different discourse modes.

The null hypothesis of the study was that there is no statistically significant difference among the essays written in different modes of discourse. In other words, an ESL writer should presumably perform equally well whether the given prompt elicits an argument, an explanation, or a description.

Choosing the appropriate statistical test is not entirely straightforward. There are different options open and it is up to the researcher to choose the most befitting test to examine the data. In order to find the most appropriate statistical analysis a researcher should take the following variables into consideration:

- Is it a similarity or a difference to be established or rejected?

- Are the data nominal, ordinal, interval, or ratio?

- Has the data collection been based on a random selection?

- Do the data consist of pairs or independent observations?

- Are the variables normally distributed?

The final decision to select a statistical test depends on these conditions. These assumptions still cannot pinpoint the most suitable statistical analysis and usually these assumptions are either ignored or misinterpreted. For example regarding the issue of normal distribution and its requirement for parametric tests, Peers (1996) points out that:

There is much misunderstanding about what is meant by **assumptions of normality**. It is often believed that, for example, to use parametric tests such as the paired t-test (for 'before' and 'after' designs) or linear regression, the response variables should be **normally distributed**. This not true. (p.253)

He further adds that "the assumptions of normality are usually based either on faith or as Siegel and Castallan(1988) put it, "rest on conjecture and hope" (p.35).

The controversial issue here is whether or not the scores are in the form of interval scores or simple ranks. As the compositions were actually compared against each other rather than against a criterion or profile, it seems that it is of a rankable nature. Besides the number of subjects in each group is limited to 12 and the assumption of normal distribution is actually violated.

According to Hatch and Lazaraton (1991, p. 355) "if you cannot meet all the assumptions of ANOVA (whether for descriptive or inferential purposes), then you would be better served by a nonparametric test such as the Friedman." In other words Friedman is a parallel test to repeated measures ANOVA.

This non-parametric test is used to compare the distributions on more than two related samples. Using this test with the SPSS, results in converting the data into ranks. The minimum number of observations required for a Friedman test is 10 (ibid. p.356) The number of observations for each group in the pilot was 12.

To investigate the probable differences among the data collected for this phase of the study, after considering all the pros and cons, a repeated-measure ANOVA was considered to be the most appropriate statistical test. And as this is the pilot study, to

compare and confirm the results a substitute nonparametric (Friedman test) was also run on the marks given for the three types of compositions.

The use of both a parametric and a non-parametric test is not because of the researcher's ignorance in using the most appropriate one. It is, rather, because of his awareness towards the current doubts and hesitations that prevail research manuals about the legitimacy of these tests. For instance, Hatch and Lazaraton (1991) believe that ordinal data is sometimes regarded as suitable for parametric and sometimes for nonparametric tests. Besides, other authors such as Siegel (1956) in his classic book *Nonparametric Statistics for the Behavioural Sciences* argues that in humanities researchers cannot produce that sort of data which is required by the parametric tests and therefore they should be abandoned. To moderate this strong claim Heyes et al. (1986, p.89) point out that despite the fact that the assumptions required for parametric tests are usually violated, some of these tests especially the t-tests "are said to be **robust**, that is they can cope with the data which does not fully meet the assumptions made by parametric tests in general."

4.2.8 Results and discussion

A repeated-measure ANOVA was applied to the data to make sure that a robust parametric test verified that the difference among the means is statistically significant and not accidental:

Table 4.2 The result of repeated-measure ANOVA for three discourse types

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	4.80	22	.22		
TYPE	4.16	2	2.08	9.52	.001**

The results are highly significant (P<0.01**) which means that the performance of the subjects on the three tests are not the same and they therefore the topics are not parallel.

Besides, the Friedman test yielded the following results, shown in table 4.3.

Table 4.3: The results of Friedman test for the three types of compositions

Mean Rank	Variable
2.14	Description
2.64	Explanation
1.33	Argumentation

Cases	chi-square	D.F	Significance
12	8.66	2	.0036**

Not surprisingly, the Friedman test confirmed the previous findings and similar to its parametric counter-part quite strongly rejected the null hypothesis (P = 0.003**). So, it can be concluded that the results obtained through three different essay titles from

the same subjects are significantly different. In other words, different modes of discourse are of different difficulty levels.

There is one final point worth mentioning and that is the underestimation, and in some cases, the dismissive attitude towards narrative writing in academic environment. Unfortunately, the experience with this study, as well as many other observations, show that many teachers believe that there is no room for narrative writing in academic life. The reason given here was that narrative is "non-academic", which is arguably a myth. In fact, Cortazzi (personal communications) has shown that reports of case studies and much other qualitative research reporting commonly use a narrative format. He has found, further, that many contemporary first year university science textbooks (e.g. in biology or psychology) consistently use narrative modes in major sections of reporting experiments and, more interestingly, these modes are used to focus on the working lives of professionals and scholars in sections of such books. Narrative, it seems, is, after all, a major mode in many areas of academic writing (including parts of the present chapter).

General language proficiency test results are largely used by universities for admitting new students. The reasoning behind this is that the students should be able to fulfil the writing tasks of higher education, which is quite clearly the case; yet, they fail to realise that an academic text is not necessarily solely an argumentative piece of writing, nor is it a descriptive one. It might be true that at macro level narration is not the main concern of the university students. However, students in many cases will need to use narration, at least as part of their writings. The subordination of narration to other modes of discourse, by no means implies that narration should be ignored in writing classes.

4.3 From pilot to main study

The findings of the pilot study and the experience gained from it were crucial for the administration of the main study, Firstly, it became clear that the topics were different in the type of discourse they would elicit and this difference was achieved successfully by simply manipulating the wording of the topics in order to make writers produce a text of that certain mode. Once again it should be recalled that it is extremely difficult, if not impossible, to elicit compositions purely within one and only one type of discourse. However, the pilot showed that it is easily possible to elicit compositions which mostly differ from one another in their general modal nature. Secondly, as a starting point, it was presumed that argument has been identified as a difficult mode to write in, simply because this was apparently demonstrated in previous experiments focusing on children at their early years of schooling. It was further assumed that argumentation may later become an easier discourse type for university students with academic maturity and writing practice. The pilot study revealed that the teacher's presumption about such cross-over effect of the modes due to maturity, brain development, and involvement in critical reasoning throughout the academic years were questionable. Thirdly, it was shown that a general impression marking method is reliable enough, at least, to score the essays for this special purpose.

With the experience and results gained from the pilot study this researcher set out to collect the data in Iran. The duration of the main data collection period was about three months from 26 November 1995 to 23 February 1996. The following section is a report on the methodology used to collect the data for the main study.

4.4. Introduction

The main purpose of this study was to investigate the EFL/ESL learner's performance when writing in different modes of discourse (i.e. argument, description, narrative, and explanation) to solve the controversy concerning whether or not an assigned topic affects the writer's performance on the elicited composition. Despite the huge bulk of research done in L1, indicating the existence of such differences, it seems that the major test batteries have taken it for granted that one's fluency in second language writing can be simply evaluated by having them write a piece of writing (Carlson et al. 1985). Personal observations and, later, the pilot study results proved contradictory to the currently accepted assumptions. In order to shed some light on the issue it was decided to fully investigate the situation in three linguistic and cognitive levels of production, recognition, and preference. A thorough account for each phase will be given here.

4.5 Main research questions

The main research questions have already been briefly introduced in chapter one. Here, they are elaborated as they will apply to the main study, following the pilot study.

Research question #1: Are some modes of discourse (i.e. *argument/explanation*) relatively more difficult to write on by EFL learners compared against other modes (*i.e. narration/description*)?

Research question #2: Do the essays written in different discourse types bear different degrees of cognitive and/or linguistic difficulty, when turned into cloze tests, to recognise and reconstruct?

Research Question #3: Do EFL writers prefer certain mode(s) of discourse over other mode(s) while choosing a topic to write on?

Consequently, four null hypotheses were proposed, two for the first question and one each for the other two:

H_0 **#1: There is no statistically significant difference between the scores of compositions written by the same EFL subjects in the two different discourse modes of argument and description.**

H_0 **#2: There is no statistically significant difference between the scores of compositions written by the same EFL subjects in the two different discourse modes of explanation and narration.**

H_0 **#3: There is no statistically significant difference between the learners' scores obtained from four cloze tests derived from four texts each written in a different major discourse type.**

H_0 **#4 There is no significant difference between the learners' choice of essay topics which will elicit texts in different discourse modes.**

To test these hypotheses a three-phase study was designed. Each phase was meant to explore one of the research questions. Therefore, the **production phase** is related to the first two research questions, the **recognition phase** and the **preference and attitude phase** probe the third and the fourth questions, respectively.

4.6 The production phase

This phase of study can be regarded as the main body of the project for it tries to simulate the real life situation in which non-native speakers of English sit for a language proficiency test. Briefly speaking, it was intended to compare four different discourse types -argumentation, description, explanation, and narration. The ideal design would be the one in which each subject had to write in all four types of discourse. Unfortunately, this was virtually impossible because only one essay can be required in one session; otherwise, the second piece of writing would be affected by the fatigue caused by the first essay. Further, it is unlikely that enough class time can be made available for students to write more than one essay in any single class session. This proved to be the case and, of course, this is normally the case in most proficiency tests, too. So, having students write four compositions meant disturbing both the teachers and the students and also the disruption of the every-day schedule for four consecutive sessions which was not acceptable to the authorities of the universities where the research was carried out. Consequently, an alternative had to be adopted: the four discourse types were broken down into two contrasting categories :

- **discourse modes hypothesised as easy to produce** -i.e. description and narration

- **discourse modes hypothesised as difficult to produce** -i.e. argumentation and explanation

The rationale for this classification is that the difference between argument and description has already been established in L1 (see chapter one) and the present pilot study confirms that a difference between the difficulty level of argumentative writing and that of descriptive writing exists. So, it was more logical to investigate and establish the case in L2, too. Besides both argument and description are genres which are admittedly widely used in EAP as well as writing element of test batteries. In brief, by contrasting description and argumentation it was intended to either support or refute the pattern already found in L1.

On the other hand, the contrast between narration and explanation seems a bit more ambiguous. First, in the pilot study, the explanatory mode had not behaved as expected. It was supposed to be a difficult mode to produce. Nevertheless, the pilot study example proved to be a relatively easy mode to produce. Besides, since the narrative mode was left out in the pilot study, there was no introductory information on the probable outcome of the study on this mode available. Besides, Kinneavy (1971, p.79) assumes exposition and narration as opposing modes where he says: "Exposition, as opposed to narration, ..." or "expository writing is also, in such contexts, opposed to narrative writing". For these reasons the two modes were put in one contrasting subgroup according to the information provided from the studies done on the case in L1 which considered narrative as an easy mode and explanation as a difficult mode of discourse.

It might seem that the result of the first group (argumentation vs description) had been predictable. However, it should be recalled that before the pilot study the starting working assumption for the researcher was that the pattern found for these two modes in L1 cannot be securely generalised for L2 situations where the subjects have been

exposed to an enormous number of argumentative texts and have, supposedly, developed a considerable level of critical thinking by learning how to argue for or against a point of view. The following is a schematic view of the researcher's assumption:

Figure4.1: The cross-over effect between an easy mode and a difficult mode of writing
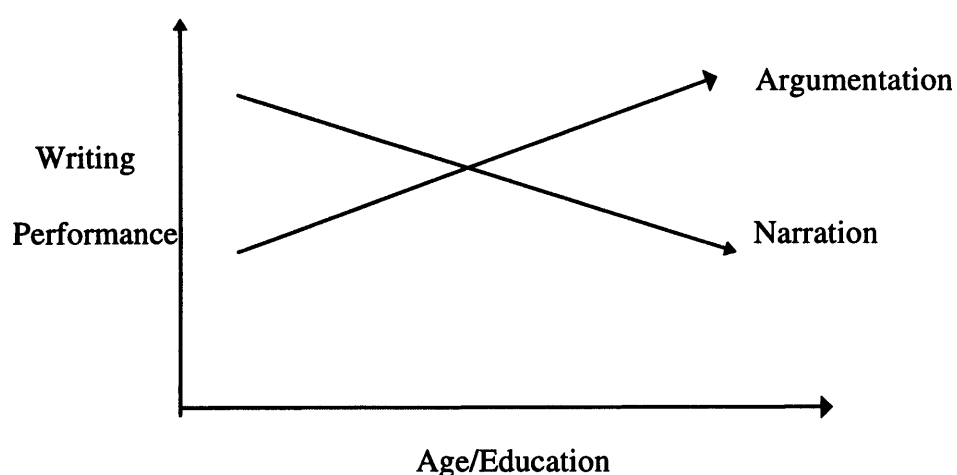
In the next step, then, each member of the former group was paired with one from the latter one. Relying on the results obtained from the pilot study and similar research done in L1, description and argumentation were put into one set, and narration and explanation into another to make two sets of contrasting discourse types. This would mean that the students involved in the field study in Iran would be grouped in the two

sets and that no student would write more than two essays. Each student would, however, write essays in hypothesised contradicting modes.

The main problem in the way of comparing different modes of discourse was to ensure that the probable observed differences were certainly attributed to discourse type differences and not other uncontrolled variables. The easiest solution seemed to be repeating the comparison at least one more time with a different set of prompts. To achieve a high degree of certainty it was decided to repeat the test *four* times, each within a different situation, and aggregate the results.

The final design of the study, after matching the correspondent contrasting compositions, was as follows:

Figure 4.2: The design of dividing opposing discourse types into sub-group classes taking part in the experiment

```
                        ┌─────────────────────┐
                        │ Modes of Discourse   │
                        └─────────────────────┘
                           ╱               ╲
                          ╱                 ╲
                         ╱                   ╲
       ┌──────────────────────┐      ┌──────────────────────┐
       │ Easy Discourse Types │      │ Difficult Discourse Types │
       ├──────────────────────┤      ├──────────────────────┤
       │ Narration            │      │ Argumentation        │
       │                      │      │                      │
       │ Description          │      │ Explanation          │
       └──────────────────────┘      └──────────────────────┘
                      ╲        ╳        ╱
       ╭────────────────────────╮   ╭────────────────────────╮
       │ Description vs Argumentation │   │ Narration vs Explanation │
       ╰────────────────────────╯   ╰────────────────────────╯
                  │                            │
       ┌──────────────────────┐      ┌──────────────────────┐
       │ Des. #1 vs Arg. #1   │      │ Nar. #1 vs Exp. #1   │
       │                      │      │                      │
       │ Des. #1 vs Arg.#2    │      │ Nar. #1 vs Exp. #2   │
       │                      │      │                      │
       │ Des. #2 vs Arg. #1   │      │ Nar. #2 vs Exp. #1   │
       │                      │      │                      │
       │ Des. #2 vs Arg. #2   │      │ Nar. #2 vs Exp. #2   │
       └──────────────────────┘      └──────────────────────┘
```
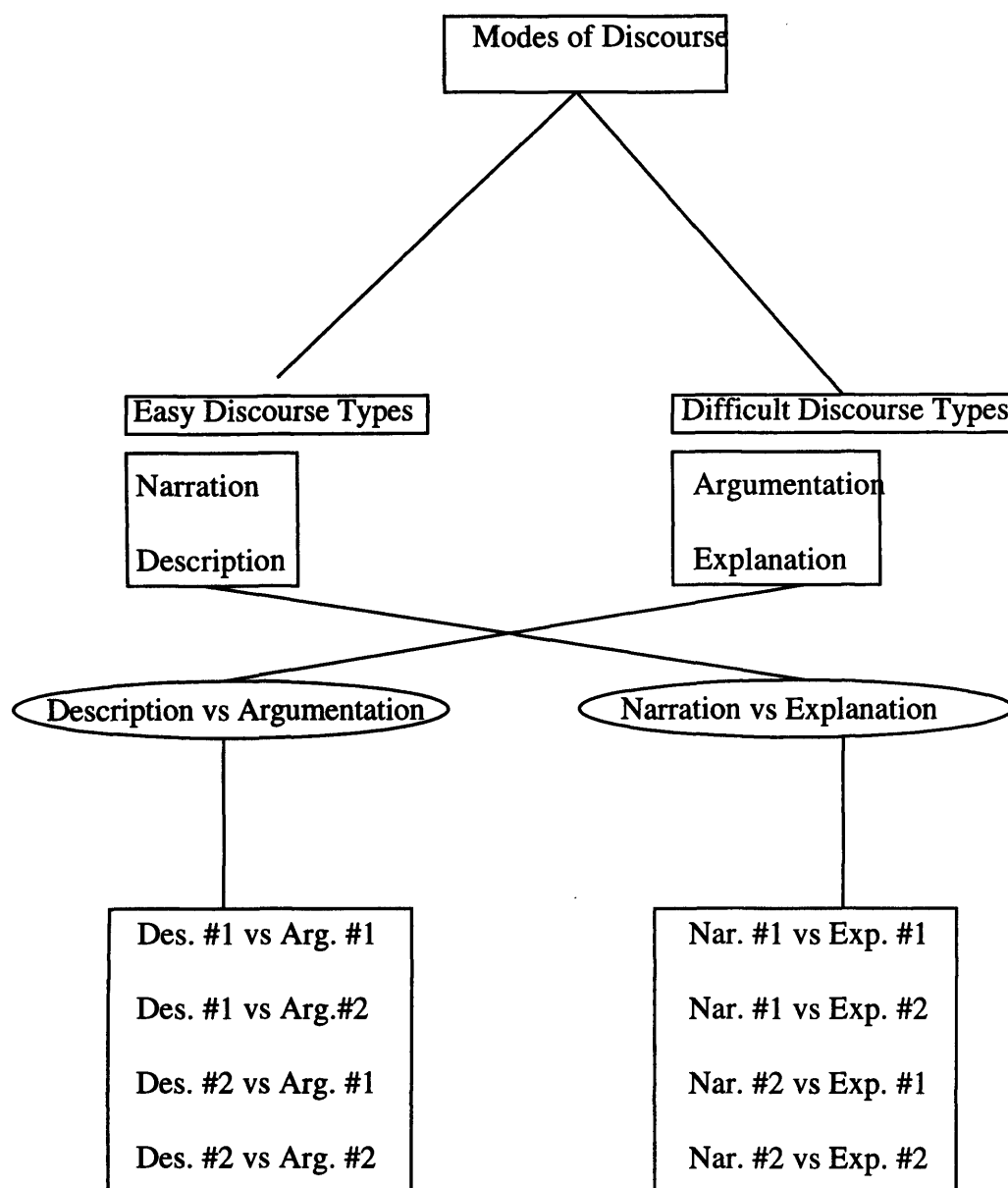
4.6.1 Variables

4.6.1.1 The independent variable

The independent variable at this stage of the study is the discourse mode in which the essays are written. Although there are eight composition prompts, there are only four discourse types i.e. Argumentation, Description, Explanation and Narration. There are two different topics to elicit essays within each mode.

4.6.1.2 The dependent variable

The scores given to the essays written by Iranian EFL learners are considered the dependent variable for the production stage of the study. These scores can vary from 0 to 20. Obviously, the former signifies a total absence of writing performance, and the latter can be interpreted as the ultimate writing performance one can expect a NNS writer to accomplish in a testing session with a limited time to write. This scale follows the normal marking system in Iran and is therefore familiar to students, teachers and raters

4.6.2 Materials and instrumentation

The process to choose the prompts was not an easy one. White (1985, p.108) points out the sensitivity of the job of the test committee for developing writing tasks:

"If it does not come up with good questions, no amount of work in scoring or follow-up administration can salvage the test. Good questions are absolutely crucial...". He further lists four characteristics for good essay prompts:

> *"Clarity*. Students will not waste time trying to figure out what is called for but will be able to get right down to work.
> *Validity*. Good students will receive high scores and weak students will receive low scores. There will be a good range of scores without too large a concentration in the middle.
> *Reliability*. Scoring of pre-test papers show considerable agreement by readers and scoring guide can be readily constructed to describe score differences.
> *Interest*. The question offers sufficient intrinsic interest so that students will write with some genuine concern and those scoring will not go mad (and hence become inaccurate) with boredom." (p.110)

In fact, a pool of topics was created by gathering topics from four different sources:

- consulting several widely used writing textbooks available e.g. *Writing Matters* (Brown and Hood, 1989) *Think and Link* (Cooper,1979); *Write it* (Dean,1988); *Writing as Thinking* (Frank,1990); *Study Writing* (Hamp-Lyons and Heasley, 1987); *In the Picture* (Hedge,1985); *Cambridge Skills for Fluency* (Littlejohn,1991); *Writing in English* (Pincas, Hadfield and Hadfield,1982); *Study Skills for Academic Writing* (Trzeciak and Mackay,1994); *Feedback* (Sherman,1994); *Practise Advanced Writing* (Stephens,1992); *Academic Writing for Graduate Students* (Swales and Feak, 1994); *Teaching and Assessing Writing* (White,1985); *Process Writing* (White and Arndt,1991); *Writing* (White and Mcgovern,1994).

- topics used in other similar research projects e.g. Hamp-Lyons and Mathias (1994).

- asking students, including Iranian ones, to list topics that they believed they would be likely to see in a language test.

- using the experience and the intuition of language teachers including the researcher and his supervisor, as well as the instructors in the LULC, to select the most appropriate topics from the pool.

Writing an essay test topic, in a sense, comprises two steps: first selecting the general topic about which the subjects are to write and second putting this topic on the paper by choosing right wordings. This is a sensitive process. Even a simple, common word like "discuss" can be interpreted in three different ways not only with regard to the writer's approach towards the subject but also each interpretation may require a different mode of discourse (Dudley-Evans, 1988).

Finally, for each type of discourse two carefully selected essay prompts were suggested. The chosen prompts are given below:

Descriptive topics

#1 **"Imagine that you are in the 25th century. Describe the people and their way of life, buildings, vehicles, etc."**

#2 **"Describe your hometown or any interesting city so that readers can have a clear image of the city in their minds."**

These two topics are more or less similar regarding the subject matter they are addressing i.e. describing a city which seems to be one of the top choices made by students. Although, the topic number one demands more degree of imagination compared to the topic number two, which mainly relies on the memory of the writer, considering the fact that most students have seen some futuristic science fiction films, they could readily use their memory or imagination to describe a city in the distant future. The 25th century was suggested to give the subjects a greater degree of freedom so that they could think of whatever they wanted to without any worry about their plausibility. In any case, this type of topic is a classic form of descriptive writing and a similar version of it is cited by White (1985, p.111): "Describe what you see when you walk round your block".

Argumentative topics

#1 **"Argue the advantages and disadvantages of commercial advertising in the media. Take a clear position."**

_#2_ **"At the present time, the human race spends huge sums of money to explore space. Do you find it a waste or do you view space projects as essential to expanding our knowledge. Discuss the issue.**

Both these topics are classical argumentative topics which look to be appealing to the learners. At least, the first topic is almost exactly the prompt that this researcher himself dealt with when taking an IELTS test before moving to Britain to study in Leicester University . The second topic was also given to the researcher as a writing course assignment during his undergraduate studies in Iran. Therefore, the two argumentative topics were chosen for they both have been used in the real world and do not merely come from the researcher's mind.

Narrative topics

_#1_ **"Write the story of the most interesting film you have ever seen. Narrate the events as they happened in the film. Your writing should look like a story.**

_#2_ **"Write the story of someone who rose from an unknown background and humble beginning to become successful and famous. Narrate the main events of his/her life story in the order in which they happened.**

Probably, finding narrative prompts, compared to other discourse types were a bit more difficult as they are not, for already discussed reasons, a common practice in academic writing classes. Surprisingly, there were numerous suggestions for the "most interesting film" topic in the initial topic pool collected which is one of the reasons for selecting this topic. Besides, the nature of the topic helps to achieve a mainly narrative piece of writing while offering the subjects have a wide range of options to choose one single interesting film from so many they have seen so far.

The second topic is a classic topic widely practised in L1 composition for which the biography of Abraham Lincoln is a typical example. However, for Iranian students this is not a very familiar topic and, therefore, could cause some of the some of the unexpected patterns for this topic which will be discussed in the next chapter.

Explanatory topics

#1 **"How can we make teaching and learning in Iranian universities more effective.**

#2 **"Considering the situation in Iran, explain how a foreign language should be learned?"**

Once again, these two topics, while being different, are closely related which makes comparisons easier. As the topic #2 directly addresses the issue of language learning and the subjects are all students of English language and literature it was possible to assume that it would be more appealing.

The main rationale for selecting these two topics is that similar topics were used by the researcher before (Nemati,1993) and they proved to be successful with Iranian students. Both topics centre on Iranian contexts which are familiar to the writers.

4.6.3 Subjects

The subjects of this phase of the study were chosen randomly from English language classes held in various Iranian universities located in Tehran wherever accessible, with no intentional preference whatsoever. Unfortunately it was practically

impossible to test the subjects' language proficiency level but, as these students have already passed a very competitive test, the National University Entrance Exam, it can be concluded that the subjects' language proficiency levels ranged from lower intermediate to advanced level for masters students. This inference is evident from the piece of writing they have produced. However, this variety of command over the English language does not affect the reliability of the research since individuals are compared against themselves, rather it increases the generalizability of the findings. The number of subjects in the largest group is as high as 28 and in the smallest group shrinks to 17. The mode for the class sizes, in three cases out of eight subgroups, is 22. Adding up these numerals, there will be 97 subjects in the **Description vs Argumentation** (Des. vs Arg.) main group and 80 in the **Narration vs Explanation** (Nar. vs Exp.) main group. The total number of subjects taking phase in this part of the study is 183 creating 366 compositions to be investigated. The number students included in each sub-group is given in the table 4.4:

Table 4.4: Number of subjects participating in each sub-group

Subgroups of easy vs difficult modes	Numberof subjects
Argumentative #1 vs Description #1	25
Argumentation #1 vs Description #2	22
Argumentation #2 vs Description #1	22
Argumentation #2 vs Description #2	28
Explanation #1 vs Narration #1	20
Explanation #1 vs Narration #2	21
Explanation #2 vs Narration #1	22
Explanation #2 vs Narration #2	17
Total number of sub-groups =8	N=177

Scoring Procedure

After collecting the compositions the difficult job of scoring had to be done. Two experienced teachers with masters degrees in English language and literature embarked on the scoring job, judging solely by the impression made by the essays. However, they had a co-ordinating session to make their decisions as close as possible. Besides, they also went through the guidelines suggested by Farhady et al. (1995; see appendix H) for marking ESL compositions on the 20-point scale with decimals of 0.25 which is the common marking scale in Iran. It was suggested that the raters should use the alternative marking system in Iran in which letter grades replace numbers. For example an A covers scores of 17-20 inclusive and a B means scores below 17 to 14 and so on.

The advantage of this alphabetical system over the numeric version of scoring is that it reduces the disagreement between the raters and helps to achieve a more uniformed set of scores and hence a higher degree of inter-rater reliability. The raters did not agree with the suggestion and maintained that they are quite comfortable with the numeric system of scoring and maintained they had developed a clear common-sense about these scores and the level of achievement each score typically signifies. Generally speaking, they used no profile. They just agreed on the criteria of what can be expected from a person as a reasonable ultimate performance. To have a more vivid and practical idea about such an achievement some of the best papers, according to the teachers of the classes, from each group were selected and analysed. Finally, there was some sort of agreement among the raters about the best possible performance they can expect from the population under study. These compositions received either 20 or 19.75 out of 20 and became the benchmark for the scoring and judging of other essays. Each rater received half of the essays, which were previously coded to avoid the use of names, and after finishing marking the first half they swapped them with the other half.

4.6.5 Reliability studies

The major concern with using totally impressionistic method for scoring writing has always been and still is the agreement between the raters arising from the criteria they consciously or sub-consciously use when judging the papers. In more technical terms the major excuse for not using impression marking is the concern over the alleged low inter-rater reliability. While, theoretically such worries may seem justifiable, in practice they do not appear to be realistic. This researcher, relying on the ample literature available on the holistic scoring and his own experience (Nemati, 1993), was quite

confident that this way of scoring, at least for the sake of the present research purposes, is quite reliable. To calculate the reliability of the scoring procedure a *Pearson-Product Moment* correlation test was used. The result was a promising 0.85 ($p<0.000$) which is by any standard (Jacobs et al., 1981) quite satisfactory for research purposes. This correlation was then adjusted by using the Spearman-Brown Prophecy Formula (Brown, 1996). The result (0.91) can be regarded as the reliability of the rating procedure.

4.6.6 Data analysis

The production phase of the study, finally, resulted in four sets of scores, one for each mode of discourse. The set of scores for argument is going to be tested against description and the scores for expository essays will be compared against the scores for narration. In other words two comparisons are needed and in each comparison two groups are being examined. As the subjects for the two opposing groups are the same, the related t-test is regarded as the most suitable statistical test. A more comprehensive discussion and the results of these comparisons will be presented in the next chapter.

4.7 The recognition phase

This phase of the study was mainly intended to examine the nature of the language produced in different discourse types. The production stage was to investigate how ESL learners produce a piece of linguistic discourse whereas this part of the research project was developed to compare and contrast the nature and features of the language already written, produced in different modes. It is necessary to point out that some nearly similar research has been done before (e.g. Reid, 1990) but these studies only

account for the linguistic similarities and differences among a limited number of topic types from a contrastive rhetoric perspective, whereas this present study mainly focuses on respondents, perceptions of the cognitive complexity involved in these linguistic genres. The previous studies just state whether or not any differences exist but do not demonstrate that more difficult texts to write are more demanding on the side of the reader, too. They do not graphically show that some discourse types need a higher expectancy grammar to be comprehended and, then, reconstructed. This study attempts to show this in a quite novel method.

4.7.1 The Variables

4.7.1.1 Independent variable

The independent variable of the recognition stage is the mode of discourse which is categorised into four modes namely: *Argumentation, Description Narration,* and *Explanation.*

4.7.1.2 Dependant variable

The score given to the learners' performances on each section of the cloze test is the dependant variable of the experiment. The range of the scores, due to the number of blanks in each passage, would from 0 to 15.

4.7.2 Method

To study this dimension of texts written in different modes of discourse a fairly novel method has been used. Four texts presumably different in nature were turned into an integrative test, known as a **cloze test**. To the best knowledge of this researcher this

aspect of differences inherent in the four major discourse types has never been put into trial. As **Cloze** is the key concept in the Recognition phase of the study, a general review of this type of language tests is given in the following section.

4.7.2.1 Cloze test; Definition and history

Deriving from the word closure, Wilson Taylor (1953), coined the term **cloze** to name a new type of language test. Closure itself is a human psychological phenomenon in which the mind tends to complete incomplete patterns e.g. a broken circle seems to be a complete circle at first sight, or as Oller (1979) explains when we look at a tree we tend to consider it as a single entity rather than a collection of leaves and branches. It is the second step when we pay attention to the components of that unit i.e. branches and leaves, etc. This can be explained by the fact that we perceive things as a complete and whole entity (or Gestalt) first and then we try to work out the missing bits and details.

Applying this principle to language, one can expect speakers of English to be able to reproduce accurately the missing parts of a sentence instead of geometric figures. This can be explained because in a cloze test the learner's **Expectancy grammar** would enable them to fill in the blank slots with appropriate language items. Expectancy grammar is a phenomenon which enables speakers of a language to predict the sequence of elements in any linguistic unit i.e. sounds, words, sentences and even larger units of discourse (Oller, 1979). This ability is not specific to L1 learners and L2 learners develop such a useful system the efficiency of which, of course, depends on the degree of their mastery over the second language. It is this degree of mastery which the cloze procedure is designed to test.

Therefore, a cloze unit can be defined, as Taylor (1953, p,416) did, as "any single occurrence of a successful attempt to reproduce accurately a part deleted from a message (any language product) by deciding from the context that remains, what the missing part should be".

Richards et al. (1992, p.55) while explaining that "in a **cloze test**, words are removed from a reading passage at regular intervals, leaving blanks" give an example of a cloze test:

"A passage used in ------------ cloze test is a -------------- of written material in ---------
words have been regularly -----------------. The subjects must then ------------ to reconstruct
the passage --------------- filling in the missing ------------." (p. 55)

4.7.2.1.1 Deletion patterns

In this example after every four given words one word is deleted. In other words, every fifth word is deleted and therefore the ratio of deletion for this passage is n=5. This way of systematic deletion of words is called the **fixed ratio method**. Naturally, by increasing or decreasing the *n* we can make a test easier or more difficult, respectively though Alderson (1980) reported that it is not possible to change the difficulty level of the test by providing more texts between the gaps beyond n=5. In other words a cloze test with n=12 is not significantly easier than a cloze test developed with the same text but based on n=6. In most versions of cloze tests commonly practised, every 7th word is deleted which can be regarded as the standard cloze test. However, for pedagogical purposes teachers usually prefer the other version of cloze test in which certain words are deliberately deleted to examine the learner's awareness of that particular point, irrespective of the interval of word deletion. The deletions can be words of a certain grammatical category such as verbs, adjectives or

even just function words. This method is known as the **variable ratio** or **rational deletion pattern** (Bachman, 1985; Cohen,1994).

Both methods have their own advantages and disadvantages. While the *fixed ratio method* keeps the integrative nature of the language inherent in a piece of prose text intact, the *variable ratio method* provides teachers with an opportunity to test whatever aspect of the language they want to examine among their learners. It can also be concluded that the former is more suitable for general language proficiency test but, the latter is more informative if used as an achievement test or as an awareness-raising or practice activity and thus, sometimes, referred to as **cloze task** rather than cloze test. In the present study, which mainly targets the subjects' language proficiency, an standard fixed ratio with (n = 7) has been implemented.

4.7.2.1.2 Scoring methods

There are mainly two methods of scoring cloze tests: The **exact word method** and the **acceptable word method**. In *exact word method* the testee gets the point only when the filled word is exactly the same as the word written in the original text. This method is arguably preferred for non-native speakers of English taking a cloze test to incorporate the original style of the writer in the test whereas the other method, the *acceptable word method*, is more recommended for native English speakers because native speakers, theoretically, have much better intuition to decide on the grammaticality and semantic appropriateness of an utterance. However, very high correlation estimates of 0.94 (Irvin, Atai and Oller; 1974) and 0.97 (Stubus and Tucker, 1974) between the two methods have been reported. While Gefen (1978) argues that it is too difficult to restore the exact missing word in a text and hence

favouring the acceptable words version, Porter (1978) maintains that the exact word version can discriminate between learners with different language proficiency levels more accurately.

To compromise between the these two versions of scoring procedures, Darnell (1968) suggested a very interesting but complicated method called **clozentropy**. In this method the test is given to a number of native speakers and then based on the received answers all acceptable alternatives are weighted according to the frequency of their occurrences in the answers. This frequency list will be the benchmark to score answers given by non-native participants. As is evident, this procedure is impractical due to its cumbersome preparation procedure.

The cloze test was initially developed to investigate and measure the readability of texts for native speakers of English (Brown, 1983) and later turned into a device to assess reading comprehension ability (Oller, 1983). The application of cloze to measure the difficulty of prose texts by Taylor (1953) was a reaction towards other flawed indexes used previously (e.g. Dale and Chall, 1948) which would analyse texts in isolation and mainly according to certain linguistic features such as sentence length. Cloze, instead, takes a variety of linguistic and non-linguistic factors incorporating text difficulty into consideration. The role of cloze in assessing reading comprehension has been verified by numerous researchers, e.g. Oller (1979), Cohen (1980), Alderson (1983) and Brown (1983), among others. Cloze tests also show very high correlation with major general language proficiency tests like TOEFL and IELTS (Kiany, 1995) and hence can be an acceptable substitute for them (Mullen, 1979).

Undoubtedly, cloze is one of the easiest tests to construct, administer and score. It boasts having the objectivity of multiple-choice tests and the communicative features

of integrative tests, simultaneously. Nevertheless, like other new methods, cloze tests have received diverse comments from researchers. While Alderson (1979) criticises cloze tests for not being sensitive to communicative abilities because of measuring only short-term grammatical constraints, Oller (1979) points out that such claims are incorrect. Besides, Chihara et al. (1979) and Jonz (1990) stress that reconstructing cloze items is not local and restricted to the adjacent words, rather in order for the subject to fill in the gaps correctly, they have to be aware of discourse flow over sentences and even paragraphs through higher order language processing. And finally supporting the methodology used for this project, Cohen (1980) believes that, among other requirements, a **textual knowledge** is necessary to fill in a cloze passage. The involvement of such awareness towards textual knowledge is of importance for the conclusions which will be drawn later in this study from the cloze results.

4.7.2.2 C-tests

Based on the fundamental principles of cloze tests, this relatively recent adaptation of cloze has been developed in Germany by Klein-Braley (1981, 1985, Klein-Braley and Raatz, 1984). Like cloze, in theory, it relies on reduced redundancy and expectancy grammar. Unlike cloze, in practice, in c-tests every second word (n=2) is deleted which makes the text very difficult to read and comprehend. In return, as a remedy, half of each deleted word is provided which works like a clue. In a C-test only exact words are considered correct and receive points.

Klein-Braley and Raatz (1984. p.144), the die hard advocates of the C-test, claim the following advantages for a C-test:

1. it is easy to construct and to score C-tests;

2. adult native speakers should obtain virtually perfect scores;

3. the deletions affect a representative sample of the text;

4. even previously untried material produces satisfactory reliability and validity coefficients; and

5. C-tests hace face validity.

Jafarpur (1995, p.195) investigating these claims, only confirms the ease of constructing and scoring C-tests and refutes the rest (p.209).

In a perhaps more fair approach, Weir (1990) lists four advantages and two disadvantages for C-tests:

advantages

1. Compared to cloze, in c-tests a larger number of words are deleted (normally 100) which are a good representation of the selected passage.

2. The scoring procedure is more objective since half of the word is provided and there is very little probability for confusion.

3. Unlike cloze tests, native speakers perform more uniformly on c-tests.

4. C-tests are economical and highly reliable

Disadvantages

1. "There is little empirical evidence of its value" (p.49)

2. As referred above, deleting every second word makes the text "heavily mutilated" and affects its face validity.

Weir (ibid) also refers to a version of this technique developed by Davies (1965) where he provides only the first letter of the deleted words. There is, therefore, some precedent for using this technique.

Considering all these arguments, it seems that deleting every second or third word may frustrate the learners who take the test and it is better to consider a larger n for deletion pattern. Besides, providing part of the missing word can improve the quality of the test and increase the objectivity of the scoring procedure and that is exactly the procedure adopted for this phase of the study.

4.7.2.3 Subjects

The 100 subjects of the Recognition phase were chosen from three universities located in Tehran: Tehran University (Faculty of Foreign Languages), Open University (Northern branch), and Open University (Southern Branch). The profile of the subjects is given in Table 4.5:

Table 4.5 : The profile of subjects taking part in the recognition phase of the study

University	Number (N)	sex (M)	(F)	level (Under)	(Post)
Tehran University	40	25	15	32	8
Open University (North)	26	0	26	26	0
Open University (South)	34	34	0	34	0
Total	100	59	41	92	8

158

4.7.2.4 Materials and instrumentation

In this phase of the study, four texts, each in a different discourse type, had to be compared. It would be very difficult, if not impossible, to find four parallel texts, equal to each other in every aspect but type of discourse. Even readability formulae cannot prove such equity. Alternatively, it was decided to use four texts written by the same person, and therefore, originated from the same level of linguistic competence, so that if any difference exists, it should be a genuine one resulting from the different prompts. An Iranian research student of English literature with a TOEFL score of 600 and an IELTS grade of band 6.5 was voluntarily assigned to produce the texts. He was not told about the research project and its underlying assumptions to keep him impartial towards different essay topics. All the aforementioned eight prompts were offered to him in pairs so that he had a choice to pick his favourite topic from each mode. He chose the following topics:

Argumentative prompt #2: "At the present time, the human race spends huge sums of money to explore space. Do you find it a waste or do you view space projects as essential to expanding our knowledge? Discuss the issue.

Descriptive prompt #2: "Describe your hometown or any interesting city so that readers can have a clear image of the city in their minds."

Narrative prompt #1: "Write the story of the most interesting film you have ever seen. Narrate the events as they happened in the film. Your writing should look like a story.

Explanatory prompt #2: "Considering the situation in Iran, explain how a foreign language should be learned?"

Once again to simulate the test environment and to provide equal conditions for all compositions, he was given only 30 minutes per essay. Besides, to create the test pressure and also make him give his best performance, he was told that these essays are of vital importance to a research project. However, this was clearly not exactly as serious as a real test. Besides, whatever the situation was, it was practically almost the same for the four different types of topic which are going to be compared and contrasted against each other. Therefore, any possible difference between a real test and the experiment would not affect the internal validity of the experiment.

For this particular cloze test, for the recognition phase of the study, a modified version of cloze test was used, similar to the one developed by Davies (1965), in which he provided only the first letter of each missing word. The rationale for choosing this method is that:

(1) it ensures more correct responses from the weaker students and this can prevent the accumulation of scores at the bottom which can cause the ceiling effect. The ceiling effect refers to situations in which tests are too easy or too difficult and the scores are very similar to each other. Considering the fact that reliability, as discussed in chapter two of this thesis, depends on the variation between the scores, the adopted method could help to achieve a more normally distributed scores and hence increase the reliability of the test.

(2) This version of cloze almost solves the discrepancy between the two scoring models i.e. The *exact word method* and the *acceptable word method*. As mentioned earlier, in the former approach some alternative but correct answers are considered wrong and receive no points. And in the latter the scorers may face so many alternative words with different degrees of correctness that can hinder the scoring procedure and affect the test results. As a solution, in this method scorers can avoid the shortcomings of the former without being confused with a wide range of acceptable answers for one single gap. After all, the choices were quite limited because of the specified first letter and eventually could not exceed two or in rare cases three realistic alternatives.

The written texts, then, were turned into standard cloze tests. In other words, the very first and last sentences of each passage were left intact and in all other sentences every 7th word was deleted. In order to limit the choice of words fitting correctly into the blanks, the first letter of each missing word was provided.

It was impossible to use the whole written composition because it would be too long, tedious and time taking which could affect the test. So, for each essay a passage comprising 15 blanks was chosen. Therefore, the final test had 60 missing words to be filled within 40 minutes. The test administration supervisors divided the allotted time into four equal intervals of 10 minutes and made the students move to the next passage at the end of each ten minutes. In this way, an even distribution of time and effort among the four passages was achieved.

4.7.2.5 Scoring procedure

As discussed before, there are mainly two methods of scoring cloze tests: The *exact word method* and the *acceptable word method*. In exact word method the testee gets the point only when the filled word is the same as the word written in the original text. In the other method, *the acceptable word method*, all the words that can fill the gap correctly will receive the point. For the reasons stated above, the latter method was chosen for scoring the cloze tests.

Subjects, after the scoring procedure, received four sets of scores, each representing their performances in reconstructing written language of a certain mode of discourse. A copy of the cloze test used in this study is provided in Appendix B.

4.7.2.6 Data analysis

Due to the design of the recognition phase, four sets are scores are collected, each representing a certain mode of discourse. As there are more than two groups being compared and the subjects of the four groups are identical, a repeated measures ANOVA is chosen as the appropriate statistical test to confirm whether the differences between these scores are statistically significant or not. The results are presented in the next chapter.

4.8 The preference and attitude phase

One of the areas in the field of evaluating writing which has drawn some attention towards itself is the issue of how the learner writers choose a topic. Despite this attention it seems that "previous quantitative attempts to examine factors affecting students choice have been inconclusive" (Polio and Glew, 1996, p.35). They further

hold that one of the reasons that may have hindered researchers in finding the factors behind the prompts that may make some topics more appealing over others is that "most of the research on the writing process examines, and not appropriately, the overall writing process from beginning to the end" (ibid). So the very first step in writing which is choosing a topic has not been studied properly on its own.

An outstanding qualitative study, as already cited, has been done by Polio and Glew (1996). They concluded that several factors affect the students' choice of topic including "their own background, knowledge, question type, and specificity of the topic" (p.35). Although they did not focus on the effects of the discourse mode it was observed that a personal narrative prompt was chosen because the subject could write immediately and faster as it does not require the writer to plan the essay before starting the actual writing. This is yet another signal that mode of discourse can make a difference in learners' choice of topic and, presumably, in the quality of their writing.

Another instance of such studies is that of Chiste and O'Shea (1988) in which they found that ESL students when offered a choice of topics, unlike native speakers, tend to choose shorter topics containing fewer words. However, as the shorter topics were placed at the top of the list it makes it look possible that the subjects were influenced by the order of the prompts in the list, a pitfall which has been avoided in the present study.

Probably the most frequently cited research of this type is that of Hamp-Lyons and Mathias (1991). They asked two trained MELAB composition readers and two ESL writing experts to rate 64 MELAB topics (administered in the period 1985-1989) on a three-point scale: easy, average difficulty, and hard. In the next step, using the

categories developed by Bridgeman and Carlson (1983) they classified the topics into

five task type categories:

1.expository/private

2.expository/public

3. argumentative/private

4. argumentative/public

5. combination

According to the experts' judgements the argumentative/public was considered as the

most difficult type of topic and the expository/private as the least difficult type.

However, when they tried to find the probable existing relations between the scores

and the rating and the topic type they realised that "mean writing score increased

rather than decreased, as topic difficulty increased, except for topics in the group

judged as most difficult" (Hamp-Lyons and Prochnow, 1991, p.62).

There are two major differences between the present study and this one. First, in the

present study it is the learners, the real world writers and test-takers who judge the

difficulty of the topic. Secondly, these researchers also admit that three types of topics

out of nine listed by Bridgeman and Carlson (1983) were not included in their study

and only half of the data set could be fit in the remaining six categories the

classification of the topics, whereas the present study is more comprehensive and

includes all four basic types of discourse including description and narration.

4.8.1 Present study: Variables

4.8.1.1 Independent variable: the independent variable is the prompt offered to the

students for rating. There are eight different prompts which comprise four different

discourse types. In other words, there are two prompts for each mode of discourse. This is to make sure that topics of a certain discourse type behave in the same way.

4.8.1.2 Dependant variable: The dependant variable is the score each topic receives according to the degree to which students favour that certain topic. These scores are based on a Likert scale and hence vary from 1 to 5.

4.8.2 Method

The data for this part of the study has been collected through a questionnaire. In this questionnaire a list of all eight composition prompts, used in the production phase, was offered to 87 Iranian university students, mainly from Tehran University, Open University and Tarbiat Modarres University, to rate them according their preference on a five-point Likert scale. The students were supposed to read the topics and according to the degree that they prefer to write on those topics in a writing test rate them from one to five, one signifying the least and five signifying the most favourite topic. The eight topics were arranged in a certain way to avoid the pitfall that is associated, as mentioned earlier, with the study done by Chiste and O'Shea (1988). In this questionnaire the topics were listed in the following order:

1. **Argumentative topic #1**

2. **Descriptive topic #1**

3. **Narrative topic #1**

4. **Explanatory topic #1**

5. **Explanatory topic #2**

6. **Narrative topic #2**

7. **Descriptive topic #2**

8. **Argumentative topic #2**

As it is shown in the list, there are two topics for each mode and the order for the second set of the topics (from explanatory to argumentative) is the reverse of the first one (from Argumentative to Explanatory). Appendix C includes a copy of this questionnaire.

4.8.3 Data analysis

After gathering the necessary data, the next step was to find out if there were any statistical significant differences among the topics and also topic types. Since there are eight topics and four topic types and the subjects of these groups are the same, therefore the groups related and dependant and hence a repeated measures ANOVA has been applied to the data.

The results of these analyses are presented in the *"preference and attitude"* section of chapter 5.

CHAPTER FIVE

RESULTS AND DISCUSSIONS

"Averages and relationships and trends and graphs are not always what they seem. There may be more in them than meets the eye, and there may be a good deal less."

Huff (1954, p. 10)

Introduction

In the previous chapters, for the convenience of readers, the research study was divided into three phases namely: *production, recognition,* and *preference.* Similarly, in this chapter the results are presented in three corresponding sections. Therefore, in the first section of this chapter the statistical results related to the compositions written by Iranian university students are presented. In the second section, the results pertaining to the analysis of cloze tests derived from four types of essays are given and finally in the third section of this chapter, readers can find the EFL learners' attitudes towards different essay prompts aimed to elicit different types of discourse. After giving the results of statistical analyses for each section, the consequences and implications of the results are immediately discussed.

The level of significance for this study, like almost all other studies in applied linguistics, as recommended by Hatch and Farhady (1982) and Heys et al. (1993) has been set to be $p < 0.05$. This means that there is only a 5% probability that the results occurred due to chance and hence can be interpreted with a certainty of 95%. This has been indicated by an asterisk (*). This is the minimum level of significance to reject a null hypothesis. It often happens that a more trustworthy result is achieved at a significance level of $p < 0.01$. This is indicated by two asterisks (**). And three asterisks (***) indicate a significance at the level of $p < 0.001$.

5.2 SECTION 1: Production

The results for the Production phase of the study include all the statistical findings related to the subjects' compositions. In regard to the discourse types, two comparisons were made: Argument versus description, and narrative versus explanation. Each of the two groups were actually compared in four separate sub-groups with four different sets of topics. The following tables (5.1, 5.2, 5.3, and 5.4) are the descriptive statistics for these sub-groups separately:

Table 5.1 Descriptive statistics for the first sub-group of Arg. vs. Des.

	Mean	Std Dev	Minimum	Maximum	N
Arg. Topic No. 2	15.48	2.95	9.00	19.50	28
Des. Topic No. 2	15.65	2.92	8.00	19.75	28

The subjects of this subgroup (shown in table 5.1) were selected from postgraduate students of TEFL in Iran. As the mean of these composition scores indicate, this group are by far more proficient than other subgroups. In fact, selected essays from this subgroup were the benchmark for comparison and scoring. The mean score for the descriptive essay is higher than that of the argumentative one.

Table 5.2 Descriptive statistics for the second sub-group of Arg. vs. Des.

	Mean	Std Dev	Minimum	Maximum	N
Arg. Topic No. 2	9.66	3.58	4.00	18.00	22
Des. topic No. 1	10.06	3.35	5.00	17.50	22

Table 5.3 Descriptive statistics for the third sub-group of Arg. vs. Des

	Mean	Std Dev	Minimum	Maximum	N
Arg. Topic No. 1	9.62	4.28	4.25	19.25	25
Des. Topic No.1	10.19	4.11	3.50	18.75	25

These two subgroups (shown in tables 5.2, and 5.3) were drawn from sophomores majoring in English language and literature. The mean scores of these subgroups are much lower than the previous one which is quite normal and predictable. However, the same pattern exists here, too. The mean score for description is obviously higher than the mean score for argumentation.

Table 5.4 Descriptive statistics for the fourth sub-group of Arg. vs. Des

	Mean	Std Dev	Minimum	Maximum	N
Arg. Topic No. 1	6.43	2.66	3.00	13.25	22
Des. Topic No. 2	7.10	2.75	3.25	15.00	22

The subjects of this sub-group (shown in table 5.4) were selected from sophomores majoring in English language at Open University in Iran. The mean score for this sub-group is the lowest among the four sub-groups who wrote essays on both argumentative and descriptive topics.

There is an interesting point here which is worth mentioning, though not directly related to this study, which concerns the lower level of education at the Open University compared to the state universities. This university has been the target of bitter criticism from several academics for admitting unqualified candidates simply for financial reasons. The Open University does not receive any budgetary assistance from the government and therefore, allegedly, admits too many applicants to meet its expenses. This is, of course, just a minute piece of evidence supporting those who oppose the policy of the Open University and it needs more substantiation.

Nevertheless, quite interestingly, in this sub-group, too, argumentative writing achieved a lower mean score than descriptive topics did. Observing the same pattern in a wide range of subjects increases the generalisability of the results. If the difference between the essays written on descriptive topics and those essays written on

argumentative topics are statistically significant then one can claim that this difference does not change with language proficiency improvement.

In order to investigate whether or not the apparent mean difference between argumentative topics and descriptive topics the four subgroups were aggregated into a main group comprising two hypothetically contrasting discourse modes of description and argument. Probably, a panoramic view of this main group might give a better understanding of the situation. Table 5.5 shows the descriptive statistics for the main group of *argument versus description* after being aggregated:

Table5.5 Descriptive statistics for the main group of Argument vs. Description

	Mean	Std Dev	Minimum	Maximum	N
Argument	10.60	4.77	3.00	19.50	97
Description	11.04	4.57	3.25	19.75	97

The next step is to put these raw numbers into a statistical analysis to ascertain whether this difference is significantly meaningful and not incidental.

5.2.1 Investigating the first null hypothesis

It was mentioned earlier in chapter four that two null hypothesis were generated from the first research question. To remind the readers, the first research question, for which this phase of study tries to find an answer, the related hypotheses are repeated here again:

Research question #1: Are some modes of discourse (i.e. *argument/explanation*) relatively more difficult to write on by EFL learners compared against other modes *(i.e. narration/description)*?

Regarding this question and the classification of discourse modes the following null hypotheses were proposed:

H_0 #1: There is no statistically significant difference between the scores of compositions written by the same EFL subjects in two different discourse modes of argument and description.

H_0 #2: There is no statistically significant difference between the scores of compositions written by the same EFL subjects in two different discourse modes of explanation and narration.

A related t-test was run to investigate any possible statistical differences between the two types of compositions (i.e. description and argument) to answer the first research question. The result is presented in table 5.6:

Paired Differences

Table 5.6 The results of related t-test for Argument vs. Description

Mean	SD	SE of Mean	t-value	df	2-tail Sig
-.4381	1.551	.157	-2.78	96	.006**
95% CI (-.751, -.126)					

172

As shown in the table there is a highly significant difference (P< 0.01**)   between argument and description essay scores written by the same subjects. Therefore the first null hypothesis is rejected and the experimental hypothesis is accepted. The conclusion is that descriptive topics, compared to argumentative topics, seem to be much easier for EFL learners to write on.

5.2.2 Investigating the second null hypothesis

The second null hypothesis was next investigated. This was derived from the first research question which, as stated above, assumes that there is no significant difference between expository and narrative topics. The following is the descriptive statistics about four sub-groups who performed on explanation and narrative:

Table 5.7 Descriptive statistics for the first sub-group of Exp. vs. Nar.

	Mean	Std Dev	Minimum	Maximum	N
Exp. Topic No. 1	8.20	3.07	2.50	13.50	21
Nar. Topic No. 2	8.25	3.29	3.00	13.75	21

In this subgroup which comprises 21 students from Open University, Southern Tehran branch, the maximum score is quite low compared to Argument/Description main group. Two reasons may be advocated to explain this situation. First, the students are from the Open University and the related problems discussed earlier might have lowered the level of scores. Second and more importantly, no postgraduates were

included in this group and that is a major reason for such differences. The mean score

for narration in this sub-group is slightly higher than mean score for explanation.

Table 5.8 Descriptive statistics for the second sub-group of Exp. vs. Nar.

	Mean	Std dev	Minimum	Maximum	N
Nar. Topic No. 2	6.28	2.46	3.00	11.50	17
Exp. Topic No. 2	7.72	2.64	3.00	13.00	17

The mean score as well as the maximum score for narrative mode in this sub-group fall

much lower than corresponding explanatory ones. This is not compatible with the

assumptions about the difficulty of these two modes. The suggested reason for such

behaviour will be discussed later.

Table 5.9 Descriptive statistics for the third sub-group of Exp. vs. Nar.

	Mean	Std Dev	Minimum	Maximum	N
Exp. Topic No. 1	7.35	3.20	2.00	13.75	20
Nar. Topic No. 1	7.37	2.83	3.00	12.75	20

Table 5.10 Descriptive statistics for the fourth sub-group of Exp. vs. Nar.

	Mean	Std Dev	Minimum	Maximum	N
Nar. Topic No. 1	8.48	3.28	2.75	13.50	22
Exp. Topic No. 2	8.49	3.13	3.25	12.50	22

These four sub-groups have to be integrated as a large group to compare two different discourse modes.

The performance of subjects in these two subgroups seems to be very much similar and the difference between the mean scores is either 0.01 or 0.02. Logically, there will be hardly any statistically significant differences between the two modes. Again, in order to investigate the probable difference between the subjects' performances in the two hypothetically contrasting discourse modes of narration and explanation these sub-groups were incorporated into a main group. Table 5.11, which shows the aggregated results, helps to show a more informative view of the subjects' performance on the two modes of explanation and narration:

Table 5.11 Aggregated descriptive statistics for the main group of Explanation vs. Narrative

	Mean	Std.Dev	Minimum	Maximum	N
Narrative	7.67	3.08	2.75	13.75	80
Explanation	7.97	3.01	2.00	13.75	80

As the table shows the overall performance on expository topics, quite contradictory to the researcher's assumptions, were better than narrative topics. Before drawing any

conclusions, the difference between these two modes has to be confirmed through statistical analysis.

Based on the rationale stated for the application of matched t-test for Argument/Description group, a similar test has been run for Narration/exposition group, too. The result of the test is given in Table 5.12.

Table 5.12 The results of related t-test for Explanation vs. Narrative

Paired Differences

Mean	SD	SE of Mean	t-value	df	2-tail Sig
.2908	1.520	.170	1.71	79	.091

95% CI (-.047, .629)

Although there is a trend here in the direction which would be needed to reject the null hypothesis, the results do not indicate a difference large enough to reject the second null hypothesis. It can be concluded that narratives and explanatory compositions do not meaningfully differ from each other at the production level.

5.2.3 Discussion

This phase of the study is supposed to answer this question:

Research question #1: Are some modes of discourse (i.e. *argument/explanation*) relatively more difficult to write on by EFL learners compared against other modes (i.e. *narration/description*)?

Due to the design of the study two null hypotheses stemmed from this question. The statistical analyses have indicated a difference between argument and description but not for narrative and exposition. It might be reassuring that a Wilcoxon matched paired, which is a non-parametric counterpart of the matched t-test, was also run which confirmed the present results for both comparisons. As table 5.13 shows, the two-tailed probability for paired samples for the Argument vs. Description is 0.003** which is even slightly stronger and more significant than that of the t-test.

Table 5.13   The results of Wilcoxon Signed-Ranks Test for Argument vs. Description

Mean Rank	Cases
47.33	26 - Ranks (Des. LT Arg.)
42.58	61 + Ranks (Des. GT Arg.)
	10 Ties (Des. EQ Arg.)
	97 Total
Z = -2.8928	2-Tailed P = .0038**

Similarly, for the other comparison, Narrative vs. Explanation, the Wilcoxon test result in table 5.14 shows 0.08, which although indicating of a trend, is still not significant at $p < 0.05$.

Table 5.14 The results of Wilcoxon Signed-Ranks Test for Explanation vs. Narrative

Mean Rank	Cases
37.38	40 - Ranks (Nar. LT Exp.)
31.72	29 + Ranks (Nar. GT Exp.)
	11 Ties (Nar. EQ Exp.)
	80 Total

$Z = -1.7189$ 2-Tailed P = .0856

The major difference between a non-parametric *Wilcoxon test* and a parametric paired or related *t-test* is that some of the data will be lost in Wilcoxon test as it just takes the ranks into consideration rather than the scores. For this reason the t-test is considered to be more robust but this does not mean that the results would necessarily differ. The difference may happen when the significance is marginal. With regard to the first research question the first null hypothesis is rejected and the second null hypothesis is accepted by both tests.

The outcome of the production phase can be stated as the two following experimental hypotheses which can be accepted:

Experimental Hypotheses:

Hypothesis #1: _THERE IS_ a statistically significant difference between the scores of compositions written by the same EFL subjects in two different discourse modes of argument and description.

Hypothesis #2: _THERE IS NO_ statistically significant difference between the scores of compositions written by the same EFL subjects in two different discourse modes of explanation and narration.

In other words, the experiment confirms that argumentative topics are relatively more difficult than descriptive ones. On the other hand, there is no significant indication that expository and narrative topics differ from each other.

The first interesting point derived from these results is that argument, compared against description, stands out as a more difficult mode of discourse even in a foreign language situation. This is quite contradictory to the researcher's prediction, discussed earlier, that at tertiary level because of enough exposure and cognitive development, learners would not have serious problems with argumentative writing. The results do not confirm this idea. Therefore, it can be concluded that writing an argument is relatively more demanding even for critically-developed minds of students at tertiary level.

Considering the lack of significant difference between narration and exposition, two tentative reasons come to the mind. First, it seems that exposition is too broad a category. In other studies like Hamp-Lyons and Mathias (1994) exposition also behaved in a strange and confusing way, at least quite contradictory to the researchers'

prediction. They were expecting expository/private topics to be the easiest but this was not the case and in order to justify such a strange observation they doubt their "colleagues' claims of possession of the knowledge and skills necessary to make judgements" (p.52). Although this cautious approach towards the expert's judgement is prudent and is emphasised in the next paragraph, it seems that such a reassessment is needed towards the present classification as well. In this study expository writing was assumed to be more difficult than narration and the result was again not in line with the expectations. This sort of mystifying behaviour has never been reported about argument or narration in L1. While argument and narration are hardly divided into sub-categories, exposition can include a range of rhetorical features which may have different bearing on the difficulty level of the topic. Quite often in the literature, exposition is broken down into smaller classifications. Hale et al. (1996), for example, distinguish four modes of discourse among academic tasks namely **Narration, Description, Exposition** and **Argument**. In their prestigious study, which is a recent one concerning the classification of writing tasks for TOEFL, they define each of these modes in just one paragraph except for exposition which, in two detailed pages, different patterns of it have been accounted for. For example an analysis, a cause/effect or a comparison/contrast relationship, and a classification are all regarded as expository writings. There is no guarantee that these patterns are of the same difficulty level; in fact, it is possible that they are not. It seems that similar research studies comparing different patterns of exposition are warranted to shed some light on the issue.

Second, one of the narrative topics was not quite familiar for Iranian students. This made narration look a bit more difficult than it could be otherwise. The selection of

this topic was actually an over-generalisation of Anglo-American academic culture into the Iranian academic situation. Topics like "the biography of a person who rose from a humble background" seem like a typical western composition subject. If this is the case, then it can be concluded that other variables, such as subject matter familiarity can override the effect of mode of discourse. This is an alarming finding for topic developers at major test batteries. This also echoes Hamp-Lyons and Mathias' (1994) warning that we cannot and should not merely rely on the expert judgement to decide on the difficulty level of topics.

The production stage of the study has shed light on the issue of topic difficulty regarding Argument vs. Description and Narration vs. Exposition but there is little here to tell us about the status of argument against narration and exposition or exposition against description. The next phase of study can be of more help since it compares all four discourse types against one another.

5.3 SECTION 2: Recognition and Reconstruction

The second part of the study is mainly concerned with the nature of texts produced in four different discourse modes and learners' cognitive engagement with these texts. In the first stage of the study probable differences in the difficulty level of writing in these modes were studied but in this phase the already produced texts have been further tested to see whether or not these texts are of different degrees of linguistic difficulty and hence have different bearings on the mind of readers or not. The question, for which the recognition and reconstruction phase of the study tries to find an answer, is reiterated here again:

Research question #2: Do the essays written in different discourse types bear different degrees of cognitive and/or linguistic difficulty, when turned into cloze tests, to recognise and reconstruct?

The corresponding null hypothesis, the third hypothesis investigated in this study, then maintains that:

H$_0$ #3: There is no statistically significant difference between the learners' scores obtained from four cloze tests derived from four texts each written in a different major discourse type.

As mentioned before, four different texts written by the same person under similar conditions were turned into modified cloze tests. These tests were later given to a number of Iranian students.

The first step would be calculating the mean score for the four sections of the cloze test each representing a different discourse mode. Table 5.15 presents the descriptive statistics reflecting the performance of subjects on four sections of the cloze test. Each section or paragraph was written in a different discourse mode.

Table 5.15 Descriptive statistics for the four different sections of the cloze test

	Mean	Std. Dev.	Minimum	Maximum	N
Expository Passage	8.32	2.68	1.00	14.00	100
Narrative passage	11.87	1.87	5.00	15.00	100
Descriptive passage	7.49	2.82	1.00	14.00	100
Argumentativ e passage	6.75	2.68	1.00	13.00	100

As shown in the table the three modes of argument, description and narration are, as expected, in order of difficulty level with argument as the most difficult text and narrative as the easiest one. The mean score for narrative (11.87 out of 15) is outstandingly high and a smaller standard deviation indicates a relatively unified performance from the subjects. The mean score for exposition falls between the means for narrative and description, higher than the former and lower than the latter.

Despite this obvious difference among the mean scores of different sections of the cloze test, it was necessary to confirm the findings through statistical analysis. To do so a repeated measures ANOVA was selected to test the hypothesis, the result of which is given in table 5.16.

Table 5.16 Repeated measures ANOVA for the variable Cloze Test by variable Mode of discourse

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	841.18	297	2.83		
TEXTTYPE	1542.57	3	514.19	181.55	.000

As the results in table 5.16 show, there is a highly significant difference (p<0.000) among the mean scores of four passages of the cloze test.

The next step is to determine where the difference lies, or to put it in a simpler way, which two of these modes are significantly different from each other. The most common way of achieving this is to compare each group with every other through a multiple comparisons test. Among the available multiple comparisons tests (e.g. the Tukey and the Sheffe tests) the Tukey test is "possibly more sensitive . . . [and] does tend to give a more accurate Type I experimentwise error rate" (Youngman, 1979, p.84). Youngman (ibid) also strongly stresses the need for a sample of at least 10 members. This test is usually intended for use with equal-sized groups. Considering these conditions and recommendations, a Tukey test with a significance level of p<0.05 was run to locate the existing differences among the groups. Table 5.17 presents the results.

Table 5.17 The Tukey Test for different types of texts used in the cloze

Mean	MODE	Grp 4	Grp 3	Grp 1	Grp 2
6.7500	Grp 4				
7.4900	Grp 3	*			
8.3200	Grp 1	*	*		
11.8700	Grp 2	*	*	*	

(*) *Indicates significant differences*

Group1= Exposition
Group2= Narration
Group3= Description
Group4= Argumentation

As table 5.17 shows, all four groups are significantly different from each other. Once again, the argumentative mode of writing, with a mean score of 6.7 proved to be the most difficult text, while narrative writing, compared to other discourse types, achieved a relatively high mean score of 11.8 and therefore is, by far, the easiest.

5.3.1 Peripheral findings

In a further investigation to make sure that texts written by the same person in different discourse types are of a different nature and do indeed possess different levels of difficulty, all four passages previously used as cloze tests were examined to determine their readability estimate. The calculation of readability estimates can help us to ascertain that the observed cloze scores are accurate and whether some of these texts are linguistically more difficult the others or vice versa.

To calculate the readability of the texts four different formulae were used: *Flesch Reading Ease, Flesch-Kincaid Grade Level, Coleman-Liau Grade Level,* and *Bormuth Grade Level.* These formulae are already installed on Word for Windows version 6 and the computer is able to calculate the readability of any selected text. The Flesch Reading Ease index is based on the average number of syllables per word and the average number of words per sentence (Microsoft Online Help Manual). The range of Flesch reading Ease is from zero to one hundred and the standard writing averages approximately 60-70. The Flesch-Kincaid grade level similarly takes the average number of syllables per word and the average number of words per sentence into account but the outcome is presented as a grade-school level. A seven to eight grade level would be considered as the standard text. Both Coleman-Liau Grade Level and

Burmuth Grade Level compute readability based on word length in characters and sentence length in words. They present the estimate as a grade level.

It is necessary to point out that for the first formula, the Flesch Reading Ease, a higher estimate indicates that the text is relatively easier, whereas for the other three readability formulae a higher estimate means that the text is more difficult and, of course, more suitable for a higher grade level. Table 5.18 presents the results for the readability estimates.

Table 5.18 Readability estimates for different discourse types used in the cloze

	Argument	Explanation	Description	Narration
Flesch Reading Ease	46.0	55.3	60.8	77.9
Flesch-Kincaid Grade Level	14.0	10.4	12.2	6.7
Coleman-Liau Grade Level	10.8	12.0	9.3	8.0
Bormuth Grade Level	10.3	10.6	10.0	9.5

Once again, all three discourse modes of Argument, Description and Narration behaved systematically and were ranked in order of difficulty except explanation which did not show a regular pattern. In this analysis all four formulas, except Flesch-Kincaid which considers Explanation slightly easier than Description, and Coleman-Liau Grade Level which put explanation at the top of Argument, consistently ranked argument as

the most difficult text, then Expository, then Description and finally Narration, for their readability level.

5.3.2 Discussion

The existence of a difference among discourse types means that the third null hypothesis of the study must be rejected and an alternative experimental hypothesis replace it. Consequently the following hypothesis is accepted:

Hypothesis #3: _THERE IS_ a statistically significant difference between the learners' scores obtained from four cloze tests derived from four texts each written in a different major discourse type.

It is obviously clear now that a cloze test derived from, for example, an argumentative text would render a lower mean score compared to a narrative text given to the same subjects. What is it, then, that causes this difference? With regard to the application of cloze test, Rye (1982) writes:

> ". . .filling in cloze deletions is not about 'seeing' patterns in the sense of seeing the patterns of a circle. Cloze procedure is essentially a _cognitive_ task. The reader has to reason and construct suggestions to fill the gap on the basis of the evidence derived from the context. It is true that there are grammatical 'patterns' in language. The reader's innate ability to produce grammatically appropriate sentences will help determine the grammatical class of the word to be produced. However, the completion of meaning, based on understanding and reasoning, is a cognitive task." (p.3)

The findings of the recognition and reconstruction phase confirm that the texts produced in different modes of discourse would be of a different nature due to two parameters. First, it seems that discourse modes like narration and description are not linguistically as difficult as argument. Of course the present study does not focus on certain traditional indicators of linguistic complexity, such as the length of T-unit, but the cloze test experiment has revealed that, as Rye (1982) admits, grammatical patterns and semantic features of the text affect performance on cloze tests. Differences in linguistic complexity can be arguably taken responsible for existing differences among the mean scores observed for different discourse types. Second, as Rye (1982) has proposed, the cloze procedure is primarily a cognitive task. When a certain type of text is cognitively more demanding then a poorer performance and a lower score naturally would be expected. This fact prevents the researcher from concluding, falsely, that argument is difficult because students are not trained to write arguments or narration is easier simply because they have been exposed to it from early stages of schooling. Arguably, this is not a matter of training and linguistic development. It is rather a matter of cognitive engagement.

With some relation of the criteria given by Rye (1982), it is possible to classify texts written in different modes of discourse into the three categories of frustration level, instruction level and independent level. According to the criteria, when a learner fills only up to 40% (in this case 6 words) of the missing words in a text turned into a cloze test then that text would be regarded as being at the frustration level. This means that even in the presence of a teacher, the learner would not be able to read that text properly without an enormous amount of help. A score of 40 to 60 percent (6-9 words for this study) shows that the text is at instruction level which means that it is suitable

to be taught in the classroom where the teacher's help is available. For a text to fall into the independent level, it has to receive 80% or more of the correct answers.

The Argumentative text used in this phase of study received an average of 6.75. The number of missing words in the text were 15. Considering the fact that 40% of 15 equals 6, then we can loosely assume that this text is at the shady area between the frustration and instructional level for these particular students. Descriptive and Expository texts received an average of 7.49 and 8.32, respectively. This means that both of them are at the instruction level. For the Narrative text, 11.87 percent of the missing words were correctly restored which shows that this text is at the independent level for the subjects who took the cloze test. As mentioned earlier, this classification is not a well-grounded one but it can roughly illustrate the difference of difficulty levels among the texts written in the four modes of discourse.

Another point worth mentioning here is related to the peripheral findings of the recognition and reconstruction phase. There has been a line of research to find measurable linguistic differences among compositions written on different topic types (e.g., Crowhurst and Piche, 1979; Freedman and Pringle, 1980; Bridgeman and Carlson, 1983; among others). These linguistic differences are usually either syntactic (e.g. average sentence length, passive voice verbs, complex sentences, etc.) or lexical (e.g. average word length, percentage of content words, percentage of pronouns, etc.). Probably one of the most outstanding studies of research of this type has been carried out by Reid (1990), in which, using a computer text-analysis programme called the Writer's workbench (WWB), she found that texts written in two different types *of comparison/contrast* and *description* were different at lexical level but no significant differences were observed at syntactic level.

The present study, while following the same principle of searching for objectively measurable differences in texts of different modes, uses a new, objective and easy way to demonstrate that different topics yield different texts. An advantage of using a readability formula over the method used by Reid (1990) is that it gives a final judgement for the difficulty of the produced text, while Reid's more detailed method sometimes may add to the present ambiguity. For instance, she found that average word length and the percentage of pronouns in descriptive text was bigger than that of the comparison/contrast text. On the other hand, she found a larger percentage of content words for comparison/contrast text. With this data, it would be very difficult to conclude which topic leads to a more complicated text. The findings of the present study demonstrate that these four formulae, more or less, yield a consistent grading of texts and their originating topics.

In sum, it can be concluded that, as the results of this phase show, some modes of discourse are more difficult to comprehend and reconstruct. And therefore, constructing them in the first place (i.e. writing the original texts) may have been more difficult tasks.

5.4 SECTION 3: Preferences & Attitudes

The third section of this research study was designed to investigate the EFL learners' preferences toward different topic types. So far, it has been shown that the modes of discourse do have a bearing on the writer's performance. It has also been shown that texts produced in these modes differ from each other with regard to linguistic and/or cognitive difficulty. This phase of the study tries to find out whether or not EFL learners really know, consciously or sub-consciously, that writing on a certain topic is

more demanding and therefore more unlikely to achieve a good mark than another one. And if they are aware of differences, what is their impression of these discourse-type-based differences and how do they rank related essay prompts? Is there any relationship between their ranking and the type of discourse to which the prompts belong ?

As a reminder, the related research question and the corresponding hypothesis are stated once more:

Research Question #3: Do EFL writers prefer certain mode(s) of discourse over other mode(s) while choosing a topic to write on?

Consequently, the fourth null hypothesis maintains that:

H_0 #4 There is no significant difference between the learners' choice of essay topics which will elicit texts of different discourse natures.

5.4.1 Investigating the fourth null hypothesis

In order to study the area of learners' preferences, a questionnaire was distributed among 87 Iranian university students, mainly from Tehran University, Tarbiat Modarres University, and Open University. In this questionnaire the eight topics used to elicit compositions in the *production phase* of the study were randomly listed and the students were asked to rank them on a Likert-type scale according to their preferences. In other words, when they found a prompt easy and appealing and they thought that they could perform better on that topic they would allocate a high grade on the Likert scale, i.e. four or five; naturally a topic which looked difficult to write on

would receive a low mark of one or two. An analogy is that the questionnaire looked

like a menu with four categories of main courses with two dishes for each course. The

subjects were required to taste and mark them and according to the perceived degree

of difficulty, they seem to be tasty. The average scores for all eight topics on a Likert-

type scale are given in the table 5.19

Table 5.19 Descriptive statistics for the eight topics from four discourse types

	Mean	Std Dev	Minimum	Maximum	N
Arg. # 1	2.79	1.29	1	5	87
Arg. #2	2.99	1.42	1	5	87
Nar. #2	3.29	1.36	1	5	87
Exp. #2	3.33	1.24	1	5	87
Exp. #1	3.40	1.31	1	5	87
Des. #1	3.51	1.39	1	5	87
Des. #2	3.62	1.27	1	5	87
Nar. #1	3.80	1.24	1	5	87

In the table 5.19 the prompts have been listed in an ascending order based on the mean

of the scores given to the topics by the subjects beginning with Argumentative topic #1

as the lowest and ending with narrative topic #1 as the highest mean score obtained.

With the exception of narrative, whose assigned levels of difficulty are split, the modes

are very clearly grouped into a perceived rank order of difficulty. However this

apparent rank order needs to be tested for its significance.

So, finally to ensure that the differences among the mean ranks are statistically

significant and are not due to accidental variation a repeated measures ANOVA was

conducted the result of which is given in table 5.20:

Table 5.20 Repeated measures ANOVA for eight different Topics

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	961.38	602	1.60		
PROMPT	65.37	7	9.34	5.85	.000***

The results are significantly high ($p < 0.001$) to confirm that there is a meaningful

difference among the eight prompts which means that learners do indeed prefer some

over others. However, each pair of these eight topics belong to four different modes of

discourse. The observed significant result shows the difference between individual prompts and it is necessary to calculate the total mean for each discourse type by aggregating the scores for paired topics. The results of these calculations are given in table 5.21, which apparently shows a clear order.

Table 5.21 Aggregated mean scores for topics paired for the discourse type

Variable	Mean	Std Dev	Minimum	Maximum	N
ARG	5.78	2.08	2.00	10.00	87
EXP	6.74	2.28	2.00	10.00	87
NAR	7.09	2.10	2.00	10.00	87
DES	7.13	2.02	2.00	10.00	87

In the next step of the statistical analysis, a repeated measures ANOVA was run to ensure that the apparent rank order of difference among the means of these four discourse types is significant. Table 5.22 shows the result:

Table 5.22 The result of repeated measures ANOVA for different discourse types

Source of Variation	SS	DF	MS	F	Sig of F
WITHIN+RESIDUAL	1096.91	258	4.25		
DISCOURSE TYPE	102.59	3	34.20	8.04	.000***

As the result of this analysis, it might be claimed that there is a significant difference between the students' perception of these four discourse types (p<0.000) . Still, it is necessary to find the location of the difference by running the Tukey's HSD multiple range test. Table 5.23 presents the location of significant differences with asterisks.

Table 5.23 The Tukey's HSD test for the total means of paired topics

Mean	Mode	nar	des	exp	arg
7.09	nar				
7.13	des				
6.74	exp				
5.78	arg	*	*	*	

As it is obviously projected in Table 5.23, argument is different from all other discourse types. Although there is considerable difference between means for explanation and narration (0.35) and also between means for explanation and description (0.39), these differences, while more relaxed post hoc tests may find them significant, are not, in fact, big enough to be considered significant by a much more conservative test like Tukey's HSD which requires a mean difference of (0.61).

Considering the fact that the repeated measures analysis of variance for both individual prompts and discourse-type-based paired prompts shows a significant difference, it can

be concluded that the fourth null hypothesis is rejected and hence the alternative hypothesis is accepted:

Hypothesis #4: "_THERE IS_ a significant difference between the learners' choice of essay topics which will elicit texts of different discourse natures."

It should be pointed out that the difference implied in this hypothesis needs to be restricted to argument and the rest of discourse modes, as located by the Tukey's HSD test. However, it was quite likely to find a wider range of difference among modes, if more lenient post hoc multiple-range tests were applied.

5.4.2 Discussion

The order of topics in the table 5.19 is quite interesting and revealing. Except for topic Nar.#2, which will be discussed separately for its strange behaviour, the rest of the topics follow a clear pattern. It is stunningly surprising that both argumentative essay prompts were scored the lowest and below the average 3. Argument, once again, was marked as the least favourable and hence proved to be the most difficult discourse type, as perceived by these students. Next to Argument and second in order of difficulty is Explanation. And, quite interestingly, Description and Narrative seems to be the most favourite types of compositions. This fascinating order of discourse modes is better represented in the Table 5.21. It is evident that prompts for narration and description are more popular among the EFL learners. The similarity of mean scores for these two modes may also indicate that subjects favoured both of them as easy modes to write in.

Narrative #2 is exceptional because of a very important factor, and probably more important than discourse mode, and that is the factor of familiarity. In the early stages of topic selection a lot of care was taken to choose homogeneous topics of similar degree of familiarity and interest so that the only remaining significant difference would be the diversity of discourse types. Nevertheless this topic behaved strangely. Topic Nar. #2, which was about the life of a famous person who rose from a humble background, was chosen according to the taste of L1 learners who write quite a lot on this sort of topic and the biography of Abraham Lincoln is a hackneyed, banal instance. However, Iranian students seldom write on this sort of issue in their composition classes. Therefore, from the very beginning this prompt was expected to defy the proposed patterns. Observing such a difference among L1 and L2 learners is quite informative and of great importance. It is a warning to the test batteries not to rely on native- speakers and their intuitions in the process of topic selection. This justification is, of course, just a hunch. As the issue of familiarity is culture-bound, it is not still clear that other L2 learners with different cultural background would react to this sort of discrepancy.

CHAPTER SIX
CONCLUSIONS

'Tis with our judgements as our watches, none
Go just alike, yet each believes his own.

Alexander Pope, Essay on Criticism, I, 9-10

6.1 An overview

Much has been said about the importance of writing skill and its role in academic

career and its paramount vitality in language teaching has been very widely

acknowledged. A brief account of such a common viewpoint has been given in chapter

two. For instance, Hamp-Lyons and Kroll (1997, p.1) reiterate that "of all the language

skills, writing is the one that can affect a student's college career the most." Testing

writing is also viewed to be of a similar status. There are many reasons to evaluate

writing such as "student placement in composition classes, programe evaluation,

instructional design needs, individual diagnosis and progress, evaluation of exit

competencies, and mastery of course content" (Reid, 1990, p.191).

There are many factors that may affect the writing ability of language learners. One of

them and probably the most important one is the prompt itself. The determining role of

topics in writing essays used to be a "neglected variable" (Ruth and Murphy, 1988)

before the second half of 1980s, but since then and throughout the 1990s more

attention has been paid towards the impact of this variable on writing ability (e.g.

Hamp-Lyons and Mathias, 1994). The prompt, does, of course, play a quite crucial

role in the assessment of writing, when, as is conventional, this is tested by asking

students to write timed essays in response to such prompts. Within this relatively

neglected variable, there are smaller but more delicate variables, some of them almost

completely unexplored. The effect of mode of discourse on ESL/EFL writing performance is one these virgin fields. Despite the existence of some evidence in L1 to confirm that certain discourse modes like argument are more difficult than description or narrative, there has been little research to either confirm or reject the validity of these findings in an EFL environment in general and EAP in particular.

The significance of EAP for this area of research into writing lies in the assumption that subjects of most experiments concerning L1, were young children with premature cognitive development and limited exposure to varieties of argumentative materials. Besides, the nature of a discourse mode like narrative as used by young pupils is much simpler and easier than what it means to adult students for whom it is more likely to be a highly sophisticated genre.

There is a kind of reasoning —apparently a rationalisation— that tries to attribute students' relative difficulty in writing argument to their insufficient exposure to this mode instead of pointing to internal and possibly inherent complications of this mode. Such an argument is clearly reflected in Freedman and Pringle (1989), where they try to answer to the question "Why students can't write arguments":

"First, from their earliest years, children are exposed to written narratives: stories are read to them and the first books they themselves read are typically stories. They do not read arguments, nor are arguments read to them. Secondly, Oral narratives offer a model for written narratives in a way that is not true of arguments."(p.76)

So, from the beginning of the present research, there was a hunch that previous findings in L1 environment may not be applicable to EFL tertiary students, since the familiarity with stories and absence of exposure to arguments may not apply to all at post-secondary levels.

This study was intended to investigate the probable differences that four basic modes of **Argumentation, Narration, Exposition,** and **Description** may cause at three levels of *production, recognition,* and *attitude.* Therefore, three research questions were to be answered. Respectively, the research questions are:

Research question #1: Are some modes of discourse (i.e. *argument/explanation*) relatively more difficult to write on by EFL learners compared against other modes *(i.e. narration/description)*?

Research question #2: Do the essays written in different discourse types bear different degrees of cognitive and/or linguistic difficulty, when turned into cloze tests, to recognise and reconstruct?

Research Question #3: Do EFL writers prefer certain mode(s) of discourse over other mode(s) while choosing a topic to write on?

On the basis of these three questions, four null hypotheses were generated, two for the first question and one for each of the other two:

H_0 #1: There is no statistically significant difference between the scores of compositions written by the same EFL subjects in two different discourse modes of argument and description.

H$_0$ #2: There is no statistically significant difference between the scores of compositions written by the same EFL subjects in two different discourse modes of explanation and narration.

H$_0$ #3: There is no statistically significant difference between the learners' scores obtained from four cloze tests derived from four texts each written in a different major discourse type.

H$_0$ #4 There is no significant difference between the learners' choice of essay topics which will elicit texts of different discourse natures.

6.2 Summary of results

To investigate the first null hypothesis 97 Iranian EFL students wrote essays on four topics, two argumentative and two descriptive. The compositions were rated by two raters and each subject received a set of two scores for their writing ability in the argumentative mode and descriptive mode. A repeated t-test showed that there is a significant difference between the means of these two sets. Hence, **the first null hypothesis was rejected**. The same procedure was repeated for comparing narrative against exposition. This time no significance was observed which means that the second null hypothesis had to be accepted. In other words, argument is more difficult for those students to write than description but narration and exposition seem to be at the same difficulty level.

Unlike the first phase of the study which compared pairs of discourse modes, the second phase of the study compared all four discourse modes against each other.

Four passages written by the same person but in different modes of discourse were turned into cloze tests and then administered to 100 Iranian EFL students. Finally each subject received a set of four scores for their performance on four passages. A repeated measures ANOVA was applied to these scores. The result showed that there was a highly significant difference between the means of scores for the four types of passages. The Tukey's HSD, a post hoc multiple range test, showed that all these passages are different from each other, ranking argument as the most difficult for those students and narrative as the easiest one. These findings were further reconfirmed by readability indices calculated for these passages.

By **rejecting the third null hypothesis** and accepting the alternative that cloze tests based on different discourse-type passages yield different results, it can be also concluded that topics which elicit different types of discourse will result in texts with different levels of cognitive and linguistic difficulty.

In the third and final stage, eight essay prompts, were presented to 87 subjects to rate them on a Likert-type scale according to their preferences to write essays on them. There were two topics of each mode in this scrambled list. Once again, the means for these four pairs of topics ranked narration as the most and argument as the least favoured topic types. Description and exposition came second and third, respectively. Although, the repeated measures analysis of variance showed that the difference among means is significant, the Tukey test only located argument as the source of difference.

Considering the statistically significant results, **the fourth null hypothesis was rejected,** too. This can be interpreted that these EFL learners, even at tertiary level, do not like argumentative topics and prefer easier modes like narration and description.

6.3 Conclusions

According to the results of the present study, which has been an attempt to fill in the existing gap in the field of second/foreign language writing with regard to the impact of mode of discourse on essay topic difficulty, the following conclusions may be drawn:

1. A significant difference among essays, elicited in different modes of discourse, has been observed. Argument is clearly more difficult to produce. Narrative, on the other hand, seems to be the easiest one. Description stands in between.

2. Exposition is very elusive and does not follow any predictable pattern, This could be due to its more complicated nature. It may require a combination of other discourse modes to write an expository essay.

3. The difference among the four discourse types is not simply a matter of training and practising in writing across these modes. They are linguistically and cognitively different. A probable reason, as suggested by Freedman and Pringle, (1998), may be that structuring in argument requires Vygotskian abstracting ability.

4. Students are consciously or sub-consciously aware of such differences and express preferences for easier modes over difficult ones.

5. Mode of discourse is just one of the several variables affecting the difficulty level of topics. There are other variables (e.g. familiarity), some of which even may have an

overriding effect. Further research is required to investigate the weight of these affecting variables.

6. Unlike what Hamp-Lyons and Mathias (1994) hypothesised, raters either do not score compositions written on a difficult topic higher than those written on an easier topic, or if there is such a compensation, it is smaller than the effect of mode difficulty.

7. This study confirms the previous findings (Hayward, 1988; Osbourne and Mulling, 1994; Polio and Glew, 1994) that students perceive different degrees of difficulty among topics. It seems that this differentiation is based on a sound judgement, a fact challenged by Ruth and Murphy (1988) and Kroll (1990b), but never proved before.

6.4 Implications for language testing

The experts in the field of language testing and related lucrative businesses are, perhaps, among the first who may be interested in this study and its results. The equality of writing tasks given to the applicants who inundate major language evaluation organisations such as TOEFL, IELTS, MELAB and others is central to the establishment of reliability and validity for these test batteries.

For these tests to maintain an acceptable degree of reliability, they should yield consistent results. To do so, the prompts given in different administrations should be parallel, with the same degree of difficulty. Accepting that topics of different discourse modes differ in their difficulty levels, it would mean that either the reliability or the validity, or both of them, are theoretically threatened. If topics given in different administrations differ in the type of discourse they elicit, then those topics cannot be

parallel. They do not bear equal linguistic and cognitive demands on test-takers. In reality, though, it seems that this is not the case, because not all four modes of discourse are likely to be used in actual examinations. For instance, it is very unlikely to see a narrative essay prompt in any language testing project. This means that language tests do not touch all areas of writing competence as four modes are concerned and, therefore, that certain constructs are not measured by writing tests. While it is a basic tenet of psychometrics that applicants must be provided with topics that evaluate their best performance, certain types of writing like narration are excluded from many writing evaluation programmes and instead applicants are forced to write on issues which may not be suitable. This is a serious flaw in the construct validity of tests (Polio and Glew, 1994) which deserves more attention in the future. Another important point here is the impact of these findings on the widely discussed portfolio method of writing assessment. It was pointed out in the second chapter that the portfolio approach is becoming exceedingly popular in some academic environments. This means that teachers are supposed to make their judgements by relying on more than one sample essay. It is then logical to recommend that teachers and even official test administers should elicit either four essays, each in a different mode of discourse whenever possible, and, if not, they should elicit one easy mode and one difficult mode of discourse. This point has been stressed by Hamp-Lyons and Kroll (1997), too.

It is also advisable to offer different choices of topics to the learners to enable them to exhibit all their potentialities. This can be done by offering topics of different discourse modes or even two topics intended to elicit one certain mode of discourse.

Last but not least, is the finding about cloze tests. Cloze tests have been in use for nearly half a century and still are highly popular and widely used. As a result of the spectacular attention this type of test has received from researchers, a great number of experiments have been done on most aspects of cloze tests. To add a little bit to this rich research, it has been observed during the course of this study that texts produced in different types of discourse yield different results. The outstanding feature of this finding is that all four texts were written by the same person under very similar circumstances. In practice, using cloze tests to evaluate the readability of texts and reading ability of students may need some adjustments, for instance to standardise the number of n for deleting words in an argumentative text should be bigger than the one for a narrative text to get two equivalent cloze tests.

6.5 Pedagogical Implications

Teaching writing may also benefit from the findings of this study. This can be viewed from both process and product points of view.

The advocates of process approaches may find it useful that, as Robinson (1988) recommended, a combination of process approach and traditional rhetoric may be more successful in an academic environment. In this case, a distinction between different types of texts and their patterns and the strategies which should be used by learners to implement these patterns would be central to the method.

Textbooks intended to be used in process approaches, usually have sections on pre-writing, brainstorming techniques, planning, outlining, drafting, revising, and editing (Hedge, 1988; White, 1988; Nelson and Murphy, 1992; Porte, 1995). In this case it might be necessary to adjust these sections to cater for all modes of discourse. If, as

this study shows, argument is the most difficult mode for EFL students, then perhaps more time should be given to planning and revising for this particular mode.

And if writing is going to be taught through a product-oriented approach, then it would be necessary to teach the learners the four modes of discourse and their expected and acceptable patterns of organising texts. This may sound somewhat old-fashioned and stereotypical but so is the product approach and nevertheless it is "still dominant in ESL teaching/learning situations today and textbooks" (El-Bacha, 1997, p. 55).

Sequencing the materials to be taught, regardless of the preferred approach, is helpful. Accepting the difference between the cognitive and linguistic burden imposed by different discourse modes, then, would mean that it is logical to start off with the most accessible or easiest narrative, followed by description, exposition, and finally the most demanding, argument. This discussion, of course, should not detract from the importance of narratives. It would be wrong to join those, criticised by Hesse(1989, p.106), who believe that "narrative is too simple, too literary, and, most significantly, too 'impractical' to merit serious attention". This sequencing is from *relatively* familiar to *relatively* unfamiliar, from *relatively* easy to *relatively* difficult. Arguably, all four modes are important for academic writing in various disciplines (Hale et al., 1996).

This study can have some positive impact on the ELT/ EAP in Iran, too. Referring to the proficiency-oriented approaches to teaching FL writing, Scott (1996, p. 147) describes two different approaches. In this eclectic situation, she believes, " some teachers may focus on developing students' command of grammar and syntax, while others may stress practice of language functions, such as describing or expressing an opinion." Presently, in Iran the dominant approach to teaching writing is the former

one. No matter how successful it may be, putting too much emphasis on isolated elements of language is harmful for the ELT in Iran where crude versions of some traditional teaching methods such as Grammar-Translation are still being practised. To eliminate this vicious circle, a strong move towards more communicative and a more meaningful viewpoint of language is inevitable. Within the domain of writing, this study can promote such a viewpoint among the teachers of English by encouraging them to shift to the latter approach i.e. teaching writing at a discourse level. This researcher strongly believes that raising teachers' awareness towards different modes of discourse and familiarising learners with different functions of language could be a remedy to this deteriorating situation.

6.6 Recommendations for further research

Earlier in this thesis, an opinion was voiced that PhD theses, despite their rigorous designs and analyses and the vigorous efforts involved, are usually ignored on library shelves and, compared to journal articles, are much less appreciated. Still, this researcher hopes that someone may find it appealing to further this line of research.

Naturally, the immediate concern would be replicating the study preferably with the same topics but different subjects. By different subjects, subjects with a mother-tongue different from Persian is meant.

Among the four discourse modes studied, exposition seemed to be quite elusive. It has been argued that this may be due to the diversity of patterns collected under the umbrella term of exposition. Further research is recommended to clear the situation by breaking exposition into smaller sub-modes. The work of Hale et al. (1996) could be a suitable starting point.

Another recommendation for people, who will embark on similar studies, is to make a clear profile of their subjects and then investigate the impact of certain features like sex and language proficiency on their performance in different modes. It is not very unlikely that differences between male and female subjects will be observed in the patterns they use for different modes.

More ambitious researchers, following Swales (1981), Hopkins and Dudley-Evans (1988), Swales (1990) and Dudley-Evans (1994), may try to study and establish patterns for *"moves"* made by weak writers and more fluent writers across discourse types. The pedagogical implications of this approach and its prospective results can be hardly exaggerated.

A very important avenue, open to be followed up by other researchers, is to eliminate the practical restrictions that prevented the design of this study to be an ideal one. The following section is a review of limitations faced by the researcher during the course of this project.

6.6.1 Limitations of the study

There are two types of flaws in this study; some of them were inevitable and some of them, to some extent, could have been avoided or at least there was room for some improvements.

Probably the most complicated task during this research was to select the topics with differences in the nature of the text they were supposed to elicit. Despite all the care taken from collecting samples from Iranian and non-Iranian students to consulting language teachers here in Britain and in Iran, only a bold researcher still may claim that the set of topics selected for this study is completely satisfactory. After all, it was this

researcher who made the ultimate decisions and finalised the wording of the prompts. Retrospectively, it seems that possible to improve the topics so that the final product would be a more representative of that certain mode.

One of the controversial issues in evaluating writing is the *rater*. Obviously, increasing the number of raters would help in getting more reliable results. Again for practical issues (the raters had to be paid) only two raters were employed which seems to be satisfactory. Usually when there is a big difference between the scores given by the two raters to an essay a third rater should judge between the two. This whole practice is labelled by Hamp-Lyons (personal correspondence) as "fraudulent". Her reason is that there is, theoretically, no justification that the third person is right and that the score they give worsens the situation. The third judge in this study was the researcher and being aware of the research aims and the hypotheses, despite all the care taken, this might have had bad effects on his impartiality. This is a flaw which , hopefully, has not affected the results (but in any case, few studies have more than three raters).

Another worrying limitation is the practical restrictions faced by the researcher during the course of carrying out the experiment in Iran. Despite all the educational improvements in recent years and the fact that more attention is paid to research in Iranian universities, unfortunately research is still viewed as inferior to teaching. In Iranian academic culture the immediate duty of academic staff is to give lectures in classes and research comes later, if at all. At any case research should not interfere with teaching because some authorities in educational institutes find research studies interfere with their curriculum. With this background it is easier to see the situation in which several classes had be involved in this research project. This became possible only with the extraordinarily sincere co-operation offered from the teachers which has

to be acknowledged. However, it was a pity that the design of the study had to be different from the desired one because of the practical issues.

The most important defect in the design of the study was the move from comparing all four modes against each other written by each student to a split version of comparing just one supposedly easy mode against one supposedly difficult one. The main reason for this negative modification was that taking the time of the class for four consecutive sessions and keeping the students occupied with writing essays was not feasible in the context. This shortcoming, however, was eliminated in the other two stages of the study i.e. cloze test and the questionnaire.

The second practical problem was finding homogeneous groups to facilitate the process of comparing the subjects' performances. Obviously a whole class had to receive the test and the composition of the class should have been left intact. As a result of inevitable differences among mixed ability groups the mean score for classes differ from one another considerably. This is not, of course, a major flaw in the design for every subject is being compared against himself or herself. The problem arising from difference among the mean scores for groups is that it is impossible to find out which sub-set of tasks were easier or more difficult. For example, from the highest mean score for a group it can not be interpreted that those subjects received the easiest set of tasks. It could be simply because they were post-graduate students, whereas the lowest mean score belonged to the freshmen which is quite natural and predictable. This ambiguity could have been avoided if it had been possible to test the subjects' language proficiency. Having had access to their proficiency level, it would have been also possible to calculate the correlation between students' language maturity and writing ability.

Another minor problem with the production stage of the research design seems to be the overwhelming majority of female students over the male ones. It is true that English language and literature is very popular among Iranian women but in this particular case the reason is that in many universities, especially in the Open University, most classes are not mixed. Although, most classes in state universities are mixed, Open University, which has been under severe criticism for serving the privileged and non-religious stratum in a strongly religious society, cautiously holds separate male and female classes. For this reason, it was just accidental that most of the classes had to be chosen from the female-only shift. There is a wealth of research to demonstrate that gender affects almost all aspects of language achievement. Ellis (1994) surveys a number of studies conducted on the gender-based differences in L2 and concludes that (ibid, p.202 and 203) " female learners generally do better than male" and that "women tackle the task of learning an L2 differently from men". So it is possible that gender has been an uncontrolled variable which may have influenced the present results. Because the overwhelming majority of students in the present sample were female, it was not possible to investigate this variable.

There are trivial problems with the second and third stages of the study, too. The main avoidable imperfection is that both of these phases were conducted using the same set of topics which was used in the production stage. An improved precautionary action would have been to use three different sets of prompts for the three stages of the experiment. This way it would have been more reassuring, with a convincing degree of generalisability, that any observed differences in the results could be attributed to the differences in modes.

There are other limitations which probably relate to the general scale of the study, and are therefore, inherent in a project of this sort. Thus, perhaps ideally all students involved would not only have written in each mode but should perhaps write more than once in each. Thus, too, the cloze text would be written by more than one person (in sets of four texts) so that students' tests would be based on several sets. While these are clearly limitations, in the Iranian context of the present field work, they would only really be overcome in a major funded study with more researchers and much more time, i.e. a different sort of project.

6.7 Concluding remarks

Only those who do not do anything do not commit mistakes. Therefore, the design of the present study, like almost any other research project which uses humanity as its subject, is far from the ideal. This is mainly because social and educational researchers do not carry out experiments in laboratories with perfect control over the guinea pigs, heat, light, chemicals, and other variables involved. Researchers in humanities and social sciences have to go out into society and educational institutes, find their delicate and sophisticated subjects who are human beings and conduct the project while negotiating and compromising with the authorities and teachers over the regulations , feasibility, time, etc.

Life is an ongoing process of learning and so is a PhD course. On average it usually takes four years for an overseas student to complete a doctorate course at British universities. Throughout this period, students are involved in the process of learning. Every single day and even every single hour teaches them something new. Reading a new book, finding a recently published article in a journal, attending classes, seminars

and conferences, and finally meeting with the supervisor all help students to broaden their knowledge of the field. Ironically, the largest and the most important contribution to the students development comes towards the end of the period from the examiners. All this happens after the novice or candidate has already chosen the topic of the project and set out to carry out the field study. Undoubtedly the most useful, realistic and practical understanding of what research is and how it should be conducted does not come from research manuals or statistics handbooks. It is, rather, the experience gained through the actual experiment that turns a relative novice into a more experienced researcher. Once again, the outcome of this development cannot have any impacts on the experiment as it is too late to re-conduct the whole experiment from the very beginning.

As said before, most of these shortcomings were practically inevitable but some of them could have been avoided. Many of the problems showed themselves during the actual process of experiment and could not have been easily predicted. At any case, the experience gained by this study throughout the past four years has helped the researcher to be capable of planning better research designs in future.

Despite these shortcomings, this beginner researcher hopes that the findings of this study contribute, no matter how little, to the important and progressing field of ESL/EFL writing and motivate other colleagues to not only replicate the study to confirm or reject the results but also further our understanding of writing ability in general and mysteries of topic difficulty in particular.

I will also keep up the hope that this study may raise awareness among the academic community that EAP in Iran needs to be revised and revisited. Presently, there is a struggle to shift ELT teachers' attention from a futile effort on teaching grammar to

the more useful skill of reading. It would be a great leap towards a more meaningful and communicative language teaching in Iran to remind the education authorities that writing , if not more vital and prestigious than reading, at least it is universally acknowledged to be inextricably bound to reading skill in academic environments. Writing and reading, as this writer has found on reading his writing, and as research shows (Grabe and Kaplan, 1996), complement and complete each other. Teachers of writing commonly remind students of the importance of reading like writers and writing like readers, i.e. to read/write with full awareness of the complementary skill. It is hardly possible for learners to be able to produce written text unless they also obtain written language input through reading.

For a reflective researcher, reading written texts about researching writing is often challenging as a form of reading, but it is much more of a challenge to write such texts. Pope's epigram (see p.191) reminds us that this thesis is written according to the writer's judgements but, like much academic writing, it awaits the judgements of readers, too.

APPENDICES

APPENDIX A: THE ORIGINAL TEXT OF FOUR PASSAGES USED FOR THE CLOZE TEST

Exploring the space or any other scientific undertaking in itself is neither good nor bad. It depends on the ends it has in view and the circumstances which endorse its immediate enactment or its postponement for times more appropriate, or even leaving it a mere theory or suggestion untested by practice. Of the last, we can think of many theories that putting them into practice can never bring any advantage to man or, in some circumstances, only add to his misery.

Programmes of space exploration are of the second kind which though potentially entail no harm to man and even could be of much help in solving the riddles of life and origin of existence, yet their benefit depends on the current circumstances and the conditions in which man lives. In cases like these, we should act according to the maxim "first things first".

****

If you come to my city from a long distance in your car, you can see the green gardens which surround it since it is located in a higher place than its surrounding. It is not an ordinary city but originally a big oasis in the less friendly environment of the desert. At the border of the city where the sand spreads, you can see the small shrubs and bushes which are called by the locals the fiery tongues because their glassy narrow leaves reflect the sunshine and appear as narrow streaks of light or flame during the day. The earth under these bushes is covered by mat of thick green grass which to the comers to the city appears a beautiful scenery that distinguishes the yellow brown sand from the green floor of these bushes. Having passed these shrubs which are here and there, you

can see young men with their double-barrelled hunting guns. Then you come to a wide extent of acres of field planted with wheat, barley, oats, as these are the city's major agricultural products.

****

"Ghost" is one of the interesting films I saw two years ago. The main character of the film is a young man who is honest, handsome and hardworking. Along with his friend he works in one of big companies as an accountant or something like this. His girlfriend is a slim, innocent-looking young woman with big eyes and short hair. She is an artist. One day the hero, I mean the main character, discovers that he has millions of dollars in his account and shares his secret with his colleague. The friend turns to be treacherous and at night, while the hero is walking in a dark street with his girlfriend, he sends an ugly-looking villain to kill him so that he can't take the whole money for himself. The hero gets killed after a futile fighting with the villain.

****

Learning a foreign language, like any other activity, arises from and is motivated by the need which man feels for a foreign language. Once the motivation is strong enough, one naturally would look for those procedures and techniques that will satisfy his need. There is no doubt that language learning should include all the four skills of learning, yet, the motivation for the learning is not the same for all those who learn the language. Some people would find speaking is of their major importance, while others' preference could be of any of the other three skills. Therefore, learning procedures should not be kept stiff and inflexible in all situations. The procedures also should vary according to age group, sex, and social and cultural background. Depending on the age, the method and appropriate techniques will change. Naturally, the inclination with

an adult will be most for reading rather than speaking which is more favoured by younger ages.

APPENDIX B: CLOZE TEST USED IN THE *RECOGNITION PHASE* OF THE

STUDY

Exploring the space or any other scientific undertaking in itself is neither good

nor bad. It depends on the ends it h . . . . . in view and the circumstances which e

. . . . . its immediate enactment or its postponement f . . . . . times more

appropriate, or even leaving i . . . . . a mere theory or suggestion untested b . . . .

. practice. Of the last, we can t . . . . . of many theories that putting them i . . . . .

practice can never bring any advantage   t . . . . . . man or, in some

circumstances, only a . . . . . to his misery. Programmes of space e . . . . . are of

the second kind which   t . . . . . potentially entail no harm to man a . . . . .

even could be of much help i . . . . . solving the riddles of life and o . . . . . of

existence, yet their benefit depends o . . . . . the current circumstances and the

conditions in which man lives. In cases like these, we should act according to the

maxim "first things first".

****

If you come to my city from a long distance in your car, you can see the green

gardens which surround it since it is located in a higher place than its

surrounding. It is not an ordinary city   b . . . . . originally a big oasis in the l .

. . . . friendly environment of the desert. At t . . . . . border of the city where the

s . . . . . spreads, you can see the small s . . . . . and bushes which are called by t .

. . . . locals the fiery tongues because their gl . . . . . narrow leaves reflect the sunshine and a . . . . . as narrow streaks of light or f . . . . . during the day. The earth under t . . . . . bushes is covered by mat o . . . . . thick green grass which to the c . . . . . to the city appears a beautiful s . . . . that distinguishes the yellow brown sand f . . . . . the green floor of these bushes. H . . . . . passed these shrubs which are here and there, you can see young men with their double-barrelled hunting guns. Then you come to a wide extent of acres of field planted with wheat, barley, oats, as these are the city's major agricultural products.

****

"Ghost" is one of the interesting films I saw two years ago. The main character of the film i . . . . . a young man who is honest, h . . . . . and hardworking. Along with his friend h . . . . . works in one of big companies a . . . . . an accountant or something like this. H . . . . . girlfriend is a slim, innocent-looking young w . . . . . with big eyes and short hair. S . . . . . is an artist. One day the h . . . . . , I mean the main character, discovers t . . . . . he has millions of dollars in h . . . . . account and shares his secret with h . . . . . colleague. The friend turns to be t . . . . . and at night, while the hero i . . . . . walking in a dark street with h . . . . . girlfriend, he sends an ugly-looking villain t . . . . . kill him so that he can't take the whole money for himself. The hero gets killed after a futile fighting with the villain.

****

Learning a foreign language, like any other activity, arises from and is motivated by the need which man feels for a foreign language. Once the motivation is strong enough, one n . . . . . would look for those procedures and t . . . . . that

will satisfy his need. There i . . . . . no doubt that language learning should i . . . . . all the four skills of learning, y . . . . . , the motivation for the learning is n . . . . . the same for all those who l . . . . . the language. Some people would find s . . . . . is of their major importance, while o . . . . . preference could be of any of t . . . . other three skills. Therefore, learning procedures s . . . . . not be kept stiff and inflexible i . . . . . all situations. The procedures also should v . . . . . according to age group, sex, and s . . . . . and cultural background. Depending on the a . . . . . , the method and appropriate techniques will change. Naturally, the inclination with an adult will be most for reading rather than speaking which is more favoured by younger ages.

APPENDIX C: THE QUESTIONNAIRE USED IN THE *ATTITUDE PHASE* OF THE STUDY

Please read each of the following essay titles and rank them, by putting a ✓ in the appropriate circle, according to how well you can write an essay on that topic. Remember that ① shows your dislike, whereas ⑤ represents your most favourite title.

☑ ✎

....
..

1. Argue the advantages and disadvantages of commercial advertising in the media. Take a clear position.　　　　　　① ② ③ ④ ⑤

....
....

2. Imagine that you are in the 25th century. Describe the people and their way of life, buildings, vehicles, etc.　　　　　① ② ③ ④ ⑤

....
....

3. Write the story of the most interesting film you have ever seen. Narrate the events as they happened in the film. Your writing should look like a story.
　　　　　　　　　　　　　　　　　① ② ③ ④ ⑤

....
....

4. How can we make teaching and learning in Iranian universities more effective? Explain your ideas.　　　　　　① ② ③ ④ ⑤

....
....

5. Considering the situation in Iran, explain how a foreign language should be learned.
　　　　　　　　　　　　　　　　　① ② ③ ④ ⑤

....
....

6. Write the story of someone who rose from an unknown background and humble beginning to become successful and famous. Narrate the main events of his/her lifestory in the order in which they happened.
　　　　　　　　　　　　　　　　　① ② ③ ④ ⑤

7. Describe your hometown or any interesting city so that readers can have a clear image of the city in their minds. ① ② ③ ④ ⑤

....
....

8. At the present time, the human race spends huge sums of money to explore space. Do you find it a waste or do you view space projects as essential to expanding our knowledge. Discuss the issue. ① ② ③ ④
⑤

....
✶✶✶✶✶✶✶✶✶✶✶✶✶✶✶✶✶✶✶✶

ﻻ ࠫThank you

Class #1

Describe your hometown or any interesting city so that readers can have a clear image of the city in their minds.

Rater #1	Rater #2
14.00	12.00
10.00	13.50
11.00	9.00
12.00	13.00
14.00	12.00
16.00	16.00
16.00	16.00
16.00	17.00
16.00	16.00
18.00	16.00
18.00	14.00
15.00	16.00
16.00	17.00
20.00	19.50
20.00	18.50
20.00	19.00
16.00	16.50
19.00	20.00
18.00	19.00

19.00	16.50
18.00	16.50
19.00	14.50
15.00	12.00
14.00	14.00
15.00	16.00
15.00	15.00
14.00	15.50
13.00	10.50
13.00	14.00
12.00	10.50
13.00	15.00
12.00	12.00
9.00	10.00
9.00	7.00

Class #1

At the present time, the human race spends huge sums of money to explore space. Do you find it a waste or do you view space projects as essential to expanding our knowledge. Discuss the issue.

Rater #1	Rater #2
15.00	11.00
10.00	12.00
14.00	10.00
8.00	10.00
15.00	12.00
18.00	14.50
16.00	15.00
17.00	14.00
15.00	16.50
18.00	17.00
17.00	13.00
15.00	10.00
16.00	13.00
20.00	19.00
20.00	18.50
20.00	18.00
18.00	14.50
19.00	20.00
19.00	17.50
18.00	14.50
19.00	16.50
18.00	17.00
16.00	15.00
16.00	13.50
15.00	13.50
14.00	15.00
17.00	16.00
13.00	10.50
13.00	10.50
14.00	10.50

13.00   14.00
13.00   13.00
10.00   9.00
 9.00   10.50

class#2

At the present time, the human race spends huge sums of money to explore space. Do you find it a waste or do you view space projects as essential to expanding our knowledge. Discuss the issue.

Rater #1	Rater #2
8.00	7.00
6.00	6.00
10.00	6.00
11.00	12.00
7.00	13.00
7.00	7.00
7.00	8.00
11.00	11.00
11.00	9.00
15.00	19.50
6.00	6.00
4.00	4.00
10.00	13.00
12.00	17.00
10.00	8.00
14.00	14.00
7.00	7.00
8.00	5.00
9.00	6.00
15.00	17.00
11.00	15.00
10.0	7.00

----

class #2

Imagine that you are in the 25th century. Describe the people and their way of life, buildings, vehicles, etc.

Rater #1	Rater #2
8.00	8.00
8.00	7.00
10.00	8.00
9.00	14.00
7.00	12.00
7.00	7.00
7.00	6.00
10.00	10.00
11.00	10.00
10.00	16.00
8.00	9.00
6.00	4.00
10.00	14.00
14.00	13.50
11.00	10.00
14.00	10.00
8.00	6.00
10.00	6.00
7.00	9.00
18.00	17.00
13.00	18.00
8.00	10.00

class #3

Write the story of someone who rose from an unknown background and humble beginning to become successful and famous. Narrate the main events of his/her lifestory in the order in which they happened.

Rater #1	Rater #2
6.00	5.00
10.00	9.00
7.00	6.50
5.00	4.50
11.00	9.00
10.00	9.00
13.00	13.50
8.00	8.50
4.00	5.00
8.00	7.00
5.00	5.50
10.00	10.00
11.00	10.00
6.00	8.00
12.00	10.50
3.00	4.00
12.00	13.50
4.00	5.00
3.00	3.00
10.00	12.00
14.00	11.00

----

class #3

**How can we make teaching and learning in Iranian universities more effective.
Explain your ideas.**

Rater #1	Rater #2
9.00	7.00
11.00	10.00
7.00	6.00
5.00	4.00
12.00	10.00
6.00	7.00
11.00	12.50
9.00	9.00
6.00	5.00
7.00	7.00
6.00	5.50
12.00	15.00
9.00	10.00
6.00	6.00
13.00	11.50
6.00	4.00
12.00	10.50
6.00	6.00
3.00	2.00
6.00	9.00
11.00	13.50

==

class #4

Imagine that you are in the 25th century. Describe the people and their way of life, buildings, vehicles, etc.

Rater #1	Rater #2
11.00	15.00
7.00	6.00
12.00	6.00
7.00	3.00
9.00	7.50
10.00	9.00
12.00	12.00
10.00	7.50
5.00	7.50
5.00	7.50
10.00	6.00
15.00	14.00
5.00	5.00
9.00	7.00
10.00	7.50
14.00	15.00
18.00	19.50
12.00	14.00
9.00	6.00
13.00	7.50
17.00	12.00
7.00	7.50
20.00	16.00
16.00	9.00
3.00	4.00

----

class #4
Argue the advantages and disadvantages of commercial advertising in the media.
Take a clear position

Rater #1	Rater #2
9.00	6.00
7.00	7.00
9.00	6.00
4.00	4.50
9.00	6.00
5.00	7.50
9.00	14.00
10.00	6.00
4.00	6.00
5.00	6.00
7.00	6.00
13.00	16.00
6.00	6.00
8.00	7.50
11.00	6.00
13.00	14.00
19.00	19.50
9.00	7.00
10.00	6.00
11.00	13.00
17.00	12.00
15.00	6.00
20.00	16.00
17.00	14.00
4.00	6.00

class #5

Write the story of someone who rose from an unknown background and humble beginning to become successful and famous. Narrate the main events of his/her lifestory in the order in which they happened.

Rater #1	Rater #2
13.00	10.00
9.00	8.00
5.00	4.00
7.00	6.00
4.00	2.00
9.00	7.50
9.00	6.00
6.00	3.00
5.00	3.00
5.00	3.00
10.00	6.00
5.00	3.00
2.00	4.00
10.00	6.00
7.00	3.00
8.00	8.00
10.00	7.00

Class #5

Considering the situation in Iran, explain how a foreign language should be learned.

Rater #1	Rater #2
13.00	11.00
9.00	7.50
8.00	7.50
10.00	7.00
4.00	2.00
9.00	8.00
14.00	8.00
8.00	4.50
5.00	2.00
10.00	3.00
9.00	5.00
9.00	4.00
3.00	5.00
11.00	7.00
9.00	4.00
8.00	7.00
10.00	10.00

class#6

**How can we make teaching and learning in Iranian universities more effective?
Explain your ideas.**

Rater #1	Rater #2
10.00	8.00
7.00	4.00
5.00	3.00
6.00	3.00
5.00	4.00
8.00	5.00
6.00	5.00
10.00	6.00
12.00	11.00
8.00	6.50
5.00	4.00
5.00	4.00
7.00	7.00
5.00	8.00
10.00	7.50
13.00	14.50
13.00	11.50
11.00	8.00
13.00	11.00
2.00	2.00

---

class #6

Write the story of the most interesting film you have ever seen. Narrate the events as they happened in the film. Your writing should look like a story.

Rater #1	Rater #2
9.00	2.50
7.00	4.00
6.00	5.00
6.00	3.00
7.00	3.50
9.00	7.50
6.00	4.00
11.00	8.00
12.00	7.00
7.00	5.00
6.00	5.00
4.00	6.00
7.00	6.50
4.00	7.00
8.00	7.50
13.00	12.50
11.00	11.00
12.00	7.50
11.00	11.50
3.00	3.00

class #7

Write the story of the most interesting film you have ever seen. Narrate the events as they happened in the film. Your writing should look like a story.

Rater #1	Rater #2
3.00	2.50
3.00	2.50
9.00	8.00
8.00	7.00
6.00	7.00
6.00	7.50
11.00	11.00
11.00	10.00
12.00	12.00
9.00	11.00
3.00	5.00
3.00	3.50
12.00	10.00
9.00	9.00
13.00	12.00
13.00	14.00
5.00	4.50
9.00	9.00
11.00	12.00
11.00	11.50
10.00	9.00
9.00	9.00

Class #7

Considering the situation in Iran, explain how a foreign language should be learned.

Rater #1	Rater #2
4.00	2.50
4.00	2.50
9.00	8.50
8.00	6.00
6.00	6.50
6.00	7.00
12.00	10.00
11.00	10.00
12.00	12.00
10.00	10.00
5.00	3.00
4.00	3.50
12.00	11.00
10.00	10.00
12.00	13.00
12.00	13.00
5.00	4.00
9.00	8.50
10.00	10.00
12.00	11.50
10.00	9.00
9.00	9.00

==

class #8

Describe your hometown or any interesting city so that readers can have a clear image of the city in their minds.

Rater #1	Rater #2
3.00	3.50
7.00	9.00
3.00	3.50
4.00	3.00
5.00	4.00
7.00	5.00
7.00	4.50
9.00	3.00
10.00	9.50
7.00	4.00
8.00	4.00
9.00	5.50
9.00	7.50
14.00	16.00
10.00	5.50
5.00	3.50
10.00	7.00
7.00	5.00
10.00	6.00
10.00	6.00
13.00	8.00
10.00	5.50

----

class #8

Argue the advantages and disadvantages of commercial advertising in the media.
Take a clear position

Rater #1	Rater #2
3.00	3.50
6.00	7.00
4.00	2.00
4.00	3.00
4.00	2.50
5.00	2.00
8.00	5.00
8.00	4.00
10.00	8.00
6.00	4.00
8.00	3.00
9.00	7.50
9.00	6.00
12.00	14.50
14.00	7.00
8.00	4.00
9.00	5.00
6.00	3.50
9.00	5.00
12.00	7.00
10.00	5.00
10.00	4.00

APPENDIX E: SCORES FOR THE FOUR PASSAGES OF THE CLOZE TEST

Expository passage	Narrative passage	Descriptive passage	Argument passage
14.00	13.00	13.00	11.00
14.00	15.00	10.00	8.00
12.00	14.00	12.00	11.00
11.00	12.00	9.00	9.00
12.00	14.00	14.00	8.00
13.00	14.00	11.00	11.00
11.00	14.00	12.00	9.00
11.00	12.00	9.00	12.00
10.00	13.00	12.00	12.00
12.00	13.00	10.00	10.00
9.00	12.00	11.00	10.00
11.00	15.00	11.00	13.00
11.00	12.00	7.00	10.00
13.00	11.00	8.00	4.00
8.00	14.00	4.00	10.00
10.00	11.00	9.00	9.00
8.00	13.00	8.00	7.00
9.00	14.00	11.00	8.00
11.00	12.00	10.00	7.00
11.00	13.00	11.00	6.00
7.00	12.00	10.00	9.00
9.00	12.00	9.00	5.00
7.00	12.00	9.00	9.00
4.00	11.00	6.00	6.00
8.00	13.00	6.00	9.00
10.00	14.00	12.00	12.00
13.00	13.00	11.00	10.00
10.00	12.00	12.00	11.00
10.00	12.00	12.00	10.00
12.00	12.00	9.00	6.00
6.00	11.00	5.00	5.00

7.00	13.00	6.00	4.00
10.00	11.00	4.00	4.00
9.00	13.00	7.00	7.00
9.00	12.00	11.00	8.00
9.00	13.00	7.00	8.00
6.00	12.00	7.00	9.00
7.00	11.00	8.00	6.00
7.00	10.00	8.00	4.00
9.00	9.00	7.00	6.00
7.00	8.00	4.00	3.00
8.00	14.00	4.00	7.00
6.00	11.00	6.00	5.00
8.00	14.00	8.00	9.00
7.00	12.00	8.00	9.00
7.00	13.00	7.00	9.00
7.00	12.00	7.00	5.00
8.00	15.00	9.00	11.00
9.00	13.00	5.00	4.00
7.00	14.00	8.00	9.00
6.00	10.00	4.00	2.00
8.00	13.00	5.00	4.00
7.00	13.00	9.00	7.00
4.00	5.00	5.00	3.00
6.00	10.00	6.00	7.00
5.00	9.00	9.00	5.00
3.00	12.00	3.00	6.00
1.00	12.00	6.00	3.00
6.00	10.00	5.00	5.00
3.00	11.00	3.00	5.00
7.00	12.00	4.00	6.00
9.00	10.00	5.00	6.00
6.00	13.00	3.00	6.00
8.00	13.00	12.00	7.00
9.00	11.00	2.00	8.00
11.00	13.00	7.00	6.00
5.00	8.00	2.00	3.00

10.00	12.00	7.00	6.00
5.00	10.00	6.00	7.00
5.00	8.00	5.00	6.00
5.00	12.00	5.00	7.00
5.00	8.00	8.00	6.00
10.00	13.00	7.00	4.00
8.00	11.00	10.00	9.00
13.00	13.00	10.00	10.00
8.00	13.00	10.00	7.00
10.00	14.00	11.00	9.00
9.00	14.00	10.00	5.00
12.00	13.00	9.00	5.00
8.00	10.00	6.00	5.00
8.00	12.00	7.00	8.00
8.00	8.00	6.00	5.00
6.00	10.00	6.00	4.00
11.00	14.00	5.00	5.00
6.00	13.00	6.00	4.00
8.00	11.00	8.00	7.00
8.00	10.00	7.00	3.00
5.00	10.00	5.00	3.00
8.00	14.00	3.00	4.00
6.00	9.00	5.00	
11.00	14.00	4.00	
8.00	8.00	6.00	
11.00	10.00	5.00	
7.00	13.00	8.00	
8.00	13.00	4.00	
11.00	10.00	5.00	
2.00	7.00	9.00	
8.00	14.00	1.00	
7.00	11.00	7.00	
6.00	10.00	6.00	
4.00	11.00	6.00	
11.00	11.00	5.00	
11.00	12.00	3.00	
10.00	12.00	6.00	

APPENDIX F: LIKERT-TYPE RATINGS GIVEN TO THE ESSAY PROMPTS BY IRANIAN EFL LEARNERS

Note: Prompts are listed as they are ordered in the Appendix C.

PROMPTS

#1	#2	#3	#4	#5	#6	#7	#8
5	5	1	3	3	4	1	5
2	4	4	5	4	5	3	3
3	1	5	5	5	4	2	3
4	3	4	3	2	5	2	1
4	3	4	2	4	4	4	5
3	5	5	2	3	4	5	4
1	5	5	2	1	2	3	2
3	1	4	3	4	3	4	5
4	3	5	4	5	3	4	5
3	4	5	3	2	5	3	3
3	4	5	1	3	3	4	1
2	2	3	3	2	2	4	3
2	1	4	2	3	1	4	2
3	4	3	5	4	5	1	3
3	4	3	2	1	3	5	1
3	2	3	2	4	1	3	1
3	2	3	3	3	2	4	3
1	5	5	5	5	5	5	5
3	5	2	5	4	4	5	3
5	2	5	3	3	4	3	2
1	5	5	1	5	5	1	5
3	4	5	5	3	3	5	1
4	3	5	3	4	2	5	5
3	2	4	3	3	4	4	3
3	1	4	4	5	4	5	2
1	3	4	5	2	3	4	3

2	1	3	5	4	5	3	4
2	4	1	5	3	2	1	1
2	4	1	2	2	2	4	5
3	4	3	4	4	3	5	4
5	5	4	5	5	5	3	5
5	5	3	5	4	2	2	4
3	5	4	5	5	5	5	5
5	5	4	5	5	3	5	5
1	4	4	4	3	5	4	3
2	4	5	5	4	5	5	2
1	5	5	3	2	5	5	4
4	1	3	5	4	2	5	3
5	4	5	4	1	4	5	5
2	3	4	4	4	5	4	2
1	5	4	5	5	3	5	1
1	4	2	3	4	2	2	1
4	4	3	5	5	4	4	3
1	5	5	2	2	5	4	1
3	1	2	3	3	2	4	4
4	4	5	3	3	5	4	3
2	5	3	3	2	4	4	4
3	5	5	1	1	3	4	4
1	5	5	4	3	3	5	2
5	4	3	5	5	1	5	3
5	3	2	3	3	1	3	2
1	5	3	2	2	3	3	5
1	5	5	3	2	4	3	2
2	5	5	3	2	1	5	3
3	4	2	3	4	2	3	1
3	2	5	4	4	4	3	4
3	1	3	5	4	4	2	2
3	5	5	3	3	3	4	3
1	1	5	4	4	2	5	5
4	5	5	1	2	4	5	1
2	5	4	2	4	5	5	3
1	4	5	3	1	1	3	4

1	5	5	1	1	1	5	1
5	2	1	4	5	1	1	4
2	4	5	2	1	5	1	1
2	3	3	1	2	3	4	4
2	1	5	1	2	4	3	1
5	2	1	3	5	1	3	3
4	4	4	3	4	2	4	3
3	4	2	5	4	3	3	4
2	3	4	2	4	5	3	1
3	3	4	3	3	4	3	2
3	4	5	3	3	2	2	1
3	4	5	1	2	4	4	3
4	3	1	4	4	5	1	5
4	3	3	4	3	3	3	3
3	5	3	5	5	5	4	4
1	3	5	5	5	5	5	4
2	2	3	2	2	1	2	1
3	5	2	4	4	3	5	2
4	4	3	3	3	1	3	2
1	1	5	5	5	3	1	1
2	5	5	4	3	5	2	5
3	4	4	5	5	4	5	5
1	4	4	2	2	2	5	3
5	2	5	4	3	3	5	4
4	1	4	5	5	2	4	1

==

APPENDIX G: SCORING BANDS FOR THE IELTS

INTERNATIONAL ENGLISH LANGUAGE TESTING SYSTEM

University of Cambridge
Local Examinations
Syndicate
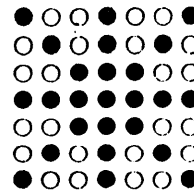
The British Council

International
Development
Program of
Australian
Universities and
Colleges

Test Report Form for Academic Course of Study

Bands

9 Expert User. Has fully operational command of the language; appropriate, accurate and fluent with complete understanding.

8 Very Good User. Has fully operational command of the language with only occasional unsystematic inaccuracies and inappropriacies. Misunderstanding may occur in unfamiliar situations. Handles complex detailed argumentation well.

7 Good User. Has operational command of the language, though with occasional inaccuracies, inappropriacies and misunderstandings in some situations. Generally handles complex language well and understands detailed reasoning.

6 Competent User. Has generally effective command of the language despite some inaccuracies, inappropriacies and misunderstandings. Can use and understand fairly complex language, particularly in familiar situations.

5 Modest User. Has partial command of the language, coping with overall meaning in most situations, though is likely to make many mistakes. Should be able to handle basic communication in own field.

4 Limited User. Basic competence is limited to familiar situations. Has frequent problems in understanding and expression. Is not able to use complex language.

3 Extremely Limited User. Conveys and understands only general meaning in very familiar situations. Frequent breakdowns in communication occur.

2 Intermittent User. No real communication is possible except for the most basic information using isolated words or short formulae in familiar situations and to meet immediate needs. Has great difficulty understanding spoken and written English.

1 Non User. Essentially has no ability to use the language beyond possibly a few isolated words.

0 Did not attempt the test. No assessable information.

Modules

A   Physical Sciences and related disciplines

B   Biological Sciences and related disciplines

C   Humanities and Social Sciences and related disciplines

**APPENDIX H: THE SCORING PROFILE RECOMMENDED BY FARHADY
et al. (1995)**

SPECIAL NOTE

THIS ITEM IS BOUND IN SUCH A

MANNER AND WHILE EVERY

EFFORT HAS BEEN MADE TO

REPRODUCE THE CENTRES, FORCE

WOULD RESULT IN DAMAGE

evidence given; problems of organization interfere.

	11-06	Shaky or minimally recognizable introduction; organization can barely be seen; severe problems with ordering of ideas; lack of supporting evidence; conclusion weak or illogical; inadequate effort at organization.
	05-01	Absence of introduction or conclusion; no apparent organization of body; sever lack of supporting evidence; writer has not made any effort to organize the composition (could not be outlined by reader).
ical elopment Jeas content	20-18	Essay addresses the assigned topic; the ideas are concrete and thoroughly developed; no extraneous material; essay reflects thoughts.
	17-15	Essay addresses the issues but misses some points; ideas could be more fully developed; some extraneous material is present.
	14-12	Development of ideas not complete or essay is somewhat off the topic; paragraphs aren't divided exactly right.
	11-06	Ideas are incomplete; essay does not reflect careful thinking, or was hurriedly written; inadequate effort in area of content.
	05-01	Essay is completely inadequate and does not reflect college-level work; no apparent effort to consider the topic carefully.
icture	20-18	Native-like fluency in the structure of English; correct use of relative clauses, prepositions, modals, articles, verb forms, and tense sequencing; no fragments or run-on sentences.
	17-15	Advanced proficiency in the structure of English; some structure problems don't influence communication, although the reader is aware of them; no fragments or run-on sentences.
	14-12	Ideas are getting through to the reader but structure problems are apparent and have a negative effect on communication; run-on sentences

	11-06	Numerous serious structure problems interfere with communication of the writers' ideas; structure review of some areas clearly needed; difficult to read sentences.
	05-01	Severe structure problems interfere greatly with the message; reader can't understand what the writer was trying to say; unintelligible sentence structure.
Punctuation, Spelling, & Mechanics	20-18	Correct use of English writing conventions, left and right margins, all needed capitals, paragraphs indented, punctuation and spelling very neat.
	17-15	Some problems with writing conventions or punctuation; occasional spelling errors; left margin correct; paper is neat and legible.
	14-12	Uses general writing conventions but has errors; spelling problems distract reader; punctuation errors interfere with ideas.
	11-06	Serious problems with format of paper; parts of essay not legible; errors in sentence final punctuation: unacceptable to educated readers.
	05-01	Complete disregard for English writing conventions; paper illegible, obvious capitals missing, no margins, severe spelling problems.
Style & Quality of Expression	20-18	Percise vocabulary usage; use of parallel structures, concise, register good.
	17-15	Attempts variety; good vocabulary; not wordy; register OK; style fairly concise.
	14-12	Some vocabulary misused; lacks awareness of register; may be too wordy.
	11-06	Poor expression of ideas; problems in vocabulary; lacks variety of structure.
	05-01	Inadequate use of vocabulary; no concept or register or sentence variety.

Component	Point	Operational Statement
Organization: introduction body, and conclusion	20-18	Appropriate title, effective introductory paragraph, topic is stated, leads to body; transitional expressions used; arrangement of material shows plan (could be outlined by reader); supporting evidence given for generalizations; conclusion logical and complete.
	17-15	Adequate title, introduction, and conclusion; body of essay is acceptable but some evidence may be.. lacking. Some ideas aren't fully developed; sequence is logical but transitional expressions may be absent or misused.
	14-12	Mediocre or scant introduction or conclusion; problems with the order of ideas in body; the generalizations may not be fully supported by the

APPENDIX I: COMPOSITION SCORES FOR THE PILOT STUDY

| Description	Exposition	Argument
5.25	5.00	4.50
6.25	6.00	5.75
5.75	6.00	5.50
4.50	5.25	4.50
4.75	5.75	4.75
5.00	5.75	4.50
6.25	5.00	3.75
3.25	4.25	3.25
8.00	8.75	8.25
6.00	6.00	6.00
5.75	5.50	4.50
5.25	5.75	4.00

APPENDIX J: SAMPLE ESSAYS IN FOUR DIFFERENT MODES

Topic 1

In our modernized world, industrial revolution, technological progresses and scientific speculations have drastically changed man's life on Earth. Every day we read or listen the hard boiled news of the everlasting competitions among the World super powers which aim at outstriping their rivals in the era of technology. Every day a considerable outlay is specified to space projects, exploration of space, sending sattelites, shuttles and aircrafts to discover unknown recesses of existence beyond our range of living, yet ironically finding a new form of life on other planets exacts a high price on living beings on this planet! Shouldn't we really think about the future life of this planet? Where are we going?

The exploitation of nature, transformation of echosystem, and the damage to the life of many living organisms which interrelatedly are connected to our own life, presage a very gloomy future for us, the lords of all beings. I'm not against scientific projects which bring better conditions of life to man's life; however, I can not stand to see that through these projects man carelessly and blindly destroys his own life. Almost everybody knows about the gradual exhustion of ozone layer and the overheating of temperature which are exclusively due to man's interference in normal cycles of life. Instead of a means of making a better life for mankind, science has become his master which leads him toward the obliteration of all signs of life on earth.

Furthermore, what is the use of spending all this money

to discover forexample, another forms of life in other planets?
Probably exploitation of human beings does not the quench
the thirst of worlds' super powers so that they aim at
rulling over those poor unknown beings too, if there are any!
It is astonishing why the developed Countries who beautifully
lecture on humanity, forget about miseries of human beings
on this planet. Many innocent children suffer from hunger,
diseases, poverty, and illiteracy. Part of the huge sums
of money wasted for space projects, can bring a little
bit sunshine to their miserable life. Life is beautiful,
and it is everybody's natural right to enjoy his being alive.
Yewis thelem, to the majority of human race happiness is only a
beautiful tale. Hunger & poverty is the underlying cause of the
destruction of forests and natural resources. Modern technology
must aid man in the direction of preventing this unnecenan
destruction of nature; however, unfortunately it has become
itself an active agent hastening the end of life on Earth.

I believe scientific space projects have accelerated
our moral degeneracy caused by industrialisation and
mechanization of our life. Now man has become machine like,
he moves farther and farther away from humanity, faith, and
God. Why the representatives of human race have become
Vladimirs, Strougons, the hairy apes, and the dangling men?
where are those high ideal heros of Renaissance thinkers
who believed man is the center of universe? Yes, he can be the
center of universe, yet through his own thoughtlessness, he
would also put an end not only to his own life, but to the

to the existence of this planet. How beautifully Alexander Pope reminds us of our state in nature when he calls man "Great lord of all things, yet a prey to all."

Argue the advantages and disadvantages of
commercial advertising in the
Machine life is a force or dialectic, if we are afraid
and be far from that, its hurt will be higher. Now,
machine is extending in Euroupian countries.
The produced merchandise are known by commercial advertisements
that originate of technology because every thing that
hasn't been known, it needs to be introduced. Introducer of
production in this world that is very large and speedy
is commercial advertisments because product without
sale is equel with rubbish and, country without
product is equel with dead economy therefore,
we can't consider like this economic for a country.
I think at commercial advertisments people's aims'
must be considered, if not, it will be misled.

(14)

What could be one's notion of a metropolis? Perhaps it could be a center of commercial transactions, industrial poles, ministral departments, and a complete set of all facilities one can think about or to sum it up, the big heart of the entire country, so to speak. To me, however, Tehran is the assemblage of contrasts..

In the morning, when the train is within the suberbs of the city you are surprised to see the dark greyish haze surrounding the city and making it almost invisible as if to be under a gloomy spell. When you are within the spell, you are able to see the appearance of the city. In down down during the rush hours, innummerable cars and busses like busy ants are crammed one after another in long queues, waiting for the traffic jam to subside so that they would be able to go toward their destination. While you are waiting to move, you look at the sky; it is no longer blue though there is no cloud. It is as if the old blue color of the ^old quilt of sky has ^been faded away under the sun, so it is only a pale grey now. Of a sudden you find yourself uneasy to breathe, an acrid, terrible, nasty smell burns your nostrils and throat till you feel your endurance is at its end, however, at last the car starts to move and gradually you get rid of this part of the city. Here and there monstrous, huge buildings have risen indicating to be of some commercial or industrial importance. The weather beaten regimented appartments with

their dirty dark colored walls are good measures to remind you of the intensity of air pollution. Sometimes very old buildings, oddly detached from the new appearance of the city attract your attention as the last reminiscences of the old day of this city. However, infront of you, toward the north of the city, there extends the stretch of high mountains; you, absorbed by the scenery, move forward the beautiful mountains.
toward

     As you move in the direction of the northern side of the city, you notice the gradual transfiguration of the scene. The number of cars especially old, shabby ones subside and instead gorgeous, fashionable automobiles move gracefully this way or that way. Now you are fascinated by the sight of fantastic, modern restaurants, cafés, shops and bazaars. Every now and then on the right or left side of you beautiful parks and plots of grass catch your eyes. As you walk along the alleyways, you are surprised at the wonderful, artistic architecture of the houses some of which resemble more or less to European buildings rather than Iranian houses. Instead of disgusting smell of poisonous gases of automobiles; you find yourself in an atmosphere filled with pleasant odors and colors. Now the gorgeous mountains are almost at reach. Absorbed in the beauties of this part of the city, you dispatch the bitter memories of your previous experience, come to the conclusion that Tehran is a beautiful city.

SPECIAL NOTE

THIS ITEM IS BOUND IN SUCH A

MANNER AND WHILE EVERY

EFFORT HAS BEEN MADE TO

REPRODUCE THE CENTRES, FORCE

WOULD RESULT IN DAMAGE

magine that you are in the 25th century.

y longing time, life will be very progressive and comfortable.
he building will increase and that time world will be crowded.
e can find many cars and vehicles. strange
ariety of food will become more and making foods
ill be computerized. at last we'll have a new world
or example spaceship are replaced instead of mechine
nd kindman by using spaceship de her/his work much
aster.
inally the warfare equipment will be more. If we use
them in a best way. we will have a good life but, if not
uman and humanity will be destroyed (ended).

The Sound of Music by Robert Wise

There was a captain who had lost his wife and had 7 children. He needed somebody to take care of his kids but anybody who came to do this duty, would go after a while. At last a nun came to his house and as usual the children begin to bother her. She tolerated them and ignored their faults. When the hero was in mission, this nun tried to teach his children musical notes which they deprived of after their mother's death. In his return, the hero and the princes who accompanied him were pleased by these naughty chidren's music so that the captain himself began singing. The nun came back to nunery but the kids went after her. Nun herself found that she would like to come back. She came back and at the same time the abovementioned woman went since she understood those 2 person fell in love with each other. The hero and heroin married and war between their country and Germany began (broke out). Since the hero was against common conditions of these days, they planned to escape. They went participated in a race based on music which held on that special day. After that they went to nunery and nuns help them go away.

SPECIAL NOTE

ITEM SCANNED AS SUPPLIED
PAGINATION IS AS SEEN

Write the story of someone who rose from

an unknown background and humble beginning

to become successful and famous. Narrate the

main events of his/her life story in the

order in which they happened

I want to say about Alibou who is a son that very

very people his father dead when he

was very child young gun his mother became

this mother sarvent in

the house of michrian but she was sick and can't

fallew work. Therefore, she have to maried

with old man who didn't let him turn the stone

How can we make teaching and learning in Iranian universities more effective? Expl your ideas

Education in Iran faces many problem especially when we consider universities. First the entering criteria are not based on the competence and capasitance of pu Another point which should be taken consideration is that preffessors are not qualified properly or at least they don't teach according to their specialities due to some deficiedcies. that we have in the respect. The third and the last problem is the universities themselves. They are not located in appropriate places and there are not enough facilities such as spacious rooms, provided in especi training instruments, etc.

In order for us to satisfy these problems, it is necessary to do some basic changes One, to change the system of the enterance examination. Next, to employ pretessors on the basis of their abiliti and the last but not the least pro viding adequate facilities for training qulified students and to encourage the students to be research rather than reader itself.

(6) 4.

ج:

Considering the situation in Iran, explain how a forign language should be learned

There are several reasens for learning of Iranian's student. The most important is reading the book and magazient and every thing that They can read. for comperhenti The secone stepe is listening to the tape for an hour or half an hour. Finally, They memory part of the subject everyday in to thair mind and the most effect thing for as the sameday learning is practis and practic

Bibliography

ACKERMAN, T. A. and SMITH, P. L. (1988) A comparison of the information provided by essay, multiple-choice, and free-response writing tests. *Applied Psychological Measurement*, 12, pp.117-128.

ALDERSON, C. (1979) The cloze procedure and proficiency in English as a foreign language. *TESOL Quarterly*, 13 (2), pp. 219-227.

ALDERSON, C. (1980) Native and non-native speaker performance on cloze tests. *Language learning*, 30 (2) pp. 59-67.

ALDERSON, C. (1983) The cloze procedure and proficiency in English as a foreign language, *in* Oller, J. W. (ed) *Issues in language testing research*. Rowley, Mass.: Newbury House, pp. 205-217.

ALLAEI, S. and CONNOR, U. (1991) Using performative assessment instrument instruments with ESL student writers, *in* HAMP-LYONS, L. (ed) *Assessing second language writing in academic context*. Norwood, NJ: Ablex. 227-240.

ANDREWS, R. (1995) *Teaching and learning argument*. London: Cassell.

APPLEBEE, A. (1982) Writing and learning in school setting, *in* NYSTRAND, M. (ed) *What writers know: The language process and structure of written discourse*. New York: Academic Press.

ARTHUR, B. (1979) Short-term changes in EFL composition skills, *in* YORIO, C.A., PERKINS, K. and SCHACHTER, J. (eds) *On TESOL*. TESOL. pp.330-342.

BACHMAN, L. F. (1985) Performance on cloze tests with fixed-ratio and rational deletion. *TESOL Quarterly*, 19 (1) pp. 535-556.

BACHMAN, L. F. (1990) *Fundamental considerations in language testing*. Oxford: Oxford University Press.

BACON, S. and FINNEMAN, M. (1992) Sex differences in self-reported beliefs about foreign-language learning and authentic oral and written input. *Language Learning*, 42, pp. 471-495.

BAIN, A. (1890) *English composition and Rhetoric*. 2nd edition. New York: D. Appleton and Company.

BALLARD, B. (1981). Language is not enough: Responses to the academic difficulties of overseas students. *Proceedings of the 2nd Study Skills Conference,* (pp. 116-128).

BALLARD and CLANCHY, (1991) Assessment by misconception: cultural influences and intellectual traditions. *in* HAMP-LYONS, L. (ed) Assessing second language writing in academic contexts. Norwood,N.J.:Ablex.

BARRA, A. (1993) Writing reflection and research may increase teacher understanding. English Teaching Forum, 31 (3) pp. 14-17.

BASHAM, C., and KWACHKA, P. (1991) Reading the world differently: A cross-cultural approach to writing assessment, *in* HAMP-LYONS (ed), *Assessing second language writing in academic contexts.* Norwood, N.J.: Ablex.

BHATIA, V.K. (1993) *Analysing genre.* London: Longman.

BIBER, D. (1988) *Variation across speech and writing.* Cambridge: Cambridge university Press.

BLOOMFIELD, L. (1933) *Language.* New York: Henry Holt.

BOWEN , J.D. ,MADSEN , H. and HILFERTY, A. (1985) *TESOL, techniques and procedures.* Cambridge: Newbury House.

BOYL, J. (1987) Sex differences in listening vocabulary. *Language Learning*, 37, pp. 273-284.

BRADDOCK, R., LLOYD-JONES, R., and SCHOER, L. (1963). *Research in written composition.* Urbana, IL: National Council of Teachers of English.

BRAINE, G. (1988). A reader reacts. *TESOL Quarterly*, 22, pp.700-702.

BRELAND, H.M and GAYNOR, J.L. (1979) A comparison of direct and indirect assessment of writing skill. *Journal of Educational Measurement*, 16, pp. 119-128.

BRIDGEMAN, B. and CARLSON, S. (1983) A survey of academic writing tasks required of graduate and undergraduate foreign students. *TOEFL research report No.15.* Princeton, New Jersey: Educational Testing Service.

BRIERE, E. (1966). Quantity before quality in second language composition. *Language Learning*, 16, 141-151.

BRITTON, J. et al. (1975). *The development of writing abilities (11-18)*. London: Macmillan.

BROOKS, C. and WARREN, R.P (1952) *Fundamentals of Good Writing: a handbook of modern rhetoric*. London: Dennis Dobson.

BROSSELL, G. (1983). Rhetorical specification in essay examination topics. *College English, 45,* 165-174.

BROSSELL, G. and HOETKER ASH, B. (1984). An experiment with the wording of essay topics. *College Composition and Communication, 35*(3), 423-25.

BROWN, D. (1983) Conversational cloze tests and conversational ability, *ELT Journal,* 37(2), pp.158-161.

BROWN, J.D. (1983) A closer look at cloze: Validity and reliability, *in* OLLER, W. J. Jr. (ed) *Issues in language testing research*. Newbury: House Publishers.

BROWN, J.D. (1988) *Understanding research in second language learning*. Cambridge: Cambridge university Press.

BROWN, J.D. (1991) Do English and ESL faculties rate writing samples differently? *TESOL Quarterly,* 25, pp. 587-603.

BROWN, J. D. (1996) *Testing in Language Programmes*. NJ: Prentice Hall.

BROWN, J.D. and BAILEY, K.A. (1984) Categorical instrument for scoring second language skill. *Language Learning,* 34, pp. 21-41.

BROWN, K. and HOOD, S. (1989) *Writing matters: Writing skills and strategies for students of English*. Cambridge: Cambridge University Press.

BRUNER, J. (1986) *Actual minds, possible worlds*. Cambridge: Harvard University Press.

BRUNER, J. (1987) Life as Narrative. *Social Research,* 54 (1), pp.11-32.

BRUNER, J. (1990) *Acts of meaning*. Cambridge, Mass.: Harvard University Press.

BURSTAL, C. (1975) Factors affecting foreign-language learning: a consideration of some relevant research findings. *Language Teaching and Linguistics Abstracts,* 8, pp. 105-125.

BUTLER, C.(1985) *Statistics in Linguistics*. New York: Basil Blackwell.

CAIRNS, W. B. (1902) *The Forms of Discourse*. Boston: Ginn and Company.

CANALE, M. (1988) The measurement of communicative competence. *Annual review of Applied Linguistics*, 8, pp. 67-84.

CANALE, M., BELANGER, M. and FRENETTE, N. (1988) Evaluation of minority student writing in first and second languages, *in* FINE, J. (ed) *Second language discourse: A textbook of current research*. Norwood, NJ: Ablex, pp. 147-165.

CARLSON, S. et al. (1985). *Relationship of admission test scores to writing performance of native and nonnative speakers of English* (TOEFL Research Report 19). Princeton, NJ: Educational Testing Service.

CARROLL B. J. and HALL, P. J. (1985). *Make your own language tests*. Oxford: Pergamon Press.

CAST, B.M.D. (1939) The efficiency of different methods of marking English compositions. *British Journal of Educational Psychology*, 9(1), pp. 251-269.

CHASTAIN, K. (1988) *The development of modern language skills: theory to practice*. 3rd ed., Florida: Harcourt Brace Jovanovich Publications.

CHIHARA, T. et al.(1979) Are cloze items sensitive to constraints across sentences? *Language Learning*, 27 (1), pp. 63-71.

CHISTE, K. and O'SHEA, J. (1988) Patterns of question selection and writing performance of ESL students. *TESOL Quarterly*, 22 (4), pp. 281-284.

COHEN, A. (1980) *Testing Language Ability in the Classroom*. Newbury: Newbury House Publishers.

COHEN, A. (1994) *Assessing language ability in the classroom*. 2nd ed., Boston, mass: Heinle & Heinle Publishers.

COOPER, C.R. (1977) Holistic evaluation of writing, *in* Cooper, C.R. and Odell, R. (ed) *Evaluating writing: Describing, measuring, judging*. Urbana, Ill.: National council of Teachers of English.

COOPER, J. (1979) *Think and link: An advanced course in reading and writing skills.* London: Edward Arnold.

COPE, B. and KALANTZIS, M. (1993*) The power of literacy: A genre approach to teaching writing.* London: The Falmer Press

CORTAZZI, M. (1993) *Narrative Analysis.* London: The Falmer Press.

CORTAZZI, M. (1994) Narrative analysis. *Language Teaching,* 27, pp. 157-170

CRIPER, C. and DAVIES, A. (1988). *Final report of the ELTS Validation Project* (ELTS: Research Reports 1). London: The British Council.

CROWHURST, M. (1980) Syntactic complexity and teachers' quality ratings of narrations and arguments. *Research in the teaching of English,* 14 (3), pp.223-231.

CROWHURST, M. and PICHE, G. L. (1979). Audience and mode of discourse effects on syntactic complexity in writing at two grade levels. *Research in the Teaching of English,* 13, pp.101-110.

CRYSTAL, D. (1987) *The Cambridge encyclopaedia of language.* Cambridge: Cambridge University Press.

CUMMING, A. (1990a) Expertise in evaluating second language compositions. *Language Testing,* 7, pp. 31-51.

CUMMING, A. (1990b) Metalinguistic and ideational thinking in second language composing. *Written communication,* 7, pp. 482-511.

CUMMING, A. (1998) Theoretical perspectives on writing. *Annual Review of Applied Linguistics,* 18, pp. 61-78.

D'ANGELO, F. J. (1976) Modes of discourse, in TATE, G. (ed) *Teaching composition: 10 bibliographical essays.* Fort Worth, TX: Christian University Press.

DARNELL, D. K. (1968) The development of an English language proficiency test of foreign students using a cloze-entropy procedure, *ERIC ED 024039.*

DALE, E. and CHALL, J.S. (1948) A formula for predicting readability. *Education Research Bulletin*, 27.

DAVIES, A. (1965) Proficiency in English as a second language. PhD Thesis, University of Birmingham.

DEAN, M. (1988) *Write it: Writing skills for intermediate learners of English, Teacher's book.* Cambridge: Cambridge University Press.

DIEDERICH, P. B. (1964) Measurement of skill in writing. School Review., 54, 584-592.

DIEDERICH, P.B. (1965) Grading and measuring, *in* JEWETT, A. and BISH, C. (eds), *Improving English Composition. Washington*, DC: National Education Association.

DIEDERICH, P. B. (1966) How to measure growth in writing ability. *English Journal*, 55, pp. 435-449.

DIEDRICH, P.B. (1967) Cooperative preparation and rating of essay tests. *English Journal*, 56, pp. 573-584.

DIEDRICH, P.B. (1974) *Measuring growth in English.* Urbana, Ill: National Council of Teachers of English.

DIEDERICH, P. B., FRENCH, J. and CARLTON, S. (1961). Factors in judgements of writing ability. *ETS Research Bulletin RB-61-1S.* Princeton, N.J.: Educational Testing Service.

DONLEY, M. (1978) Marking advanced essays. *English Language Teaching Journal*, 32 (2), pp. 115-118.

DUDLEY-EVANS, T. (1988) A consideration of the meaning of "discuss" in examination questions, *in* P. Robinson (Ed.), *Academic writing: Process and product: ELT Documents No. 129,.* Hong Kong: Modern English Publications and The British Council, *pp. 47-52.*

DUDLEY-EVANS, T. (1997) Genre: How far can we, should we go? *World Englishes*, 16 (3), pp. 351-358.

DUNLOP, I. (1969) tests of writing ability in English as a foreign language. *ELT Journal.*

EL-BACHA, N.S. (1997) Patterns of lexical cohesion in EFL texts. Ph.D. Thesis, University of Leicester.

ELBOW, P. (1973) *Writing Without Teachers*. New York: Oxford University Press.

ELLIS, R. (1994) *The study of second language acquisition*. Oxford: Oxford University Press.

EVOLA, J., MAMER and LANTZ (1980) Discrete point versus global scoring for cohesive devices, *in* OLLER, J. W. Jr. and PERKINS, K. (eds), Research in language testing. Mass: Newbury House Publishers Incorporated, pp. 177-183.

FAIGLEY, L. , et al. (1985) *Assessing writers' knowledge and processes of composing*. Norwood, NJ: ablex.

FAIGLEY, L. (1986). Competing theories of process: A critique and a proposal. *College English,* 48, pp. 527-542.

FARHADY, H. JAFARPOOR, A. and BIRJANDI, P. (1995) *Testing Language skills; From theory to practice*. Tehran: SAMT publication.

FARZANEHNEZHAD, M. (1992) Cohesive devices and writing proficiency. M.A. thesis, Tehran University.

FELDT, L. S. and BRENNAN, R. L. (1989) Reliability, *in* LINN, R. L. (ed) *Educational measurement*. New York: ACE/Macmillan.

FLAHIVE, D. and SNOW, B.G. (1980) Measures of syntactic complexity in evaluating ESL compositions, *in* OLLER, J. W. Jr. And PERKINS, K. (eds) *Research in language Testing*. Mass: Newbury House Publishers Incorporated, pp. 171-176.

FRANK, M. (1990) *Writing as thinking: A guided process approach*. New Jersy: Prentice Hall

FREEDMAN, S. W. and CALFEE, R. C. (1983) Holistic assessment of writing: Experimental design and cognitive theory, *in* MOSENTAL, P., TAMOR, L., and WALMSY, S. (eds) *Research in writing : Principles and methods*. Longman, pp. 75-98.

FREEDMAN, A. and PRINGLE, I. (1981) Writing in the college years: some indices of growth. *College composition and communication*, 31, pp. 311-324.

FREEDMAN, A. and PRINGLE, I. (1984) Why students can't write arguments? *English in Education*, 18(2), pp. 73-84.

FREEDMAN, A. and PRINGLE, I. (1989) Contexts for developing argument, *in* ANDREWS, R. (ed) *Narrative and argument*. Milton Keynes: Open University Press, pp. 73-84.

FRIES, C. (1945), *Teaching and learning English as a second language*. Ann Arbor: University of Michigan Press.

FROWE, I. (1989) Arguing, *in* Andrews, R. (ed) *Narrative and Argument*, Milton Keynes: Open University Press, pp. 55-63.

GARDNER, R. and LAMBERT, W (1972) *Attitudes and motivation in second language learning*. Rowley, Mass.: Newbury house.

GERANPAYEH, A. (1994) Are score comparisons across language proficiency test batteries justified? An IELTS-TOEFL comparability study. *Edinburgh Working Papers in Applied Linguistics*, 5, pp.50-65.

GIOVANNINI, G. (1943) "The four forms of composition." *Dictionary of World Literature*. New York: The Philosophical Library, pp.117-118.

GODSHALK, et al., (1965) *The measurement of writing ability*. New York: College Entrance Examination Board.

GRABE,W. and KAPLAN, R. B. (1996) *Theory and practice of writing*. Longman: New York

GUILFORD, J. P. and FRUCHTER, B. (1978) *Fundamental Statistics in Psychology and Education*, New York: McGraw-Hill.

HALE, G. et al. (1996) A study of writing tasks assigned in academic degree programs. *TOEFL Research Report #44*. Princeton, NJ: Educational Testing service.

HALLIDAY, M. A. K. (1989) *Spoken and written language*. Oxford University Press.

HAMP-LYONS, L. (1986a). *Assessing second language writing in academic settings*. Ph.D. Thesis, University of Edinburgh, Scotland.

HAMP-LYONS, L. (1986b). No new lamps for old yet, please. *TESOL Quarterly*, 20(4), pp. 790-796.

HAMP-LYONS, L. (1987). Performance profiles for academic writing, *in* K. BAILEY, K., CLIFFORD, R. and DALE, E. (eds) *Language Testing Research*. pp. 78-92. Monterey, Cal.: Defense Language Institute.

HAMP-LYONS, L.(1988a) Proficiency, profiling and L2, in PORTER, D. (ed), ELTS: *Research reports (ii): Proceedings of the Invitational Conference on the ELTS Validation Project*. London: British Council.

HAMP-LYONS, L. (1988b). The product before: Task-related influences on the writer, *in* ROBINSON, P. (ed), *Academic writing: Process and product*. ELT Documents No. 129, (pp.35-46*)*. London: Macmillan/British Council.

HAMP-LYONS, L. (1989 a) English as a Second Dialect, Bilingual and ESL writers' essay test strategies: Pragmatic failure in a key literacy event. Paper presented at the NCTE Conference, Baltimore.

HAMP-LYONS, L. (1989b). *Preparing for the (TOEFL) Test of Written English*. New York: Newbury House/Harper & Row.

HAMP-LYONS, L.(1989 c). Raters respond to rhetoric in writing, *in* DECHERT, H. and RAUPACH, G. (eds), *Interlingual processes*, Tubingen: Gunter Narr Verlag. pp. 229-244.

HAMP-LYONS, L. (1990) Second language writing: Assessment issues, *in* KROLL, B. (ed) *Second language writing: research insights for the classroom*. New York: Cambridge University Press, pp.69-87.

HAMP-LYONS, L.(1991). Reconstructing "academic writing proficiency, *in* HAMP-LYONS, L. (ed), *Assessing second language writing in academic contexts*. Norwood, N.J.: Ablex.

HAMP-LYONS, L and HEASLEY, B (1987) Survey review: Textbooks for teaching writing at the upper levels. *English Language Teaching Journal*, 38 (3), pp. 209-216.

HAMP-LYONS, L and HENNING, G. (1991) Communicative writing profiles: An investigation of the transferability of a multiple-trait scoring instrument across ESL writing assessment contexts. *Language Learning*. 41 (3), pp. 209-216.

HAMP-LYONS, L. and KROLL, B. (1997) TOEFL 2000- writing: composition, community, and assessment, *TOEFL Monograph Series*. Princton, NJ: Educational Testing service.

HAMP-LYONS, L. and MATHIAS, S. P. (1994) Examining expert judgements of task difficulty on essay tests. *Journal of Second Language Writing*, 3(1), pp. 49-68.

HAMP-LYONS, L. and MATHIAS, S. P. (1991) The difficulty of difficulties: Prompts in writing assessment, *in* SARINEE, A. (ed) *The regional language centre seminar on language testing and language programme evaluation*; *Current developments in language testing, 1990*. pp.58-76.

HAMP-LYONS, L., and REED, R. (1990). *Development of the new Michigan Writing Assessment*, Report to the College of LS & A, University of Michigan. Ann Arbor, MI: English Composition Board.

HARPIN, W. (1976) *The second "R": Writing development in the Junior school.* London: Allen & Unwin.

HARRIS, D. P. (1969) *Testing English as a Second Language*, McGraw-Hill.

HARRIS, J. (1993) *Introducing writing*. London: Penguin Books Limited.

HARTOG, P. (1936) English composition at the School Certificate Examination: and the "Write anything about something for anybody" theory. *in* SADLER, M. et al.(eds) *Essays on examinations*. London: Macmillan.

HARTOG, P. et al. (1941) *The marking of English essays*. London: Macmillan

HATCH, E. and LAZARATON, A. (1990*) The research manual: designand statistics for applied linguistics*. New York: Newbury House Publication.

HATCH, E. and FARHADY, H. (1982) *Research design and statistics for applied linguistics*. Rowley, Mass.: Newbury House.

HAYWARD, M. (1988) The rhetoric of essay prompts. Paper presented at the annual meeting of the conference on college composition and communication (39th, St. louis, March 17-19).

HAYWARD, M. (1990) Evaluation of essay prompts by nonnative speakers of English. *TESOL Quarterly*, 24 (4), pp. 753-758.

HEATON, J. B. (1975) *Writing English language tests*. London: Longman.

HEATON, J. B. (1990) *Classroom testing*. London: Longman.

HEDGE, T. (1985) *In the picture*. Walton-on-Thames: Nelson.

HEDGE, T. (1988) *Writing*. Oxford and New York: Oxford University Press.

HENNING, G. (1987) *A guide to language testing*. New York: Newbury House.

HESSE, D.(1989) Stories in essays, essays as stories, *in* Anderson, C. (ed) *Literary nonfiction*. Carbondale: Southern Illinois University Press.

HEYES, M. et al. (1986) *Starting statistics in psychology and education*. Oxford: Oxford University Press.

HIROKOWA, K., and SWALES, J. M. (1986). The effects of modifying the formality level of ESL composition questions. *TESOL Quarterly*, 20, pp. 343-345.

HIRSCH, E. (1977) *The philosophy of composition*. Chicago, IL.: The University of Chicago.

HIRSCH, E. and HARRINGTON, D. P. (1981) Measuring the communicative effectiveness of prose. *in* FREDRIKSON, C. and DOMINIC, J. (eds) *Writing: The nature, development and teaching of written communication. Vol.2: Writing: Process, development and communication*. Hillsdale, N. J. :Erlbaum, pp.189-207.

HOEY, M. (1983) *On the surface of Discourse*. London: Allen and Unwin.

HOLLAND, R.M. (1976). Piagetian theory and the design of composing assignments. *Arizona English Bulletin, 19,* pp.17-22.

HOOVER, M. and POLITZER, R. L. (1980) Bias in composition tests with suggestion for a culturally appropriate assessment technique, *in* FARR WHITEMAN, M. (ed) *Writing: The nature, development and teaching of written communication. Vol.1: Variation in writing* . Hillsdale, N.J.: Erlbaum. pp. 197-204.

HOPKINS, A. and DUDLEY-EVANS, T. (1988) A genre-based investigation of the discussion sections in articles and dissertations. *English for Specific Purposes*, 7, pp.113-121.

HOROWITZ, D. M. (1986a) Essay examination prompts and the teaching of academic writing. *English for academic writing*, 5 (2), pp. 107-120.

HOROWITZ, D. M. (1986b) What professors actually require: Academic tasks for the ESL classroom. *TESOL Quarterly*, 20 (3), pp. 445-462.

HUFF, D.(1954) *How to lie with statistics.* Middlesex: Penguin Books Ltd.

HUMBURG, T.J. (1984) Holistic evaluation of ESL compositions: can it be validated objectively? *TESOL Quarterly*, 18, pp.87-107.

HOMBURG, T.J. (1989) Holistic evaluation of ESL compositions. *TESOL Quarterly*, 18, pp. 87-107.

HUNT, K. (1965) *Grammatical structures written at three grade levels.* Urbana, Ill.: national Council of teachers of English.

INGRAM, E. (1977) Basic concepts in testing. in Allen & Davies (eds).

IRVIN, P., ATAI, P. and OLLER, J. (1974) cloze, dictation and the test of English as a foreign language, *Language Learning.* 24, pp.245-252.

JACOBS, H.J. et al. (1981) *Testing ESL composition: A practical approach.* Rowley, Mass: Newbury.

JACOBS, S. E. (1982). *Composing and coherence.* Washington, D. C.: Center for Applied Linguistics.

JAFARPUR, A. (1995) Is C-testing superior to cloze. *Language Testing,* 12 (2), pp.194-216.

JOHNS, A. (1988) *Another reader reacts. TESOL Quarterly,* 22 (4), pp.705-707.

JOLLY, D. (1984) *Writing Tasks; an authentic-task approach to individual writing needs.* Cambridge: Cambridge University Press.

JONZ, J. (1990) Another turn in the conversation: What does cloze measure? *TESOL Quarterly.* 24 (1), pp. 245-252.

KAMEEN, P. (1979) Syntactic skill and ESL writing quality, in YORIO et al. (eds), *On TESOL '79: the learner in focus.* Washington: TESOL. pp. 343-350.

KAPLAN, R. B. (1966) Cultural thought patterns in intercultural education. *Language Learning,* 16, pp. 1-21.

KAPLAN, R. (1987). Cultural thought patterns revisited. in CONNOR. U. and KAPLAN, R.B. (eds), *Writing across languages: Analysis of L2 text.* Reading, Mass: Addison-Wesley. (pp. 9-22).

KELLER-KOHEN, D. and WOLFE, A. (1987) *Extending writing in the college of literature, science and the arts: Report on a faculty survey.* Ann Arbor: University of Michigan English Composition Board.

KESHAVARZ, M.H. (1996) Islamic Republic of Iran, Language policy, *in* DICKSON, P. and A. CUMMING, A.: Profile of Language Education in 25 Countries. Slaughi, NFER.

KIANY, G. R. (1996) Reliability and validity of cloze test. in CORTAZZI, M., RAFIK GALEA, S. and HALL, B. (eds) *Aspects of language teaching, learning and research methodology in the context of education,* Leicester: University of Leicester, pp.179-197.

KINCAID, G. L. (1953). Some factors affecting variations in the quality of students' writing. Ph.D. Thesis, Michigan State University, East Lansing, MI.

KINNEAVY, J. L. (1971) *A Theory of Discourse: The aims of discourse.* Englewood.

KLEIN-BRALEY, C. (1981) Empirical investigations of cloze tests. PhD Thesis, University of Duisburg.

KLEIN-BRALEY, C. (1985) A close-up on the C-test: A study in the construct validation of authentic tests. *Language Testing,* 2(1), pp.76-104.

KLEIN-BRALEY, C. and RAATZ, U. (1984) A survey of research on the C-test. *Language Testing,* 1(2), pp. 134-146

KODA, K. (1993) Task-induced variability in FL composition: language specific perspectives. *Foreign Language Annals,* 26 (3), pp. 332-346.

KRAPLES, A. R. (1990) An overview of second language writing process research, *in* KROLL, B. *Second Language writing: research insights for the classroom.* cambridge: Cambridge University Press.

KRESS, G. (1982) *Learning to write.* London: Rutlege and Kegan Paul.

KROLL, B. (1990a) *Second language writing: Research insights for the classroom.* Cambridge: Cambridge University Press.

KROLL, B. (1990b) What does time buy? ESL students performance on home versus class compositions, *in* KROLL, B. (ed) *Second language writing: research insights for the classroom*. Cambridge: Cambridge University Press, pp.140-154.

LABOV, W. (1972) *Language in the inner city*. Philadelphia, P.A.: University of Pennsylvania Press.

LABOV, W. and WALETSKY, J. (1967) Narrative analysis: Oral versions of personal experience, *in* HELM, J. (ed), *Essays on the verbal and visual arts*. Seattle: American Ethnological Society, pp. 12-44.

LABOV, W. and FANSHEL, D. (1977) *Therapeutic discourse: Psychotherapy as conversation*. New York: Academic Press.

LA BRANT, L. L. (1933) A study of certain language developments of children in grades four to twelve, inclusive. *Genetic Psychology Monographs*, XIV, pp. 387-491.

LARSEN-FREEMAN, D. (1978) An ESL index of development. *TESOL Quarterly*, 12, pp. 439-448

LEWIS, M. (1993) *The lexical approach*. Hove, England: Language Teaching Publications.

LITTLEFAIR, A.B. (1991) *Reading all types of writing: the importance of genre and register for reading development*. Milton Keynes: Open University Press.

LITTLEJOHN, A. (1991) *Writing, books 1 and 2*. Cambridge: Cambridge University Press.

LLOYD-JONES, R. (1977) Primary Trait Scoring, in COOPER, C. R. and L. ODELL. *Evaluating Writing: Describing, measuring, judging*. New York: State University of New York at Buffalo.

LONGACRE, R. (1976) *An anatomy of speech notions*. Lisse: Peter de Ridder.

LUNZER, E. A.(1968) *Development in learning: behavior: learning: education*. London: Staple Press

MADSEN, H.S. (1983) *Techniques in testing*. Oxford: Oxford University Press.

MAGNAN, S. S. (1985) Teaching and testing proficiency in writing: Skills to transcend the second language classroom, in OMAGGIO, (ed) *Proficiency, curriculum, articulation: The ties that bind*. Middle-Bury, Vt: Northeast Conference.

MATSUHASHI, A. (1982) Explorations in the real-time production of written discourse, *in* NYSTRAND, M. (ed) *What writers know: The language, process, and structure of written discourse*. New York: Academic Press, pp.269-290

McARTHUR, T.(1996) *The Oxford companion to the English Language*, Abridged edition. Oxford: Oxford University Press.

McDONOUGH, J. (1986) English for academic purposes: Aresearch base? *English for Specific Purposes*, 5(1), pp. 17-25.

MELLON, J. C. (1969) *Transformational sentence-combining: a method for enhancing the development of syntactic fluency in English composition*. NCTE Research Report No. 10. Champaign, Ill.: National Council of Teachers of English.

MOFFETT, J. (1968) *Teaching the universe of discourse*. Boston: Houghton Mifflin Company.

MORRIS, C. W. (1946) *Signs, Language, and Behavior*. Englewood Cliffs, N.J.: Prentice-Hall, Inc.

MUCHIRI, M. et al. (1995) Importing composition: Teaching and researching academic writing beyond North America. *College Composition and Communication*, 46, pp. 175-198.

MULLEN, K. A. (1977) Using rater judgement in the evaluation of writing proficiency for non-native speakers of English, *in* Brown, H., Yorio, C. A. and Crymes, R. A. (eds) *On TESOL '77: Teaching and learning English as a second language: Trends in research and Practice*, pp. 309-320.

MULLEN, K. A. (1980) Rater reliability and oral proficiency evaluation, *in* OLLER, J. and PERKINS, K. (eds) *Research in language testing*, pp. 91-101.

MUNRO, L.A. (1991) A comparative study of different methods of evaluating assessment writing. M.Ed. Thesis, University of Edinburgh.

NELSON, G. L. and MURPHY, J. M. (1992) Writing groups and and the less proficient ESL students. *TESOL Journal*, 2(2), pp.23-26.

NEMATI, M. (1993) The role of word selection in writing. MA Thesis. Tarbiat Modarress University, Tehran.

NEMATI, M. (1996) The effects of discourse on EFL/ESL writing performance, *in* CORTAZZI, M., RAFIK GALEA, S., and HALL, B. (eds) *Aspects of language*

teaching, learning and research methodology in the context of education, Leicester: University of Leicester, pp. 162-167.

O'DONNEL, W. (1968) *An investigation in the role of language in a physics examination.* Monograph No. 7. Moray House Publications. Edinburgh: Oliver & Boyd.

OLLER, J. W. Jr. (1979) *Language Tests at School: A Pragmatic Approach.* New York: Longman Group, Ltd.

OLLER, J. W. Jr. (1983) *Issues in language testing research.* Rowley, Mass.: Newbury House.

OSBURNE, A. and MULLING, S. (1994) Essay prompts and the ESOL students. Issues in Applied Linguistics, 5 (1), pp.103-109.

PARK, Y. M. (1987) The influence of the task upon writing performance. Ph.D. Thesis. University of Illinois, Urbana-Champaign.

PAULSTON, C.B. and BRUDER, M.N. (1976) *Teaching English as a second language: Techniques and procedures.* Cambridge, Mass.: Winthrop Publishers, Inc.

PEACOCK, C. (1986) *Teaching writing: A systematic approach.* Worcester: Billing and Sons Limited.

PEERS, I. (1996) *Statistical analysis for education and psychology research.* London: The Falmer Press.

PERKINS, K. (1980) Using objective methods of attained writing proficiency to discriminate among holistic evaluations. *TESOL Quarterly,* 14 (1), pp.61-69.

PERKINS, K. (1983) On the use of composition scoring techniques, objective measures and objective tests to evaluate ESL writing ability. *TESOL Quarterly,* 17 (4), pp.651-671.

PERRON, J. D. (1977) The impact of mode on written syntactic complexity, Parts 1-3. *Studies in Language Education:* Reports 24, 25, 27. Athens: Department of Language Education, University of Georgia.

PINCAS, A. (1962) Structural linguistics and systematic composition teaching to students of English as a second language. *Language Learning,* 12, 185-194.

PINCAS, A. (1982) *Teaching English Writing.* London: Macmillan.

PINCAS, A., HADFIELD, G. and HADFIELD, C. (1982) *Writing in English.* London: Macmillan.

POETKER, J. S. (1977) Practical suggestions for improving and using essay questions. *High School Journal,* 61, pp. 7-15.

POLIO, C. and GLEW, M. (1996) ESL writing assessment prompts: How students choose. *Journal of Second language Writing.* 5(1), pp.35-49.

POLLITT, A. et al. (1985) What makes exam questions difficult? An analysis of "O" grade questions and answers. *Research Reports for Teachers,* No.2 . Edinburgh: Scottish Academic Press.

PORTE, G. K. (1995) Writing wrongs: copying as a strategy for understanding EFL writers. *ELT Journal,* 49(2), pp. 144-151.

PORTER, D. (1978) Cloze procedure and equivalence, *Language Learning,* 28 (2) pp. 333-341.

PRINGLE, E. (1981) Towards a taxonomy of given-new information, *in* COLE, P. (ed) *Radical pragmatics.* London and New York: Academic Press, pp. 225-255.

QUELMALZ, E. and CAPELL, F. (1979) *Defining writing domains: effects of discourse and response mode.* California University, Los Angeles, Centre for the Study of Evaluation. (ERIC Document Reproduction Service No. ED 212 661)

QUELLMALZ, E., CAPPEL, F. and CHOU, C. (1982) Effects of discourse and response mode on the measurement of writing competence. *Journal of Educational Measurement,* 19, pp. 241-258.

RAIMES, A. (1983a) Anguish as a second language? Remedies for composition teachers, *in* FREEDMAN, A., PRINGLE, I. and YALDEN, J. (eds) *Learning to write: First language/ second language.* London: Longman, pp.258-272.

RAIMES, A. (1983b) Tradition and revolution in ESL teaching. *TESOL Quarterly,* 17, pp. 253-252.

RAIMES, A. (1983c) *Techniques in teaching writing.* Oxford: Oxford University Press.

RAIMES, A. (1987) Language proficiency, writing ability and composition strategies: A study of ESL college student writers. *Language Learning,* 37(3), pp. 459-468.

RAIMES, A. (1990) The TOEFL test of written English: Causes for concern. *TESOL Quarterly,* 24 (3), pp. 427-441.

REED, W.M., BURTON, J.K. and KELLY, P.P. (1985) The effects of writing ability and mode of discourse on cognitive capacity engagement. *Research in Teaching of English*, 19 (3), pp. 283-297.

REID, J. M. (1990) Responding to different topic types: A quantitative analysis from a contrastive rhetoric perspective, *in* KROLL, B. (ed) *Second language writing: Research insights for the classroom.* Cambridge: Cambridge University Press, pp. 191-210.

REID, J. (1993) *Teaching ESL writing.* New Jersey: Prentice Hall.

RICHARDS, J. C., PLATT, J and PLATT, H. (1992) *Longman dictionary of language teaching and applied linguistics.* 2nd ed., Essex: Longman Group Limited.

ROBB, T., ROSS, S. and SHORTREED, I. (1986) Salience and its effect on EFL writing quality. *TESOL Quarterly*, 20 (1), pp. 83-95.

ROBINSON, T. (1985) Evaluating foreign students' compositions: The effects of rater background and of handwriting, spelling and grammar. PhD Thesis, University of Texas at Austin.

ROBINSON, P. (1988) Academic Writing: Process and product. *ELT Documents 129.* Reading: Modern English Publications and the British Council.

ROSEN, H. (1969) An investigation of the effects of differentiated writing assignments on the performance in English composition of a selected group of 15/16 year old pupils. Ph.D. Thesis, University of London.

RUBIN, D. and PICHE, G, (1979) Development in syntactic and strategic aspects of audience adaptation skill in written persuasive communication. *Research in the Teaching of English*,18, pp. 293-316.

RUTH, L., and MURPHY, S. (1988). *Designing writing tasks for the assessment of writing.* Norwood, N.J.: Ablex.

RYE, J. (1982) *Cloze procedure and the teaching of reading.* London: Heinmann Educational.

SCHOLES, R. (1981) Language, narrative and anti-narrative, in MITCHELL, W. J. T. (ed) *On narrative.* London: University of Chicago Press, pp.200-208.

SCOTT, V. M. (1996) *Rethinking foreign language writing.* Boston, Mass: Heinle & Heinle Publishers.

SELIGER, H.W. and SHOHAMY, E. (1989) *Second language research methods.* Oxford: Oxford University Press.

SHERMAN, J. (1994) *Feedback: essential writing skills for intermediate students.* Oxford: Oxford University Press.

SHOHAMY,E., GORDON, C. and KRAEMER, R. (1992) The effect of rater's background and training on the reliability of Direct writing tests. *The Modern Language Journal,* 76, I (27-33).

SHUY, R. and FASOLD, R. (1973) *Language attitudes: Current trends and prospects.* Washington D. C. : Georgetown University Press.

SIEGEL, M. E. (1982) Responses to student writing from new composition faculty. *College Composition and Communication,* 33 (3), pp. 302-309.

SIEGEL, S. (1956) *Nonparametric statistics for the behavioral sciences.* New York, NY: McGraw-Hill.

SIEGEL, S. and CASTELLAN, N. J. (1988) Nonparametric statistics for the behavioral sciences. 2nd edition, New York: Mcgraw-Hill.

SILVA, T. (1990) Second language composition instruction: Developments, issues and directions in ESL, in KROLL, B. (ed) *Second language writing : Research insights for the classroom.* Cambridge: Cambridge University Press, pp. 11-36.

SILVA, T., LEKI, I. and CARSON, J. (1997) Broadening the perspective of mainstream composition studies: Some thoughts from the disciplinary margins. *Written Communication,* 14, pp. 398-428.

SMITH, W. and SWAN, B. (1977) Adjusting syntactic structure to varied levels of audience. *Journal of Experimental Education,*46, pp. 29-34.

SPACK, R. (1984) Invention strategies and the ESL college composition student. *TESOL Quarterly,* 18, pp. 649-670.

SPAAN, M. (1989). *Essay tests: What's in a prompt?* Paper presented at the 23rd Annual TESOL Convention, San Antonio, Texas, March.

SPAAN, M. (1993) The effect of prompt in essay examination, in DOUGLAS, D. and CHAPELLE, C. (eds) *A new decade of language testing research: Selected papers from the 1990 language testing colloquium.* Alexandria, VA: TESOL, pp. 98-121.

SPOLSKY, B. (1995) *Measured words: the development of objective language testing*. Oxford: Oxford University Press.

STEPHENS, M. (1992) *Practise advanced writing*. Harlow: Longman.

STUBUS, J.B. and TUCKER, G.R. (1974) The cloze test as a measure of English proficiency. *Modern Language Journal*, 58, pp.239-242.

SWALES, J. (1981) *Writing scientific English*. Walton-on-Thames: Thomas Nelson.

SWALES, J. (1984a) Research into the structure of introductions to journal articles and its application to the teaching of academic writing, *in* WILLIAMS et al. (eds).

SWALES, J. (1984b) Thoughts on, in and outside the classroom, in JAMES (ed).

SWALES, J. (1990) *Genre Analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

SWALES, J and FEAK, C.B. (1994) *Academic writing for graduate students: Essential tasks and skills*. Ann Arbor: The University of Michigan Press.

TAHRIRIAN, M.H. (1986) Attitudes and methodology in teaching English as a foreign language. *Indian Journal of Applied Linguistics*, 13 (1), pp. 93-100.

TAKASHIMA, H. (1987) To what extent are non-native speakers qualified to correct free composition? A case study. *British Journal of Language Teaching*, 25 (1), pp. 43-48.

TANNEN, D. (1989) *Talking voices: Repetition, dialogue, and imagery in conversational discourse*. Cambridge: Cambridge University Press.

TAYLOR, B.P. (1980) Teaching composition skills to low level ESL students, *in* CROFT, K. *Reading on English as a second language: For teachers and teacher trainers*. Cambridge, Mass.: Winthrop Publishers, pp. 367-385.

TAYLOR, W. L. (1953) Cloze procedure: a new tool for measuring readability, *Journalism Quarterly*, 30, pp. 415-433.

TEDICK, D. (1993) *A multidimensional exploration of raters' judgements of ESL writing*. Paper presented at TESOL Conference, Atlanta.

THORNDIKE, R. L. and HAGEN, E. (1969) *Measurement and evaluation in Psychology and education*. New York: Wiley.

TRZECIAK, J. and MACKAY, S. E. (1994) *Study skills for academic writing*. New York: Prentice Hall.

VANN, R., LORENZ, F. and MEYER, D. (1991). Error gravity: Faculty response to errors in the written discourse of nonnative users of English, *in* HAMP-LAYONS, L. (ed), *Assessing second language writing in academic contexts*. Norwood, N.J.: Ablex.

VAUGHN, C. (1991) Holistic assessment: What goes on in the rater's mind? *in* Hamp-Lyons (ed) *Assessing second language writing in academic contexts*. pp. 111-125. Norwood.NJ: Ablex.

VEAL, L. R. and TILLMAN, M. (1971) Mode of discourse variation in evaluation of children's writing. *Research in the Teaching of English*, 5(1), pp.37-45.

WEAVER, F. (1973). The composing processes of English teacher candidates: responding to freedom and constraint. Ph.D. Thesis, University of Illinois.

WEIGLE, S. C. (1994) Effects of training on raters of ESL compositions. *Language Testing*, 11, pp.197-223.

WEIR, C. J. (1983) *Identifying the language problems of overseas students in tertiary education in the UK*. Ph.D. Thesis, University of London, England.

WEIR, C. J.(1990) *Communicative Language testing*. London: Prentice Hall.

WEIR, C. J.(1993) *Understanding and developing language tests*. New York: Prentice Hall.

WHITE, E. (1984). Holisticism. *College Composition and Communication*, 35, 400-409.

WHITE, E. (1985) *Teaching and assessing writing*. San Francisco: Jossey-bass.

WHITE, H. (1980) The value of narrativity in the representation of reality. *Critical Inquiry*, 7, pp. 5-27.

WHITE, R. (1988) *Teaching writing skills*. London: Longman Groups Limited.

WHITE, R. and ARNDT, V. (1991) *Process writing*. London: Longman.

WHITE, R. and McGOVERN, D. (1994) *Writing; student's books*. New York: Prentice Hall International.

WHITE, E. and THOMAS, L.L. (1981) Racial minorities and writing skills assessment in the California state universities and colleges. *College English*. 48, pp. 276-288.

WIDDOWSON, H. G. (1978) *Teaching language as communication.* Oxford: Oxford University Press.

WILKINSON, A. (1983). Assessing language development: The Credition Project, *in* FREEDMAN, A., PRINGLE, L. and J. YALDEN (eds) *Learning to write: First language/second language* (pp. 67—86).New York: Longman.

WILKINSON, A. (1989) Our first great conversationalists. *English in Education,* 23 (2), pp. 12-24.

WILKINSON, A. (1990) Argument as a primary act of mind. *English in Education,* 24 (1), pp. 10-21.

WISEMAN, S. (1949) The marking of English compositions in grammar school selection. *British Journal of Educational Psychology,* 19, pp.200-209.

WISEMAN, S. and WRIGLY, J. (1958) Essay reliability: The effect of choice of essay title. *Educational and Psychological Measurement,* 18, pp. 129-138.

WITTE, S. et al (1990) *Holistic evaluation: Issues, Theory, and practice.* New York: Guilford Press.

WITTE, S. and CHERRY, R. (1986) Writing processes and writing products, in COOPER, C. and GREENBAUM, S. (eds) *Studying writing: Linguistic approaches vol. 1.* Newbury Park: Sage Publications, pp. 112-153.

WOOD, R. (1990) *Assessment and Testing: A survey of research.* Cambridge: Cambridge University Press.

YOUNGMAN, M. B. (1979) *Analysing social and educational research data.* London: McGrow Hill Book Company (UK) Limited.

ZAMEL, V. (1976) Teaching composition in the ESL classroom: What we can learn from research in the teaching of English. *TESOL Quarterly,* 10, pp. 67-76.

ZAMEL, V. (1983) The composing processes of advanced ESL students: six case studies. *TESOL Quarterly,* 17(2), pp. 165-187.

ZAMEL, V. (1987) Recent research on writing pedagogy. *TESOL Quarterly,* 21(4), pp. 697-715.