Test-Takers' Perceptions:
Behind the Construct of a Test

By
Elena Bogdanova Aminkova

A Master Thesis submitted to
The Faculty of Graduate Studies and Research

In partial fulfillment of
The requirements for the degree of

Masters of Arts

School of Linguistics and Applied Language Studies

Carleton University

Ottawa, Ontario

August, 2004

© Elena Bogdanova Aminkova

All rights reserved

Library and
Archives Canada

Bibliothèque et
Archives Canada

Published Heritage
Branch

Direction du
Patrimoine de l'édition

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 0-612-98947-X

Our file Notre référence

ISBN: 0-612-98947-X

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

PAGINATION ERROR.

ERREUR DE PAGINATION.

TEXT COMPLETE.

LE TEXTE EST COMPLET.

Abstract

The main focus of the study is on test-takers' accounts of their perceptions of a high-stakes test (the CAEL Assessment): more specifically the relationship between test-takers' proficiency and accounts of their perceptions of the test. The test-takers' accounts of their perceptions that were taken into special consideration were test difficulty, fairness, the entire testing experience, etc. The findings of the study suggest that test-takers' perceptions of a high-stakes test can be affected by a number of characteristics (e.g. test-takers' proficiency levels, test-takers' immediate pragmatic goals for taking the test, test-takers' emotional state during the test, test characteristics, etc) and can provide valuable insights about what tests actually do. Test-takers' perceptions are useful resource for test development and use.

Acknowledgments

I would like to thank to everybody who helped me along the way with this thesis, to the people who were there for me listening, and giving advice.

Special, huge thanks to my supervisor Dr. Janna Fox who helped me find my thesis topic, design and conduct the study, encouraged me during the research, incubation, and writing stages, and helped enormously with the revisions of the numerous drafts.

I would like also to thank Dr. Devon Woods for his help during the entire course of my M.A. degree and for helping me refine the questions for the semi-structures interviews.

Prof. Natasha Artemeva is another person who has a great contribution for the existence of the thesis. She kept on insisting that academic writing was not something that I was incapable of doing.

A special thanks to Wendy Fraser for her help in conducting the study (especially for coming in on weekends).

Another round of thanks goes to my friends Allison Moro and Krista Jack for being there for me in the difficult moments, and to the previous and current GSA Executive and staff for making me feel at home at work.

Last but not least in importance, an enormous thanks goes to my parents Dr. Bogdan Aminkov and Penka Aminkova who made my academic studies in Canada possible, for their constant material and moral support all the way during

my M.A. studies. Thanks to my brother Konstatin and grandparents Elena and Vasil Valkovi for being there for me even half a world apart.

And now once the rounds of thanks are over, I would like to share something that I learned while working on this thesis. Contrary to the commonly shared view that academia is “an ivory tower” and that academic work and the people working there are isolated from the “real world”, my experience while writing the thesis proved something different. Writing a Master’s Thesis like any other human activity is social action (Miller, 1984) in its essence, both the writer and the product are influenced by the social context they are in and they also influence this context through their contact with people in it.

Having said this, I would just like to add that the existence of this thesis is also a proof of the incredible resilience of the human nature regardless of what happens in the social context people find themselves in. My final conclusion is that we do not realize what we are able of overcoming until we are faced with it.

Table of Contents

1. Chapter One: Introduction.....	1
1.1. Theoretical Background	6
1.1.1. SLA and SLL Research: Affective Factors.....	6
1.1.2. Perceptions.....	10
1.2. Research questions.....	14
1.3. Initial Constraints of the thesis.....	15
2. Chapter Two: Literature Review.....	17
2.1. Test-takers' reactions to tests.....	17
2.2. Research gaps.....	21
2.3. Features of test-takers' feedback and reactions.....	25
2.3.1. Anxiety.....	27
2.3.2. Time limits.....	30
2.3.3. Attitudes and positive reactions.....	32
2.3.4. Ways of improving testing experience.....	34
2.3.5. Test items/ format.....	36
2.3.6. Test validity.....	41
2.3.7. Test difficulty.....	42
2.3.8. Difficulty factors.....	44
2.4. Test-takers' perceptions of task/test difficulty.....	50
2.4.1. Planning/preparation time.....	51

2.4.2. Test format/items.....	54
2.5. The Cognitive approach in language testing.....	59
2.6. The Social approach in language testing.....	59
3. Chapter Three: Methodology.....	75
3.1. Method.....	75
3.2. Participants.....	76
3.3. Instruments.....	79
3.4. Procedures.....	85
3.5. Analysis.....	88
3.5.1. Phase One: Analysis of Post-test questionnaire.....	88
3.5.2. Phase Two: Analysis of Semi-Structured Interviews and Journals.....	93
4. Chapter Four: Results.....	94
4.1. Phase One: Post-Test Questionnaire.....	94
4.1.1. Frequency analyses.....	94
4.1.2. Crosstabulation with Chi-square analyses.....	99
4.2. Phase Two: Semi-Structured Interviews and Journals.....	108
4.2.1. Participant One: Wendy.....	108
4.2.2. Participant Two: Dave.....	110
4.2.3. Participant Three Chloe.....	111
4.2.4. The Interviews before taking the test.....	113

4.2.4.1. Emerging themes.....	121
4.2.4.2. Recommendations.....	124
4.2.5. The Journals.....	124
4.2.6. The Interviews after taking the test.....	127
4.2.6.1. Additional comments.....	133
4.2.6.2. Emerging themes.....	137
4.2.7. Outcomes: distinctive participants' profiles.....	138
4.2.7.1. Wendy.....	138
4.2.7.2. Dave.....	141
4.2.7.3. Chloe.....	145
4.2.7.4. Common comments.....	148
4.2.8. Results Summary.....	149
5. Chapter Five: Discussion.....	151
5.1. Phase one and Phase Two- A Comparison.....	151
5.2. Discussion of the Research Questions.....	159
5.2.1. Findings.....	165
6. Chapter Six: Conclusion.....	174
6.1. Limitations of the thesis.....	177
6.2. Suggestions for the CAEL post-test questionnaire.....	180
6.3. Suggestions for the CAEL Assessment.....	181
6.4. Suggestions for testing in general.....	183
6.5. Directions for future research.....	184

7. References.....	185
8. Appendices	
7.1. Appendix A: Post-test questionnaire.....	196
7.2. Appendix B.....	198
7.2.1. Semi-structured interview questions before the test.....	199
7.2.2. Semi-structured interview questions after the test.....	200
7.3. Appendix C.....	201
7.3.1. Tables: Response Rates.....	202
7.4. Appendix D: Ethics Approval Letter.....	204

Tables

Table 1: Structure of the study.....	76
Table 2: Test-takers' five most commonly reposted first languages.....	77
Table 3: Test-takers' five most commonly reported intended area of study.....	77
Table 4: Overall test-takers' CAEL bandscores.....	99
Table 5: Crosstabulation of test-takers' perceptions of time in relation to test fairness and topic.....	100
Table 6: Crosstabulation of test-takers' perceptions of having enough time and that of the overall testing experience.....	101
Table 7: Chi-square analyses of test-takers' perceptions of room comfort and the test.....	102
Table 8: Chi-square analyses of test-takers' perceptions of sound quality of tape and their perceptions of the test.....	102
Table 9: Chi-square analyses of test versions and test-takers' perceptions of the test.....	103
Table 10: Crosstabulation of test version and test-takers' overall bandscores.....	104
Table 11: Crosstabulation of test-takers' overall bandscores and perceptions of test fairness.....	105
Table 12: Crosstabulation of test-takers' overall bandscores and perceptions of topic importance.....	106
Table 13: Crosstabulation of overall bandscores and perceptions of topic importance.....	106

Table 14: Wendy-CAEL bandscores.....	109
Table 15: Dave-CAEL bandscores.....	111
Table 16: Chloe-CAEL bandscores.....	112
Table 17: Response rates (RR) of test-takers' perceptions analyzed for frequency of response.....	202
Table 18: Response rates (RR) of test-takers' bandscores and perceptions of the test.....	202
Table 19: Response rates (RR) of test-takers' perceptions of test time and the test.....	202
Table 20: Response rates (RR) of test-takers' perceptions of test topic/version and the test.....	203
Table 21: Response rates (RR) of test-takers' perceptions of room comfort and the test.....	203
Table 22: Response rates (RR) of test-takers' perceptions of sound quality of the tape and the test.....	203

Abbreviations

CAEL Assessment – Canadian Academic English Language Assessment

DAM – discursive action model

ESL – English as a Second Language

L2 – second language

OLT – Oral Language Test

OPI – Oral Proficiency Interview

RGS – Rhetorical Genre Studies

RR – response rate

SATI – Scholastic Aptitude Test I

SLA – Second Language Acquisition

SLL – Second Language Learning

SOPI – Simulated Oral Proficiency Interview

TOEFL – Test of English as a Foreign Language

TWE – Test of Written English

Transcriptions notations

Muffled = ()

Pause = //

Long pause = ///

P: = Participant

R: = Researcher

CHAPTER ONE: INTRODUCTION

My interest in language testing is to a great extent due to my rather extensive personal experience as a language test-taker, and some more recent experience in test development and marking. My experience as a test-taker was mainly related to writing norm-referenced tests such as the Test of English as a Foreign Language (TOEFL) and the Scholastic Aptitude Test I (SAT I) which are used for application purposes in American universities and colleges. Later, I had the opportunity of becoming acquainted with a criterion-referenced test – the Canadian Academic English Language (CAEL) Assessment (Fox, 1995; 2001, 2003), which is also a high-stakes test used to screen students for admission. This time my experience was enriched by the opportunity to view a high-stakes test from test raters' and developers' points of view.

In addition, my practical experience was backed up by a Language Testing course at the 400 level that I took during the first year of my M.A. degree in Linguistics and Applied Language Studies at Carleton University with Dr. Janna Fox. The main purposes of this course were to become familiar with the basics of language testing, to develop practical skills in language test development, to write a critical evaluation of an already existing language test in terms of reliability and validity, and to develop a sample language test. Equipped with theoretical knowledge, I was able to explore further the subject of language testing by taking a testing tutorial course with Dr. Fox during the Fall term of the second year of

my M.A. degree. For this course I was required to write a paper on an area of language testing research of particular interest to me.

The resulting paper I chose to write attempted to bridge different disciplinary interests: 1.) research on writing in Rhetorical Genre Studies (Freedman and Medway, 1994; Giltrow, 2002; Miller, 1994), 2.) research in second language studies on affective factors in language learning (Arnold, 1999; Casado, 2001; Dornyei, 2001; Gardner, 2001; Geen and Shea, 1997; Krashen, 1981; McLeod, 1997), 3.) and research in language testing on task difficulty and test-takers' perceptions and feedback on language tests (Bachman, 1990; Brindley and Slater, 2002; A. Brown, 1993; J.D. Brown, Hudson, Norris and Bonk, 2002; Elder, Iwashita, and McNamara, 2002; Iwashita, McNamara and Elder, 2002; Fulcher and Reiter, 2003; Robinson, 2001; Scott, 1986; Shohamy, 1982; Skehan, 1996; Stansfield, Kenyon, Paiva, Doyle, Ulsh and Cowles, 1990; Zeidner and Bensoussan, 1988). I intended in the tutorial paper to examine the key factors that could have influence test-takers' performance on tests and to explore the implications of test-takers' perceptions and feedback for language testing. Although the focus of the paper was primarily on language testing, the purpose of bringing into the conversation research from other areas of applied language studies was to approach the issues that test-takers face not in isolation, but by considering the social context within which test-takers engage in testing.

In this thesis I extended the scope of the above-mentioned paper. The primary objective of this thesis was to examine test-takers' perceptions of a high-

stakes test, the Canadian Academic English Language (CAEL) Assessment. The main focus of the study is on test-takers' accounts of their perceptions of a high-stakes test (the CAEL Assessment): more specifically the relationship between test-takers' proficiency and accounts of their perceptions of the test. The test-takers' accounts of their perceptions that were taken into special consideration were test difficulty, fairness, the entire testing experience, etc.

First, I analyzed a sample drawn at random from CAEL test-takers' responses to post-test questionnaires, which are routinely filled out by test-takers after writing a CAEL test. Then, I interviewed three students planning to take the CAEL test about their perceptions before and after taking the test. Prior to taking the test these test-takers were asked to record their test-related experiences in a journal focusing on their expectations, and feelings about the test as well as their motivation, anxiety, etc. The data obtained through the research instruments (post-test questionnaires-Appendix A, Interview questions - Appendix B, and journals) was compared. The test-takers' journals were used to triangulate (Nachmias and Nachmias, 1987) the data obtained from the CAEL post-test questionnaires and the interviews. As a result of the study, it was argued that test-takers' perceptions are an important source of information about what tests actually do.

In the following section, the theoretical background to the study is reviewed. It includes an overview of literature in Second Language Acquisition

(SLA) and Second Language Learning (SLL)¹ regarding affective factors in language learning, and definitions of perception from different points of view.

It is my strong belief that language testing research in test-takers' feedback and perceptions can benefit from research in SLA and SLL. As Elder et al (2002) state learners' perceptions, attitudes and affective reactions towards language learning and the influence of these on learners' learning and therefore performance are better researched in SLA and SLL than in language testing. SLA and SLL have long traditions of using learners' perceptions to examine the development of language learning. Shohamy (2000), in an article on the relationship between SLA and language testing, finds that language testers are not sufficiently taking test-takers' attitudes towards tests into consideration, nor are they then sufficiently considering the needs of institutions and groups which use tests. In addition, research has identified interfaces between the two fields (see e.g. Bachman and Cohen, 1998; Chapelle, 1998; Shohamy, 2000; Swain, 2001). As Bachman and Cohen (1998) note despite the different research focuses of SLA and SLL, and language testing a number of researchers believe that there are common research questions concerning both fields. Chapelle (1998) adds that language tests can serve as SLA and SLL elicitation devices.

¹ Krashen (1981) distinguishes between language acquisition and language learning. He defines language acquisition as resulting from a subconscious process similar to the way children acquire their first language. On the other hand, language learning results from formal instruction. It is a conscious process of learning about the language. According to Krashen (e.g. 1981) learning is less important than acquisition, acquisition and learning belong to different domains and the transition from one to the other is not possible. Other researchers contrary to Krashen's view (e.g. 1981) consider that there is a possibility of transition between learning and acquisition, and vice versa, and use the two terms interchangeably (Bialystok, 1994; Ellis, N., 1994; Ellis, N., 1998; Ellis, N., 1999; Ellis, N., 2002; Ellis, R., 1994a).

The need for more cooperation between the two fields is clearly demonstrated in the works of Shohamy (2000) and Swain (2001). In her study Shohamy (2000) concludes that language testing might be relevant in theory but not in practice, as it does not address properly issues such as language complexity and because it violates basic SLA principles. Swain (2001) points out that test-takers are also interested in language learning, and that, in fact, examining test-takers' performance might help both SLA and language testing researchers to understand the strategies behind learners'/ test-takers' performance.

1.1. Theoretical Background

1.1.1. *Second Language Research (SLA) and Second Language*

Learning (SLL) Research: Affective Factors

It is important to note that like research in language testing, research in SLA and SLL has traditionally used cognitive approach² (see e.g. McNamara, 1995, 1997, 2001). It distinguishes between cognitive and affective variables/factors³, which impact both the process of language learning⁴ (Wenden, 1998) and inevitably learners' perceptions of language learning as well as all processes related to language learning such as language testing from this perspective. Cognition is viewed as a mental process, or a faculty, by which knowledge is acquired, "it is a sober examination of the facts in pursuit of veracity" (McLeod, 1997, p 5). Affect, which is considered to be the noncognitive aspect of human behavior, is used as a generic term to address a wide variety of phenomena such as emotions, feelings, attitudes, beliefs, moods, motivation, intuition, drives, instincts and perceptions (McLeod, 1997).

According to Dulay, Burt and Krashen (1982) one's affect - and therefore perceptions of a given thing, event or situation - reflects the way a thing, event or situation corresponds to one's needs or goals, and the impact it has on one's

² This is not surprising given the fact that the move from the predominantly cognitive perspective to the social one has been taking place over the past 10-15 years.

³ Variables/factors – refer to one and the same concepts as these terms are used interchangeably in the literature reviewed.

⁴ The cognitive and affective factors/ variables are distinguished on the basis of the distinction between cognition and affect made within the framework of the cognitive approach. In fact, the social and more specifically an ecological approach to language testing (Fox, 2002, 2003), on which the thesis draws, does not make a distinction between cognitive and affective factors.

emotions. However, it is important to point out that affect and perceptions are not synonymous for feelings and/or emotions (McLeod, 1997).

The affective variables/factors in SLA and SLL are classified as either individual or relational (Gardner, 2001):

a.) Individual Affective Factors

The individual affective factors focus on learners' characteristics such as anxiety, motivation, attitudes, perceptions, self-confidence, self-esteem, cognitive and language learning styles and strategies, language aptitude, reactions to pedagogical techniques, the learning environment, and learning materials.

Individual affective factors also depend on personality variables such as: the need for achievement, risk-taking, empathy (Gardner, 2001), and attributes such as gender, social class, professional experience, and proficiency (Elder et al., 2002).

Research in language testing generally acknowledges that learner factors such as anxiety, confidence and motivation can influence test takers' performance and interact with the tests' / tasks' characteristics (see e.g. Elder et al., 2002).

b.) Relational Affective Factors

It is important to point out that the relational affective factors are formed by learners' social identity (i.e. belonging to a certain socio-cultural group) (Dornyei, 2001).

The significance of the affective factors for learners' language acquisition is well defined by Krashen's (1994) affective filter hypothesis.

Affective Filter

The role and influence of affective factors have been a focus of various studies in SLA and SLL (Arnold, 1999; Casado, 2001; Dornyei, 2001; Gardner, 2001; Geen and Shea, 1997; Krashen, 1981; McLeod, 1997). According to Krashen's Affective Filter Hypothesis (1994) under certain circumstances affective factors/variables can cause a mental block that prevents the input from reaching the language acquisition device⁵. Actually, the main effects of affective factors on learners' acquisition and writing can be summarized either as beneficial:

- when the affective factors facilitate the completion of the task. In this case the affective filter is low and the input can reach the language acquisition device (Krashen, 1994).

or as negative:

- when the affective factors inhibit the completion of the task (Krashen, 1981; Wenden, 1998). In this case the affective filter is high and the input cannot reach the language acquisition device (Krashen, 1994).

Krashen (1994) claims that learners with high motivation, self-confidence, positive perceptions of the learning process and all activities associated with it, a good self-image, and a low level of anxiety are more likely to be successful in second language learning.

⁵ Chomsky (1965) defines the language acquisition device as a genetic mechanism for language acquisition present in infants, which enables them to develop linguistic competence.

On the other hand, low motivation, low self-esteem, negative perceptions of the learning process and all activities associated with it, and high levels of anxiety can combine to “raise” the affective filter and form a “mental block” that prevents the input from being used in the process of language learning (Krashen, 1985).

Rosenberg and Hovland (1960, cited in McLeod, 1997) state that anxiety is a component of the concept called attitude. However, affective reaction is only a part of an attitude. As a result of an affective reaction, concepts or objects can be evaluated or labeled as good or bad, positive or negative (McLeod, 1997). Eiser (1987, cited in McLeod, 1997) argues that the relationship between one’s attitude and future behavior/experience is logical and based both on personal and social factors. In fact, both attitude and perceptions might be also defined not as a response, but rather as readiness to behave in a certain way (McLeod, 1997).

In that light, the abovementioned affective factors in language learning can be applied to a large extent to test-takers’ performance on language tests. First, within the current social context of language learning, testing is an activity inherent to the process of language learning. Second, as already mentioned in this section, learners’ perceptions can have influence on their language learning and consequently could have implications for their test performance (e.g. see e.g. Shohamy, 2000) and their perceptions of their test performance (e.g. see A. Brown, 1993).

In the following section of this chapter, I examine the concept of perception from different perspectives.

1.1.2. Perceptions

Within language testing research, perceptions have traditionally been considered in cognitive terms. Recently, however, the term has been reconsidered from a social point of view as well.

a.) The Cognitive View:

From a cognitive point of view perception is the process of interpreting and organizing “incoming sensory information” (Volpe, 1995, p. 83) from the environment i.e. the surrounding social context. Volpe (1995) also adds that perception is a creative-constructive process, which links sensory experience to meaning. Perception is a strictly mental process, which cannot be directly observed, however it can be verified in one’s behavior and verbal reports.

Moreover, perception operates to establish a relationship between present experiences and similar ones in the past. It is not based only on present input but also on concepts previously associated with this input (Volpe, 1995).

Another important aspect about perception from the cognitive perspective is the correlation between perception and categorization. The organizational factors of perception are the factors of thought (Volpe, 1995) thus any given type of experience can become a focal point of attention. As a result, conditions

facilitating the link between a given perception and already existing experience is established, then new experience is normally recognized and classified as belonging to a certain category. The categorization of perceptions happening in our brains allows researchers to detect recurrent patterns and trends in one's behavior, verbal reports and accounts.

b.) The Social View:

Perceptions in Discursive Psychology

Discursive psychology (see particularly Edwards and Potter, 1992) defines perceptions from a social point of view, which is central to the ecological approach in language testing⁶ adopted in this thesis. Within the framework of discursive psychology, discourse is viewed as social practice, i.e. a real-world phenomenon that can be studied and not just as a theoretical abstraction.

Moreover, discourse is viewed as central to both interaction and cognition.

Participants in conversational events are performing communicative, interactional work where memory is not only recalling an isolated event. On the contrary, memory has to do to with communicative actions and interests. Everyday communicative/interactional i.e. discursive practices such as talk and writing contain an abundant number of formulations such as perceptions, knowledge, and inferences describing different events. Therefore, within the "ecological" tradition (Bronfenbrenner, 1997, cited in Fox, 2003; Fox, 2003) perceptions can be viewed

⁶ The ecological approach to language testing (Fox, 2002, 2003) is discussed in detail further in the thesis.

as social practices and can be studied as functions in relation to the context of the social actions in which remembering occurs.

Perceptions from a Social Perspective

Bakhtin (1981), whose ideas are central to the social view to writing (see for example Freedman and Medway, 1994; Miller, 1994), talks about “krugozory” which means the circle of one’s own vision” (p.425). He states that “krugozory” is always something very specific and differs from individual to individual or, in other word, one’s own vision pertains only to this individual only. When this concept is applied on a more global scale and to the scope of society, it is called a “belief system”.

c.) The View in SLA and SLL

In SLA and SLL, learners’ perceptions might be due to learners’ beliefs and attitudes, to the features of a particular task and/or test. Attitudes are considered to arise from beliefs. Actually beliefs are also defined as “psychological entities formed as a result of past experience and are linked to one’s future behavior, reactions and experience”. (McLeod, 1997, p. 76) Woods (2003) defines beliefs from a social constructionist perspective on language learning (McGoarty, 1998, cited in Woods, 2003) as interconnected and structured, formed by means of social interaction, and situated in social contexts, and integrated in learners’ thought and action.

At this point it is important to add that research on test-takers' feedback and affective reactions to tests indicates that test-takers' benefit from and are pleased with the opportunity of expressing their feelings and attitudes towards tests they have taken (Zeidner and Bensoussan, 1988; Zeidner, 1988).

Some studies are more specific, (Madsen and Murray, 1984, cited in Scott, 1986; Scott, 1986; Zeidner, 1988; Zeidner, 1990) stating that talking about test-taking experience and releasing aversive emotional reactions lessens test-takers' anxiety levels and may even help their test performance since test-takers normally regard test-taking experiences as personally meaningful events. Zeidner (1990) claims that the act of assessing a test *in situ* can serve as a catharsis for test-takers and they can function better during tests. Bradshaw (1990) concludes that retroactive reporting on tests generally leads to identifying issues causing post-test resentment and underperformance.

The literature on test-takers' feedback/ perceptions and difficulty factors in language testing discussed in Chapter Two of this thesis includes mainly cognitively based studies despite the fact that the overall approach in this thesis is socially informed. The reason for this is that most of the available research in language testing, for example research on task/test difficulty, effects of planning time and test-takers' feedback draws on the cognitive approach (Bachman, 2002; Brindley and Slatyer, 2002; Bradshaw, 1990; A. Brown, 1993; J. D. Brown et al., 2002; Elder et al, 2002; Foster and Skehan, 1996; Fulcher and Reiter, 2003; Iwashita et al., 2001; Norris, Brown, Hudson and Bonk, 2002; Robinson, 2001;

Skehan and Foster, 1997; Wigglesworth, 1997, 1998). There are only a limited number of studies to date within a socially informed perspective (e.g. Chalhoub-Deville, 2003; Fox, 2003 and McNamara, 1995, 1997; Norton and Stein, 1998).

1.2. Research Questions:

The reason for undertaking the study was to compare test-takers' perceptions of their performance on a high-stakes test (the CAEL Assessment) before and after taking the test. More specifically the following questions were addressed:

- 1.) How do test-takers' perceive a high stakes test?
 - 2.) What characteristics (e.g. test-takers' language proficiency, test-takers' perceptions of their language proficiency, test difficulty, fairness, the testing situation itself) seem to be related to test-takers' perceptions?
 - 3.) What is the relationship, if any, between test-takers' proficiency and their perceptions of the test?
-
- 4.) Is it possible to use test-taking strategies in order to compensate for insufficient language proficiency and knowledge?

This chapter includes an overview of SLA and SLL research on affective factors in language learning, and definitions of perception from different perspectives. Chapter Two presents a theoretical overview of language testing research on test-takers' feedback and reactions to tests. It reviews some of the

affective and cognitive factors considered to affect test-takers' performance and their overall impression of the testing process. Chapter Two discusses literature regarding test-takers' perceptions and feedback in language testing, research on difficulty factors, and those factors thought to be related to test-takers' perceptions of test difficulty. It also includes a detailed discussion of the origins of the ecological approach to language testing (Fox, 2002, 2003) and examines studies drawing on it. Chapter Three discusses the methodology of the study. This chapter also examines an overview of the undertaken analysis in the thesis. Chapter Four includes the results from Phases One and Two of the study conducted for the purposes of this thesis. Chapter Five presents the discussion and Chapter Six - the conclusion.

1.3. Initial Constraints of the Thesis:

Although the thesis reviews literature on test-takers' feedback and reactions to tests, it does not include overview of research on the relationship between test-takers' personal characteristics (such as gender, nationality, language learning background) and their test performance and/or perceptions of language tests. Although undoubtedly test-takers' personal characteristics can be related to their test performance and perceptions of tests, the available literature on this issue (Bradshaw, 1990; A. Brown, 1993; A. Brown and Iwashita, 1996; Clapham, 1998; Hill, 1998; Ginther and Stevens, 1998; Purpura, 1998; Zeidner and Bensoussan, 1988) seems rather controversial and does not offer conclusive

findings and noticeable trends as to the effect of test-takers' personal characteristics on their performance or perceptions of tests. Although the analysis of the interviews with the three participants raises issues such as washback, test-takers' language learning and test-preparations strategies, discussion of such literature is beyond the scope of this thesis. The issue of test topic and previous knowledge also emerges. However, Phase One suggests that test topic did not impact test-takers' perceptions of the test. In addition, research on the CAEL test (e.g. Jennings, Fox, Graves, and Shohamy, 1999) demonstrates that the test topic did not have significant effect on test-takers' scores.

The review of research in SLA and SLL in the Introductory Chapter is not comprehensive. It was included with the purpose to provide the reader with a definition of affective factors/variables, anxiety, motivation, etc in the area of language studies.

Similarly, the discussion of Rhetorical Genre Studies (RGS) and discursive psychology done in relation to the ecological approach of language testing (Fox, 2002, 2003) is not thorough. The literature in question was added in order to provide the reader with some background on the origin of the ecological approach to language testing (Fox, 2002, 2003).

CHAPTER TWO: LITERATURE REVIEW

2.1. Test-Takers' Reactions to Tests

Chapter Two reviews language testing research on test-takers' feedback and reactions to tests. It examines some of the affective and cognitive factors⁷ identified in language testing and more specifically the ones that are considered to that could test-takers' perceptions of test difficulty. When discussing test-takers' perceptions of test difficulty, it is important to note that most of the studies discussed in this chapter draw on a cognitive approach. In brief the cognitive approach to language testing regards the context in which testing takes place as a separate variable from language users/test-takers and/or their language abilities. The social approach to language testing and more specifically the ecological approach (Fox, 2002, 2003) regards the social context as inseparable from the participants and actions involved in the testing process. It is the ecological approach to language testing that was adopted for the study in this thesis. The nature and the differences between the cognitive and social approaches are discussed in greater detail in this chapter.

Early research in language testing (Shohamy, 1982) recognizes that testing has impact, which reaches further than fulfilling its primary purpose of measuring test-takers' language knowledge, abilities and/or proficiency. Test-takers'

⁷ The cognitive and affective factors/ variables are distinguished on the basis of the distinction between cognition and affect made within the framework of the cognitive approach. In fact, the social and more specifically an ecological approach to language testing (Fox, 2002, 2003), on which the thesis draws, does not make a distinction between cognitive and affective factors.

feedback, perceptions, reactions, motivation, perseverance, attitude to the testing experience, self-esteem, self-confidence and anxiety are factors that may have significant implications for test-takers' performance on a test as well as for the test assessment accuracy. Research in language testing has not proven conclusively what is the relationship, if any, between the above-mentioned factors and test-takers' test performance. However, Shohamy (1982) argues that because learning a second or foreign language involves social and psychological factors such as these their effect might even be intensified in testing setting.

Nevo (1985) states that measurement specialists in testing are generally inclined to consider test-takers' feedback, attitudes, perceptions, motivation, and emotional dispositions for the purposes of test administration and research. Zeidner (1990) when reviewing research in classroom testing adds that test-takers' feedback and reactions should be given their due importance for assessing test validity of language courses' curriculum and/or determining classroom testing policies and procedures. A number of researchers in testing highlight the use of test-takers' feedback (Alderson, 1988; A. Brown, 1993; Cohen, 1984; 1988, 1997, cited in McNamara, 1997; Hill, 1988; Nevo, 1985; Shohamy, 1982; Stansfield, 1991; Zeidner, 1987, 1988, 1990, Zeidner and Bensoussan, 1988). Kenyon and Stansfield (1991, cited in McNamara, 1997) comment on the usefulness of test-takers' feedback and note that it's easier to obtain as compared to raters' feedback for instance. According to Hill (1998) and A. Brown (1993) test-takers' feedback can have both practical and theoretical applications. The former has to do with an

effective measurement of the face validity of tests⁸. At this point it needs to be said that test-takers' feedback has not traditionally been viewed as a legitimate source of evidence for validity of tests' content (Bachman, 1990). However, Bachman (1990) emphasizes that the importance of test-takers' feedback to test face validity is significant and should not be undermined. Hill (1998) highlights the significance of face validity especially for high-stakes tests, which could determine key outcomes in test-takers' lives. Zeidner (1988, 1990) repeats numerous times that taking into account test-takers' feedback is "evidence superior" (Zeidner, 1988, p. 216) for test face validity.

A possible theoretical application of test-takers' feedback suggested by Hill (1998) is using feedback to support statistical analysis performed for identifying problematic areas of tests and testing situations. Zeidner and Bensoussan (1988) note that there is a certain amount of research dealing with test-takers' affective reactions to tests, reactions to test formats/items as well as various other aspects of classroom testing. Indeed, test-takers' feedback can be obtained regarding individual test items or item types, entire tests and entire test batteries. As far as classroom tests in particular are concerned, Zeidner (1998) points out that examinee inventories could be used in evaluating classroom testing techniques and procedures such as daily quizzes, homework assignments, term papers, presentations, etc. Zeidner and Bensoussan (1988) offer a fairly

⁸ Face validity is defined as a measure of "the knowledge or abilities it claims to measure, as judged by an untrained observer" (Davies, Brown, Elder, Hill, Lumley and McNamara, 1999, p. 59) such as the test-taker or the institutions which plans to administer it. Research literature does not generally consider face validity a good argument for validity.

comprehensive classification of how test-takers' feedback could be applied to some aspects of the testing process. There are four main areas:

- The first concerns the entire testing process and test-taking experience as reflected in test-takers' perceptions, attitudes, beliefs and affective reactions before, during and after a test.
- The second has to do with the different aspects of a test, its face validity and fairness and has implications for test development, evaluation and administration. For instance, information provided by test-takers can be used to modify or correct testing conditions perceived by students as causing problems or being disadvantaging and negative.
- The third area is in identifying flaws and problems in tests and testing procedures. Zeidner (1988) points out that the use of test-takers' feedback could go even beyond that and have implications for test-takers' preparation behaviour, which on its own may influence performance and attainment on tests.
- The fourth area is related to the "subjective ecology" (Zeidner and Bensoussan, 1988, p. 100) of the testing process itself and the importance of giving test-takers opportunities to express their opinions, feelings, attitudes and emotional dispositions towards a test.

Other researchers identify additional aspects of the testing process that would benefit from test-takers' feedback. Alderson (1988) notes that test-takers' feedback could be used for the development of test specifications and the

modification of test items. Stansfield, Kenyon, Paiva, Doyle, Ulsh and Cowles (1990) suggest using test-takers' feedback for revising test instructions. All of these suggestions could be added to the second aspect of Zeidner and Bensoussan's (1988) classification, which includes issues such as test face validity and fairness, test development, evaluation, and administration. Hill (1998) adds to this list, the identification of threats to test validity and fairness such as detecting sources of cultural bias, insights into exposing cheating and/or other types of subversive behaviour.

2.2. Research Gaps

A. Brown (1993) points out that despite the availability of research in the area of test-takers' feedback and reactions, there is significant criticism of the methodology of obtaining, analyzing and applying test-takers' feedback in the various aspects of the testing process. Elder et al. (2002) highlight that in SLA and SLL research learners' feedback/ perceptions are widely addressed compared to the area of language testing. There are a number of research gaps in the practical application of test-takers' feedback as well.

For example, the majority of the most prominent research (Brutch, 1979, cited in Scott, 1986; A. Brown, 1993; Elder et al., 2002; Hill, 1998; Kenyon and Stansfield, 1991, cited in McNamara, 1997; Scott, 1986; Shohamy, 1982; Stansfield et al., 1990; Zeidner and Bensoussan, 1988) in this area is based only on or deals predominantly with oral tests or compares test-takers' reactions to one type of test item versus another such as essay versus multiple-choice type exams

(Zeidner, 1987, 1998, 1990). Learners'/test-takers' perceptions are perhaps best used for examining difficulty of listening tests and tasks (Elder et al., 2002).

It is unfortunate that the application of research on test-taker's feedback seems to be neglected, as it could be a valuable source of information for tests' subjective qualities. Zeidner and Bensoussan (1988) in a fairly early article examining test-takers' attitudes to an oral test of English state that face validity of cognitive tests can be measured using test-taker's feedback. They identify research gaps in the implementation of test-takers' feedback in test development, validation and assessment of test-takers' attitudes to oral versus written tests. They particularly emphasize the fact that there is no systematic approach to the development and validation of test-takers' feedback so that testing specialists could obtain insight into the "subjective ecology" (Zeidner and Bensoussan, 1988, p. 100) of testing situation. Bradshaw (1990) and A. Brown (1993) in much later articles also observe that there is a lack of uniform policy in implementing test-takers' feedback and reactions to tests.

Other studies point at various other gaps in the research and implementation of test-takers' feedback and reactions. Zeidner (1990) when assessing three studies of undergraduate and graduate college students' attitudes towards key facets of classroom tests such as test item formats, and varied testing models and conditions, concludes that there is much more data available about school children's' attitudes, preferences and dispositions towards varying test formats than about those of older students' and test-takers'. This article also recommends that further research be undertaken to determine when feedback

should be obtained, how often and for what types of tests. Zeidner (1990) also suggests that the implementation of test-takers' feedback could bring a change in some of the current methods and procedures in testing, if necessary. This, on its part, might have positive academic and attitudinal consequences for students and test-takers.

Early research in test-takers' feedback (Zeidner, 1988 and 1990) reports that test-takers' feedback is a crucial but neglected source of information in regards to test qualities and characteristics from the point of view of tests' major clientele - the test-takers. Bradshaw (1990) points out that due to not so abundant research in the area of test-takers' feedback, it is difficult to draw any general conclusions about the value and possible practical applications of test-takers' feedback. She also attributes this fact to the contradictory, in her opinion, results obtained from various studies on test-takers' feedback to the absence of consistency in methodology of gathering and interpreting research data. In A. Brown's opinion (1993) this situation is due to a lack of research in the practical application of test-takers' feedback in test development. She states that there is no uniformly established method for analyzing and revising oral performance test tasks based on test-takers' feedback.

Bradshaw (1990) when discussing test validity and in particular test-takers' perceptions of test validity, notes that validity is a scarcely researched test characteristic. She argues that the available literature on test-taker's perceptions and reactions to test validity account for test-takers' emotional state during tests and influences test-takers' opinions after that. Zeidner (1988, 1990) highlights the

importance of test validity not only in terms of content, construct and predictive test validity but also in terms of its perceived test validity by test-takers. Other researchers also take a very strong stand on the issue of test validity. Hill (1998), for example, claims that obtaining feedback for the purposes of test validation should be a normal practice in test development. Elder et al. (2002) points out that unlike SLA studies where learner/ test-takers' perceptions occupy a central place, in language testing learner/ test-takers' perceptions receive just marginal treatment. Test-takers' reactions are not considered to be a central source of data for test validation purposes, which is left to test development and analysis experts (Elder et al., 2002). In the general case, test-takers' feedback and reactions especially regarding test difficulty are used only in relation to test face validity or "appearance of validity" and therefore are not considered as important and/or crucial to the test validation process (Bachman, 1990). However, test-takers' attitudes to a test can affect on their performance and thus can have implications for test validity or at least perceived test validity. Therefore, a number of researchers (Nevo, 1985; Spolksy, 1995; Elder and Lynch, 1996, cited in Elder et al., 2002) admit that if test-takers' attitudes towards tests affect their performance this dependence may lead to drawing "unwarranted inferences" (Elder et al., 2002, p. 350). Interestingly, Messick (1989) also recommends using test-takers' perceptions as evidence for construct validity.

2.3. Features of Test-Takers' Feedback and Reactions

Zeidner and Bensoussan (1988) when discussing test face validity identify three main aspects of the testing process with regard to which test-takers' feedback/reactions can be retrieved.

The first aspect is the testing situation itself including test takers' perceptions of emotional atmosphere, presence of examiner, examiner-test-taker rapport, physical facilities and environment, test administration procedures, clarity of instructions, and allotted time. The second aspect is test characteristics such as perceived fairness of tests, readability and comprehension of test items, perceived interest of test content and procedures. Scott (1986) adds to this classification test format, test length, perceptions of test validity and difficulty. Bradshaw (1990) states that test characteristics and/or testing situation can affect test-takers' performance and even cause post-test resentment. The third aspect is test-takers' individual characteristics, for instance anxiety, motivation, attributions of success/failure, achievement expectancies, test attitudes and preferences. Indeed these test-takers' individual characteristics coincide with what other researchers define as affective reactions to tests. It seems reasonable to assume that both testing situation and test characteristics would have influence on test-takers' affective reactions to tests.

Elder et al. (2002) refers to learner factors such as anxiety, confidence and motivation when reviewing difficulty factors in language testing. The mentioned learner factors are considered to produce different levels of stress and to lead to different attention span while performing test tasks. This way learner factors come

into interaction with the test characteristics. The outcome of this interaction can have implications for learner's/test-takers' test performance.

Prior to examining in detail the nature of test-takers' reactions to tests, however, it is important to mention that the testing literature offers another classification distinguishing between test-takers' cognitive perceptions as opposed to their affective reactions to tests. Test-takers' cognitive perceptions mainly deal with judgments (Jones et al., 1980, cited in Zeidner and Bensoussan, 1988 and Scott, 1986) as to the perceived reliability and validity of tests, perceived relation of test content to the received instruction, and correctness in reflecting test-takers' language proficiency, and test ease/ difficulty. In addition to the already mentioned test-takers' affective reactions (i.e. anxiety, motivation, self-esteem, self-confidence and motivation) Scott (1986) identifies emotive reactions including perceived pleasantness of testing experience, frustration, test fairness, and perceived success. However in the testing literature some of these test-takers' reactions are classified both as cognitive and affective, for example the perceptions of test validity and reliability and the perceptions of test fairness (Bradshaw, 1990; Scott, 1986).

Based on the literature reviewed for this thesis, the field of language testing seems to view test-takers' feedback, and perceptions as a combination of cognitive and affective reactions to tests. Cognitive and affective reactions on their part can be due either to test-takers' individual characteristics or to the testing process itself be it testing situation or test characteristics or the interactions between any of them. This chapter will examine in detail some of test-takers'

affective and cognitive reactions most frequently mentioned in testing research literature.

2.3.1. Anxiety

Anxiety is one of most frequently researched test-takers' affective reactions. Scott and Madsen (1983) mention that ESL/EFL literature "is virtually mute" (p. 271) regarding test affect, but that there is very extensive body of research on test anxiety in psychology. Since then the situation in ESL/EFL literature has changed, however, the importance of psychological research on anxiety should be given due importance in the area of language testing. Research in psychology (Sarason, Davidson, and Lighthall, 1960, cited in Shohamy, 1982) distinguishes the direct relationship between anxiety and test performance. Anxiety can influence test-taker's performance both in a positive and in a negative way or in other words it can have a facilitating or a debilitating effect (Alpert and Haber, 1960, cited in Scott, 1986). High-test anxiety normally leads to lower performance as test-takers under stress experience cognitive interference and preoccupation (Sarason et al., 1960, cited in Shohamy, 1982). Whether anxiety in a testing situation will have a facilitating or a debilitating effect depends on the testing situation itself, test characteristics and/or test-takers' individual characteristics. However, the interaction between these factors is only partially known (Bradshaw, 1990).

Research on test anxiety also distinguishes between *trait* and *state* anxiety. *Trait* anxiety is viewed to be a stable personal characteristic towards a particular

situation or task. Conversely, *state* anxiety is considered to be more transient and dependent on perceived threat, it is caused by features of a given situation or task (Bradshaw, 1990; Spielberg, 1966, cited in Scott, 1986). Generally, people with higher *trait* anxiety are more likely to experience *state* anxiety more frequently since they are more sensitive to potential danger or threat. However, there is no dependency between the two types of anxiety, the presence of one does not predispose the presence of the other. According to Madsen (1982) *state* anxiety seems to have more debilitating/negative effect on people with higher *trait* anxiety. Other researchers report that test-takers experiencing high levels of anxiety tend to underperform on tests (Alpert and Haber, 1960, cited in Scott, 1986; Nevo and Stetz, 1985).

Zeidner (1988), in relation to classroom tests, suggests several explanations why students associate testing experience with anxiety and aversive emotional reactions. These reactions, first of all, are maybe due to the high degree of importance that students attribute to achievement on classroom tests. Second, they can be a result of the perceived difficulty of tests, and third, maybe a consequence of the “relatively speeded testing conditions” (p. 221).

What seems to be a common finding in language testing studies on test-takers’ feedback and reactions to language tests, which coincides with the previously mentioned findings within research in psychology, is the link established between test-takers’ anxiety and performance (Hill, 1998; Madsen, 1982; Madsen and Murray, 1984, cited in Scott, 1986; Scott, 1986; Zeidner and Bensoussan, 1988). Zeidner (1988) in a study examining students’ reactions to

classroom testing, using a “group associative examinee feedback technique”⁹, finds out that students’ tend to link classroom tests to aversive emotional reactions and states, and more importantly test anxiety is the major and most salient meaning component.

The other reactions and states belonging to this category are fear, tension, worry, irritability, headache and nervousness. Besides test anxiety, other major meaning components¹⁰ associated with classroom tests are: achievement (success, failure, grades), exam preparation (studying, cramming, preparing), difficulty (easy, difficult), items (questions, answers), knowledge (knowledge, facts, concepts), thought, speed (fast, quick, too little time, time) teacher, cheating, and attributions (interesting, boring, long, short). Zeidner (1988) also identifies ten most popular individual associations in regards to classroom tests. In general, they are again related to aversive emotional reactions and states and include fear, grade, failure, emotional arousal, study, tension, stress, success, knowledge and questions.

⁹ The “group associative examinee feedback technique” is based on “associative group response test” (Szalay and Deese, 1978, cited in Zeidner, 1988), which measures meaning components related to different referent stimuli.

¹⁰ These meaning components are associated with “group associative examinee feedback technique” (Zeidner, 1988). They are responses provided by more than one test-taker and are consequently grouped into common themes (Zeidner, 1988).

Scott and Madsen (1983) who strongly argue in favor of measuring test affect in evaluating language exams state that extremely frustrating experiences on tests and quizzes can result in creating negative classroom atmosphere, lead to lower achievements and problems with discipline and dropout.

In addition to anxiety, testing literature as already mentioned recognizes factors such as nervousness, and frustration. A. Brown (1993) concludes that anxiety and frustration can have serious adverse effects on test-takers' performance and even decrease the fairness of a test in test-takers' opinion. Cohen (1984, cited in Bradshaw, 1990) considers anxiety/nervousness to be a stronger cause for test-takers' adverse reactions than low English language proficiency. Scott (1986) who interchangeably talks about test-takers' nervousness and anxiety concludes that these states are generally caused by reactions to the testing situation, test content, time constraints, perceptions of validity, test length, test format, testing environment, familiarity with test instructions and format.

2.3.2. Time Limits

As already mentioned, the effect of anxiety on test-takers' performance appears to depend on a combination of factors such as testing situation itself, test characteristics and/or test-takers' individual characteristics (Bradshaw, 1990). Zeidner and Bensoussan (1988) in review of research on anxiety in testing settings identify time constraints as one of the primary factors for test anxiety, which on its part could have implications for test-takers' performance. Scott (1986) in her study of Brazilian students' affective reactions to oral tests

establishes direct relationship between students' emotional state and in particular anxiety and the time constraints of the test. The same findings are supported by Hill (1998), Madsen (1982) and Madsen and Murray (1984 cited, in Scott, 1986).

Both in the studies of Madsen (1982) and Madsen and Murray (1984, cited in Scott, 1986) time limits and most importantly insufficient time are concluded to be the main factors for test-takers' anxiety and frustration. Hill (1983, cited in Scott, 1986) reports the effects of time limits on performance and the link between time limits and anxiety. She found that high-test anxious children working under time constraints made three times more errors than when the time constraints were removed. Moreover, these children needed twice as much time as those who were identified as low test anxious. Once again when the time constraints were removed the high-test anxious children took approximately the same time as the low-test anxious children and their performance was as good as that of the low-test anxious children. A. Brown (1993) observes negative feedback from test-takers again in relation to time limits of a test. In this research study the allotted time for the test is deemed as inadequate and considered to prevent test-takers from demonstrating their real language abilities. Fox (2003) researching bias detection in two versions of the CAEL Assessment identifies time as the factor most frequently mentioned by the test-takers.

However, other researchers do not come up with the similar findings. Galassi, Frierson Jr and Siegel (1984, cited in Scott, 1986) conclude that time limits account only to a small extent for test-takers' anxiety and therefore performance on a test. Schwarzer et al. (1982, cited in Scott, 1986) state that

depending on the time allotted for a test, test anxiety might have indirect effect on test-takers' performance. Bradshaw (1990) reports inconclusive results when introducing longer time limit for completing a test. In her study after the introduction of a longer time limit test-takers' reactions did not change in any significant or explainable way.

2.3.3. Attitudes and Positive Reactions

The motivation and attitude of test-takers can influence their reactions to tests and their performance (A. Brown, 1993). Shohamy (1982) differentiates between test-takers' attitudes and the rest of the affective reactions to tests. She points out that test-takers' attitudes together with anxiety are of significant importance to test-takers' performance especially when certain types of test items such as cloze tests are involved. Shohamy (1982) also points out that test-takers' attitudes towards the entire testing experience might have serious consequences not only for their attitude to tests in general but also for their attitude to the learning of the language tested.

Both early and later testing research (Foster and Skehan, 1996; Hill, 1998; Savingnon, 1972, cited in Shohamy, 1982; Scott, 1986; Shohamy, 1982) identify reasons for test-takers' positive reactions and attitudes to tests, considering testing experience as a pleasant one and liking tests. The main factor, which contributes to that, is when a testing experience, in test-takers' opinion, corresponds to real life situations or is regarded as a learning experience regardless of the test difficulty. Foster and Skehan (1996) report that discourse salience can affect test-

takers' level of accuracy especially in terms of syntax and morphology. In this case, it is speculated that the higher-level degree of precision and accuracy in test-takers' opinion is due to the fact that test-takers use language in tasks having more personal significance to them. Brutch (1979, cited in Scott, 1986), Shohamy (1982) and Scott (1986) add that test-takers' think that completing communicative tests allow them to demonstrate their real language knowledge. Communicative tests are regarded to correspond to real life situations, which makes the testing experience to be perceived as more relaxed, more pleasant, less anxious and even less intimidating. Hill (1998) observes that test-takers with relevant work experience (i.e. professionals) reacted more positively to a test¹¹ than the test-takers who lacked such professional background (i.e. non-professionals) regardless of the fact that the former group did not perform better than the latter group.

¹¹ The test in question is the **access** test, which is a test of English language proficiency written by prospective migrants to Australia (Hill, 1998).

2.3.4. Ways for Improving Testing Experience

A number of researchers in language testing make recommendations as to how to reduce the test-takers' anxiety and improve the testing experience in general. Shohamy (1982) in her investigation of affective factors in language testing makes a number of recommendations in regards to testing procedures that seem to be preferred by test-takers. She summarizes that both test format and test content should be perceived by test-takers as a real-life experience, as a learning tool, should create low anxiety, and offer a relaxed atmosphere. In addition to evoking positive reactions in test-takers, test providing real-life experience/performance can motivate test-takers' and students and enhance their preparation.

Thus a test can achieve much beyond its primarily purpose which is measuring test-takers' language knowledge, abilities and/or proficiency, it can serve as a learning tool as well (Shohamy, 1982).

In this relation, Zeidner (1988) also proposes solutions for improving classroom-testing experience for students. Students should be encouraged to view classroom testing more as a feedback technique designed to monitor their progress in acquiring the material they are supposed to learn rather than as a measure of their academic potential, abilities and status. Zeidner (1988) supports this idea by referring to research in anxiety. Covington (1986) claims that test-takers perceive testing situations as threatening to their sense of self-worth and in the same time determining for their reputation of competency. Moreover, there is a link between the perceived risk of failure and students' idea of their own

intellectual capabilities. As far as test characteristics are concerned, test-takers' anxiety can be reduced and their performance can be enhanced by offering test-takers' the opportunity to choose between alternative test items and/or test formats (Zeidner, 1990). Greater familiarity with test items/formats also can lead to decrease in test anxiety (Scott and Madsen, 1983). Another possible solution suggested by Scott and Madsen (1983), more applicable to classroom than to high-stakes testing, is to provide extensive exposure to new test items/formats before administering the actual test. In fact, in Scott and Madsen's (1983) study the test frustration and anxiety is reduced with each test administration both with the high and low proficiency groups. It is important to highlight that after readministering the test the test-takers' attitudes to the test was more positive regardless of their scores. In addition, the test items, which were rated most favorably, were not the easiest ones. These findings are of interest in relation to the outcome of the study undertaken in this thesis.

Averill (1973, cited in Zeidner, 1990) states that through retesting test-takers achieve higher perceived sense of control over testing situations. This reduces the experienced psychological stress and as a result the testing experience is regarded as less conflicting. Seligman and Miller (1979, cited in Zeidner, 1990) add that when given a choice, test-takers' experience higher sense of self-control and freedom. The presented solutions by Scott and Madsen (1983) and Zeidner (1988, 1990) although made in regards to classroom testing settings can be to some extent taken into consideration and applied to high-stakes testing situations.

As already mentioned earlier in this chapter, test-takers' reactions to tests depend on a combination of factors such as the testing situation itself, test characteristics and/or test-takers' individual characteristics (Zeidner and Bensoussan, 1988). The following section of this chapter will examine the relationship between test items/ formats and test-takers' reactions. It will also look into the factors accounting for test-takers' perceptions of other test characteristics such as test validity and difficulty.

2.3.5 Test Items/ Format

Scott (1986) discusses test-takers' affective reactions to different test items/ formats. The reviewed research shows that test-takers' perceptions of test items/ formats could influence both their affective and cognitive reactions towards a test, especially test-takers' perceptions of anxiety, test validity and test difficulty (A. Brown, 1993). Zeidner (1988) recommends that test evaluators working with teachers in the development of classroom tests should be more receptive to test-takers' preferences and emotional dispositions to test items and should take these preferences into account when choosing test formats. It is important to point out that test-takers' are able to distinguish between their perceptions of test item difficulty from their attitude (positive or negative) to test items/formats (Sauvignon, 1972, cited in Scott, 1986).

Different studies (Sauvignon, 1972, cited in Scott, 1986; Shohamy, 1982; Zeidner and Bensoussan, 1988) show that test-takers' prefer oral communicative tests, as these tests are thought to test and retrieve test-takers' actual language

knowledge. Oral interviews/tests as compared to discrete points tests (Brutch, 1979, cited in Scott, 1986) and cloze tests (Shohamy, 1982) are also viewed by test-takers as having less anxiety load, to be more relaxing, more pleasant, and closer to real-life experience (Shohamy, 1980, cited in Scott, 1986). According to some studies written tests are considered to be more pleasant, valuable, fair, with lesser anxiety load, and more reflective of test-takers' language abilities. Oral test on their part are viewed by test-takers as more interesting (Zeidner and Bensoussan, 1988; Zeidner, 1990). As it can be seen, both some oral (Shohamy, 1982) and written tests (Zeidner and Bensoussan, 1988; Zeidner, 1990) can be received as having a lesser anxiety load.

Test-takers' perceptions of oral test can vary on the basis of the particular type of oral test and its administration, for instance there are semi-direct/tape-based oral tests and live ones. A comparison of live Oral Proficiency Interview (OPI) and Simulated Oral Proficiency Interview (SOPI) discussed in A. Brown (1993) demonstrates that test-takers' preferred the live test as they felt less nervous about the testing procedure and therefore perceived it as less difficult.

These findings are confirmed by other studies also stating that test-takers tend to regard tape-based tests as more stressful (Hill, 1998; McNamara, 1990, cited in Elder et al., 2002). However, McNamara (1990 cited in Elder et al., 2002) suggests that these reactions can be due to particular test task rather than to be a reaction to tape-based oral tests only. He speculates that the time limit might be a factor as well since some of the test-takers regarded speaking for a certain period of time on a given topic as "artificial".

Another possible explanation suggested by Hill (1998) is that language background might be a variable in test-takers' reactions to test items/formats. Basically, it is presumed that test-takers feel more nervous when they are not familiar with certain language learning techniques and methods of testing. In this study subjects from Asian countries felt more nervous on the live interview format as they were not familiar with communicative language teaching and therefore they were more comfortable with the format of the language laboratory than with the live interview. However, other studies such as Zeidner (1988) do not establish such dependence.

The variation in test-takers' perceptions of written tests may also depend on the particular type of written tests as it is the case with oral tests. As already said, test-takers' generally do not favor discrete-point and cloze tests and seem to prefer communicative written tests involving writing essays and letters (Bradshaw, 1990; Brutch 1979, cited in Scott, 1986). Undoubtedly, as A. Brown (1993) points out different findings can be due to differences in test content, test format, test-takers' individual characteristics, differences in test purposes and/or different methods used for obtaining test-takers' feedback.

For instance, Zeidner (1990) evaluating undergraduate students' attitudes towards essay versus multiple choice item formats, report different results from the above-mentioned ones. Students view essays as more reflective of their true achievements, more fair, allowing individual expression of ideas, and with less potential for cheating. However, essays are also perceived as relatively long and requiring over learning and having more knowledge on order to succeed.

Multiple-choice items on their part are considered to be easier to prepare for, requiring less preparation time, and less time consuming during exams. However, another study again conducted by Zeidner (1988) reveals slightly different results, graduate students tend to perceive essay exams and multiple-choice exams in a similar way. In addition, they regard written exams as easier, and evoking greater success expectancies. Another study by Zeidner (1987) on junior-high school students' perspective on essay versus multiple-choice classroom exams provides more insight into the students' reasoning behind their attitudes. In general, and surprisingly enough as compared to other studies, students reacted more positively to multiple-choice type exams than to essay type exams. Multiple-choice type exams are considered to be less difficult, less complicated, clearer, more interesting, less tricky, less anxiety evoking eliciting higher success expectancies, and making students feel more at ease. In addition, they were judged as:

- easy to prepare for, as they require less time and effort for preparation and less memorizations of factual material
- more convenient to use as it is possible to guess the correct answer with a chance of success and because there is no need to express answers in written form. The students who preferred essay type exams in this category said that they saw in essay type exams the opportunity to express their ideas freely and accurately.
- having greater chance for success as they are considered easier to prepare for, having of answer choices gives students more

security and confidence, students can guess the correct answer, there is no danger of lowering the scores because of spelling mistakes, and these exams require less time and effort. The students who preferred essay type exams in this category say that they can demonstrate the extent their knowledge to a maximum degree.

- having lesser anxiety load as essay type exams require extended additional effort and emotional energy for selecting, organizing and expressing ideas in essay form. Essay type exams do not give clues to the right answer, they are relatively long in length and require relatively complex responses.

Overall, the multiple-choice item types are viewed more favorably because they offer answer choices to choose from, have convenient answer form, there is no need to organize and write answers, and offer the opportunity to guess the correct answer.

The essay answer choices received positive reactions into two categories - perceived value and fairness. They are viewed as:

- better reflecting students' knowledge of the exam material as they give students optimal options for expressing their ideas and knowledge.
- more fair because they do not offer the probability to guess the right answer so the scores obtained reflect students' real knowledge. The students who regarded multiple-choice item types as fairer said that this item type does not allow getting scores for partial answers.

Perhaps, students'/test-takers' varying reactions to same test items/formats are due to the interactions between students'/test-takers' individual characteristics and:

- their testing experience
- their perceptions of certain test items/formats
- and/or characteristics of the particular tasks their reactions/feedback was obtained in regard with

2.3.6. Test Validity

As already mentioned test-takers' reactions towards test items/formats can have implications for their perceptions of test validity and difficulty. For instance, Scott (1986) states that test validity is an important factor determining if test-takers would like a test or not. Zeidner (1988) regarding classroom tests concludes that if test-takers perceive a test as valid and fair, this would lead to improved examinee-examiner cooperation and rapport, optimal motivation and decreased negative emotional reactions and "debilitating motivational dispositions" (p. 231). Bradshaw (1990) summarizes the importance of validity saying that test validity is of concern to test-takers and can account for their emotional state during and after tests. Bradshaw (1990), in her study of test-takers' reactions to a placement test found, out that test-takers' rated as more valid the test items/ formats, which they viewed as most difficult. Provided that test-takers perceive tests as testing the language skills they are supposed to acquire, test-takers can react positively to tests and even view them as fair and valid even

when they consider tests to be very difficult, when they feel that they did not perform well and/or when they feel they were not prepared for the testing experience (Savignon, 1972, and Shohamy, 1980 cited in Scott, 1986).

Moreover, workplace orientation of tests is crucial for test face validity as these tests are viewed by test-takers as more appropriate (Hill, 1998).

Providing test-takers with planning time appears to be of concern both to test-takers and test developers. Basically, test-takers are faced with the dichotomy to promote “best performance” of test-takers on one hand, and to provide as authentic as possible testing situation, on the other hand (Wigglesworth, 1997, 1998). The source of concern here is that in everyday speech events, planning time is not usually available except for pauses, monitoring and repetitions, which are normally present in everyday speech events and, in a way, provide some time for online planning. Research on planning time in testing situations, which will be addressed in greater detail later in this chapter, proves that in the general case planning time does not have any crucial effect on test-takers’ discourse and therefore does not put them in a disadvantaged position.

2.3.7. Test Difficulty

In regards to test difficulty, A. Brown (1993) points out that time limit, vocabulary difficulty, speed, and lack of clarity of test instructions, lack of familiarity with test format/items/type are most salient characteristics that test-takers assign to test difficulty. She also adds that test difficulty may also depend

on test-takers' individual characteristics such as amount of time they have spent in the target language country, and the scores they obtained on a particular test.

However, before examining test-takers' perceptions of test difficulty it is essential to the background of this study to review literature on difficulty factors in language tests. As mentioned in Chapter One, the research on difficulty of listening tests and tasks (e.g. Bachman, 2002; Brindley and Slatyer, 2002; Elder et al., 2002; Fulcher and Reiter, 2003; Iwashita et al., 2001; and Norris et al., 2002), is more abundant. Despite this fact, it has appeared only recently (Fulcher and Reiter, 2003). One of the first sources that deals with the issue of task difficulty is Pollitt (1991, cited in Fulcher and Reiter, 2003) who discusses how tasks' difficulty in performance tests could vary. However, this source does not suggest any approach to assessing difficulty level of tests/tasks. Bachman's (1990) model of test method of facets and Weir's (1993, cited in Fulcher and Reiter, 2003) performance conditions are among the earliest methods for classifying difficulty level of test tasks. Language testing has also borrowed classificatory criteria developed and employed in SLA research (Fulcher and Reiter, 2003) that is one-way versus two-way tasks and communicative goal and activity type. However, according to Fulcher and Reiter (2003), Tarone (e.g. 1983, cited in Fulcher and Reiter, 2003) and Ellis (1985, cited in Fulcher and Reiter, 2003) are the ones who have put in the spotlight the issue of task difficulty in speaking tests.

The number of factors and variables that can be related to test performance goes beyond learners'/test-takers' perceptions only (Elder et al., 2002). Test difficulty and in particular task difficulty can have significant impact on test-

takers' performance. Therefore, the following section of this chapter will outline difficulty factors in language tests. Language testing research on task difficulty, similarly to other areas of research in language testing, is mainly cognitively based.

2.3.8. Difficulty Factors

Bachman (1990) proposes several types of factors influencing test performance: language ability of the test-takers, test characteristics, test takers personal characteristics, and other random factors. Skehan (1996 and 1998, cited in Elder et al., 2002) proposes a three-feature task difficulty model for determining task difficulty. Prior to looking into this model, it is important to explain the importance of task-based tests and the study of task characteristics. First of all, task-based test methods elicit performance in conditions as similar as possible to real world situations and allow for evaluation according to real world criteria. The evaluation of task difficulty is important as performance on one task can be used to estimate performance on other similar tasks (J.D. Brown et al., 2002). Skehan's model of task difficulty includes:

- **code complexity** – the syntactic and lexical difficulty of language input.
- **cognitive complexity** – the processing demands of tasks, and the availability of relevant schematic knowledge.

- **communicative stress** – the result of differentials in the time pressure, the modality of task performance, and the scale, or number of participants involved.

J.D. Brown et al. (2002) drawing on Skehan's (1996) model, try to operationalize performance difficulty factors that would allow for generalizations about test task performance and thus for generalisations about likely abilities in performing other tasks. Their definitions of the three processing components developed by Skehan are the following (J.D. Brown et al., 2002):

code command – depends on the kind of language and information required for successful task performance where the most important factor is the quality of the language that must be understood and/or produced by test-takers. This component is subdivided into two characteristics: range and input/output sources. The range is the extent to which the linguistics code (vocabulary, items, grammatical structures, pragmatic features of tasks, structure of the information that is presented and produced) is inherent to the language of a given task (i.e. if it has a greater or lesser degree of "spread"). The input/output sources are the degree to which test-takers must "decode" numerous sources of information each of which presents differences in the code.

cognitive operations – include the amount and kind of information processing that test-takers must engage in for successful task performance. This component is subdivided into two characteristics: input/output organization and input/output ability. The input/output organization is the extent to which information must be significantly organized, reorganized or shifted regarding

certain parameters. The input/output ability is the degree to which test-takers need to search the information upon which task performance is to be based.

communicative adaptation – is determined by the type of communicative language activity that is required and also the number of moderator variables that can drastically affect the relative difficulty of a given communicative act. This component is subdivided into two characteristics: mode addresses and response level. The mode addresses is the extent to which a task is deduced to have a productive element which is inherently connected to task performance success. The response level is the extent to which test-takers must interact with input in an “on-line” sense, i.e. to process information relatively immediately.

However, in another article by J.D. Brown et al. (2002) using the same operationalization of performance difficulty factors, the final results do not prove that the cognitive processing factors are very useful for estimating actual performance difficulty differences among test tasks:

We did not find that the cognitive processing factors as operationalized in our original task difficulty framework proved particularly useful for estimating actual performance difficulty differences among the tasks. As such, initial evidence does not support the use of this framework as an inferential basis for generalizing from examinees' performances on specific tasks to likely abilities with related tasks. Nevertheless, the observation of some degree of relationship between the cognitive factors and task-dependent performance evaluation (i.e. those based on 'real-world' criteria identified by external informants) suggests a fruitful area for research on language performance assessment (p. 415).

Again in regard to task-based test difficulty, Robinson (2001) differentiates between task complexity and task difficulty as well as task conditions, which tend to interact during task performance and to influence task

performance. Task complexity is defined as “the result of the attentional, memory, reasoning, and other information processing demands imposed by the structure of the task on the language learner” (Robinson, 2001, p. 29). In addition, the differences in information processing demands, due to design characteristics, are normally fixed and invariant. Task complexity involves cognitive factors, which are divided into resource-directing factors: number of task elements (+/- few elements), immediacy of information provided (+/-here-and-now), reasoning demands of the task (+/-no reasoning demands), and “resource-depleting” factors: planning time (+/-planning), number of tasks (+/-single task), prior knowledge (+/-prior knowledge).

Task difficulty involves learner factors, which on their part are divided into affective variables such as confidence, motivation, and anxiety and ability variables such as intelligence, aptitude, cognitive style, and proficiency. The affective variables are more temporary and likely to change and to affect the size of resource pool availability. The ability variables, however, are more permanent, fixed determinants of resource pools and more stable over a course of instruction.

Task difficulty is determined by the factors contributing to differences between them and the extent of available cognitive resources and thus it can be of help in explaining “*between learner variance*” (Robinson, 2001, p. 32) in test-takers’ performance on tasks. In addition, the relationship between task complexity and task difficulty is not a fixed one since first, test-takers may differ in intelligence and aptitude and second, may have different attentional, memory, and reasoning resource pool limits. The resource pool limits can be influenced by factors such as

motivation, for instance, greater motivation in one test-taker can temporarily expand the currently available resource pool limits.

Task conditions also called the interactive demands of tasks (Robinson, 2001) include both participation factors and participant factors. Participation factors for example are the direction of information flow (one way/two way), and communicative goals (one or many solutions) of task performance. Participant factors include gender, familiarity with each other and task role, relative status, task goal, and task interpretation.

To summarize the main characteristics of the above-mentioned classifications of test/tasks difficulty factors influencing test performance, there are two main approaches to task difficulty (Bachman, 2002). The first approach recognizes a number of task characteristics independent from test-takers' abilities (see Bachman, Davidson, and Milanovic, 1996). Brindley and Slatyer (2002) offer an example of those task characteristics, they include:

- amount of lexical overlap between the text and the response format
- length of text preceding the information required to respond
- length of required response
- repetition of tested information
- whether responses and repetitions of information are verbatim or paraphrases

The second approach to task difficulty identifies "difficulty features" (Bachman, 2002, p. 463) which are considered to be combinations of task characteristics and test-takers' abilities (such as Elder et al., 2002; Norris et al.,

1998 cited in Elder et al., 2002; and Skehan, 1996). An illustration of these “difficulty features” are once again presented in Brindley and Slatyer (2002):

- the nature of the input
- the nature of the impact
- the individual listener factors

Bachman’s (2002) argument in relation to those two approaches to task difficulty is that they are problematic since they are not proposed on the basis of test tasks, although claiming to do so, but on the basis of interactions between certain test takers and certain test tasks (i.e. those classifications cannot be applied to the testing process in general). Therefore the solution suggested by Bachman (2002) is to investigate how test takers interact with test tasks and what is the effect of this interaction on test performance.

Psycholinguistic Categories of Difficulty versus Pragmatic Categories of Difficulty

Fulcher and Reiter (2003) propose a refreshing approach to identifying task difficulty. They distinguish between psycholinguistic and pragmatic categories of task difficulty, stating that those categories are abstract in predicting task difficulty and fail to account for test performance variations. Fulcher and Reiter (2003), on their part, examine a number of pragmatic categories, which they consider to explain why particular tasks present more difficulties for certain test-takers than others. Their assumption is that the pragmatic characteristics of tests/tasks influence test performance because these characteristics are linked to the social and cultural demands presented by that particular test/task. Test-takers’

performance on a test/task may be reduced if the pragmatic demand of that task/test is realized in a different way and/or is culturally inappropriate in test-takers' L1 context.

In fact, Fulcher and Reiter's (2003) approach, without arguing for its validity as compared to the rest of the difficulty factors classifications discussed so far, comes very close to the issue of learners'/test takers' perceptions of test difficulty which is emphasized in this thesis.

Since perceived difficulty of a test appears to be an important variable (Bradshaw, 1990), depending on many factors, the following part of this chapter will review in a greater detail test-takers' perceptions of difficulty as reflected in language testing research.

2.4. Test-Takers' Perceptions of Task/Test Difficulty

The available research in language testing suggests that test difficulty and especially perceived test difficulty may influence test-takers' performance. For instance, Elder et al. (2002) recognizes that researching test-takers' perceptions of test difficulty may reveal how powerful the influence of test difficulty is on test-takers' performance. In fact, it is speculated that test-takers might be able to predict what makes a test item difficult, which may be helpful for the purposes of test development and bring more insight into the existing theories of test difficulty. Moreover, test-takers could point at other unnoticed difficulty features, which could also affect their performance. However, research in language testing dealing with test-takers' perceptions and feedback, does offer certain information

as to how test-takers perceive task difficulty and its impact on their performance. It is important to repeat that factors such as anxiety, confidence and motivation, termed by Elder et al. (2002) as learner factors, can produce different levels of stress, which on their part can affect test-takers' attention span when completing tests.

As already mentioned earlier in this chapter test-takers' perceptions, feedback and reactions can be obtained in relation to three main aspects, first, testing situation, second, test/task characteristics and third, test-takers' individual characteristics. These aspects can influence test-takers' perceptions, and reactions to tests separately, or they can interact with each other and thus influence test-takers' perceptions, and reactions.

2.4.1. Planning/Preparation Time

As far as test/task characteristics are concerned, the effect of planning/preparation time on test takers' perceptions of task difficulty (Hill, 1998) seems to be one of the most determining factors. On the basis of this, Elder et al. (2002) speculates that providing more time will result in a decrease of the perceived test difficulty.

In this respect, there is a group of studies examining the effect of additional planning time on test-takers' performance. However, it is salient to point out that these studies approach the issue of test/task difficulty from a cognitive perspective, which is not within the scope of the research of this thesis. Researchers investigating the effect of planning time in classroom situations

observe that the complexity of students' discourse when provided planning time was affected but not its accuracy (see e.g. Ellis, 1987; Crookes, 1989; Hulstijn and Hulstijn, 1984; Ortega, 1995, cited in Wigglesworth, 1998). However, Wigglesworth (1998) examining the effect of planning time in testing situations finds that test-takers' discourse is improved in terms of accuracy but only for difficult tasks. He concludes that planning time might have a major importance in reducing the load on test-takers' cognitive resources but only for difficult tasks. It is also speculated that the reverse findings in testing situations, i.e. improved accuracy compared to improved complexity in classroom situations, is due to interaction between test-takers' proficiency levels and the cognitive loads of the tasks. When simple tasks are performed planning time does not benefit test-takers because their cognitive capacity is not reached. Therefore, planning time is a factor, which may have influence on test-takers' perceptions of difficulty, but perhaps only for difficult tasks. However, there is also the issue of individual differences such as how (well) test-takers use planning time as well as their proficiency levels.

Another study conducted by Wigglesworth (1997), which in fact is a continuation of the one discussed above demonstrates similar findings as the above-mentioned. However, in this case test-takers' proficiency levels are taken into account. It appears that high-proficiency test-takers benefit from the extra planning time when performing difficult tasks. Low-proficiency test-takers, on their part, do not seem to make use of it. It is possible that high-proficiency test-

takers are focusing more on the form and complexity of their discourse whereas low-proficiency test-takers are focusing on the content (Wigglesworth, 1997).

It is also possible that test-takers with different proficiency levels use the planning time in different ways, i.e. they undertake different activities and focus on different aspects of tasks they need to complete. These different approaches in using the allotted planning time may account for a relationship between test-takers' language competence and their strategic competence. In addition, this finding may reflect how test-takers with different language proficiency levels use their strategic competence (Wigglesworth, 1997). Iwashita et al. (2001) does not find any consistent difference in test-takers' performance related to fluency and accuracy when planning time is introduced.

Elder et al. (2002) report only weak correlations between test-takers' perceptions of task difficulty and actual task difficulty as reflected in test-takers' performance. The qualitative part of this study gives some explanation as to why test-takers might consider planning time a facilitating factor. Test-takers with different levels of proficiency find telling a story easier when they are provided with 3 minutes of planning time. They could use the time to organize their ideas and write them down. However, this perceived ease on the part of the test-takers is not reflected in their actual test performance. A possible explanation offered by Elder et al. (2002) is that individual differences in the quality of planning done maybe accountable.

When it comes to planning time, there is also the issue of undetailed and detailed planning. It appears that when undetailed planning is used student

discourse shows progress both in accuracy and fluency. However, detailed planning seems to pose too big of a challenge to students, and as a result they are unable to control all aspects of their performance and output language. Basically, accuracy is lowered as students engage into the production of overambitious discourse (Foster and Skehan, 1996).

2.4.2. Test Format/Items

Another important test/task characteristic that is frequently investigated and researched in language testing literature is test format/items. The effect of different test format/items has already been discussed in this chapter. I will just repeat that certain test items/formats are perceived as being easier than others. Semi-direct/tape-based oral tests are generally perceived by test-takers as more difficult and/or stressful than live oral tests (A. Brown, 1993; Hill, 1998; McNamara, 1990, cited in Elder et al., 2002; Stansfield, 1991). Written communicative tests are more preferred than discrete-point and cloze tests (Bradshaw, 1990; Brutch, 1979, cited in Scott, 1986). More specifically, essay-type exams are considered to be fairer than multiple-choice exams, creating less potential for cheating, and allowing test-takers to express their knowledge and ideas better. Multiple-choice exams are viewed as requiring less preparation time and being easier to prepare for (Zeidner, 1988).

A. Brown (1993) offers a lengthier list of factors influencing test takers' perceptions of task difficulty. She refers to: inadequate response time, unfamiliar vocabulary, speed of voices on the tape, lack of clarity in instructions, unclear

prompts, too much input material to process and lack of familiarity with the task type (p. 351).

Another important part of test takers' perceptions of task/test difficulty is related to test takers' own characteristics interacting with the test/task characteristics. In that respect, language proficiency (Iwashita and Elder, 1997 cited in Elder et al., 2002; Scott and Madsen, 1983) is obviously the main factor influencing test takers' perceptions and their performance. According to Iwashita and Elder (1997, cited in Elder et al., 2002) test-takers' language proficiency is the most powerful test-taker personal characteristic. Scott and Madsen (1983) observe that candidates with lower proficiency tend to respond less positively to oral tests than candidates with higher proficiency. Brooks (1999, cited in Elder et al., 2002) observes similar dependency between test takers' attitudes toward the format of assessment and their performance. Other researchers such as Bradshaw (1990), A. Brown (1993), Shohamy (1982) and Zeidner (1988, 1990) detect a relationship between test-takers' performance on given tasks and their perceptions of the features of these tasks. Test-takers' with lower proficiency tend to be less favorable to task features higher proficiency test-takers'.

The reported tendency undoubtedly can be due to various factors such as familiarity with the format of a test/task, experience of taking the test/ doing the task, etc. However, language proficiency seems to be the leading factor for test takers' perceptions of test/task difficulty and their performance. Nevertheless, Elder et al. (2001) find that test-takers' proficiency is not linked to the enjoyment

experienced when performing tasks and thus enjoying the testing experience is not reflected in test-takers' performance.

There are a number of studies, which employ Skehan's (1996) model of task difficulty (discussed in the Introduction), which differentiates three different factors: code complexity, cognitive complexity and communicative demand. These studies examine relationships and dependencies between task difficulty, task complexity and test-takers' perceptions of task difficulty. Perhaps, researchers report different and at times contradictory results depending how Skehan's (1996) model is applied and operationalized in different tasks.

For instance, Norris et al. (2002) report that test-takers' familiarity with tasks might have been a determining factor accounting for their performance. However, it is important to point out that although these test-takers' perceptions of tasks have given rise to some interesting comparisons about the link between task difficulty and test-takers' performance success, the relationship between test-takers' self-ratings of task ease/difficulty and their performance on them was moderate (Norris et al., 2002). Overall, Norris et al. (2002) did not consider the cognitive processing factors as operationalized in the original task difficulty framework useful for predicting and assessing actual performance difficulty differences among tasks. However, they suggest that further investigation of the relationship between cognitive factors and task-dependent performance evaluations might prove very useful.

As far as test-takers' perceptions of difficulty are concerned as cross-referenced with estimates of task cognitive complexity, Robinson (2001)

concludes the difference of tasks' cognitive complexity is matched by test-takers' perceptions of differences in task difficulty. Task complexity does influence test-takers' perceptions of task difficulty. Test-takers tend to have less confidence in their abilities when completing more complex task versions, they also regard them as more difficult and stressful. However, increasing task complexity as approaching authentic target task demands apparently does not influence test-takers' interest, and/or motivation to complete the task. Moreover, test-takers' affective reactions to task seem to influence their production, in particular their fluency and the lexical variety of their discourse. Test-takers' motivation in particular was found to be positively related to greater lexical variety especially with simple task versions. Another factor that apparently have strong impact on test-takers' perceptions of task difficulty is their role in a test, either as speaker or hearer (Robinson, 2001). Hearers tend to judge simple and complex versions as less difficult and less stressful as compared to speakers due to the higher confidence in their language abilities.

The perspective and immediacy dimensions of task conditions appear to account for test-takers' perceptions of difficulty as well. Test-takers' report that telling a story from a personal perspective and having pictures when telling a story made the task easier for them. Therefore it can be assumed that test-takers' individual factors different from language ability such as visual acuity, memory, personality, and learning style may also influence their perceptions of test difficulty (Elder et al., 2001).

Robinson (2001) concludes that adjusting task cognitive demands has significant impact on test-takers' production, interaction as well as perceptions of difficulty. The most intriguing finding here is that task complexity influences only some aspects of test-takers' perceptions of difficulty such as stress experienced by test-takers but not motivation and interest in completing tasks. Elder et al. (2001) on other hand, take a different stance on the issue of test-takers' feedback. They conducted a study on the difficulty of oral tests exploring four dimensions of task conditions: immediacy, adequacy, perspective and planning time. Elder et al. (2001) conclude that test-takers' perceptions of difficulty are obviously not related to hypothesized difficulty of different task conditions. Therefore, they recommend that test-takers' feedback should not be relied on to a significant extent, neither for the purposes of test design, nor for test validation. This is especially true because test-takers' feedback can be influenced by test-takers' personal characteristics such as gender, social class, professional experience and/or task characteristics as well.

As already demonstrated, the available research in language testing on test-takers' feedback/reactions and in particular test/task difficulty uses a predominantly cognitive approach. Therefore, the following section of this chapter briefly examines the origins and the differences between the cognitive and social approaches in language testing.

2.5. The Cognitive Approach in Language Testing

Cognitively based research in language testing mainly includes quantitative clinical and/or experimental studies involving manipulation and/or measurement of different variables (such as planning time, response time, adequacy, perspective, immediacy, repetition of information, speech rate, test type, number of hearings, role, input source, item type, etc.). For instance, the majority of studies that deal with task/test difficulty take into consideration three main factors: code complexity, cognitive complexity and communicative demand identified by Skehan (1996 and 1998; cited in Elder et al., 2002). Within the cognitive approach there is a link between language ability and language user. It is presented like this: "ability-in language user". However the context in which language use and language testing take place is separate, that is to say that "ability-in language user" and "context" are viewed as two very important but separate entities or variables (Chalhoub-Deville, 2003).

2.6. The Social Approach in Language Testing

The social approach to language testing, which is not that widely applied, is relatively more recent (e.g. Chalhoub-Deville, 2003; Fox, 2003; McNamara, 1995, 1997; Norton and Stein, 1998). It views language as situational, socially imbedded, and inseparable from the actions, activities and participants involved in the testing process. The change from the cognitive to the social approach in language testing and second language studies started in the mid 90s. While the cognitive approach to language testing draws on numerical data, the social

approach looks for typified recurrent patterns. The same development of a social view can be observed in psychology. Within the broad discipline of psychology, sub-disciplines such as ecological, social and discursive psychology have appeared.

One of the first articles that marks the beginning of this change in the field of language testing is Norton and Stein's (1998) article *Why the "Monkeys Passage" Bombed: Tests, Genres, and Teaching* examining cultural bias in a reading test. The article outlines some of the principles of the social approach drawing on the notion of genre, which is viewed as resulting from certain social occasions. The main contribution of this article resides in the implications of the social approach for language testing and teaching that it discusses. Other studies such as those of Bachman and Cohen (1998), Chapelle (1998), and Chalhoub-Deville (2003) undoubtedly demonstrate a transition stage between the cognitive and social approaches both in language testing and second language studies.

Bachman and Cohen (1998), and Chapelle (1998) examine interfaces between SLA and language testing. Bachman and Cohen (1998) identify research questions, empirical approaches, and interfaces such as variability and performance, which are common to both fields. As far as variability is concerned, its main sources are considered to be 1.) individual differences in the language abilities acquired or measured, 2.) individual differences in the strategies and other processes that learners and test-takers use, 3.) variation in the tasks and context and their effects on language use and test performance.

In this article, test performance is related to communicative language competence or communicative language ability. The authors perform thorough examination of the development of the theories concerning the communicative language competence including language use strategies and ways for their measurement. Although Bachman and Cohen (1998) take into account the importance of learners' individual language abilities, learners' individual language strategies as well as the effect of task and context variation, their work occupies a midpoint position as far as the transition between the cognitive and social approaches are concerned. What prevents this work from crossing the divide on the side of the social approach is the way certain concepts are defined. For example, language use strategies are defined as "mental operations or processes that learners consciously select and apply when attempting to accomplish language use tasks" (Bachman and Cohen's, 1998, p. 14). Strategic competence is considered to include meta-cognitive strategies. In addition, when examining the relationship between learners' strategies and their performance on tests, Bachman and Cohen's (1998) examine the relationship from a cognitive point of view by referring to a classic cognitive study - that of Purpura (1998).

As demonstrated all the processes and strategies involved in language learning and testing, according to these authors, are essentially cognitive phenomena, which to some extent might be influenced by social features such as variation in tasks and context. However, this work contains a very encouraging recommendation for further research of the interfaces between SLA and language

testing. This is the suggestion for combining qualitative and quantitative research methods.

Chapelle's (1998) work is interesting mainly because of the definition of language ability as a construct provided. The construct is defined as "a meaningful interpretation of observed behaviour" (p. 33). When discussing performance and vocabulary, the author refers to definitions produced by trait theorists, behaviorists, and interactionists. In fact, the use of these definitions marks Chapelle's (1998) work as transitional between the cognitive and social approaches in SLA and language testing.

Chalhoub-Deville (2003) in review of different available models of the L2 construct makes a fairly good distinction between the cognitively based models and the social interactional perspective on L2 construct. Her article also can be regarded as a transitional one between the cognitively based studies on language testing and the socially oriented ones. In fact, the article even discusses a social-cognitive L2 construct representation. A main criticism to it and a reason why I would not classify it as belonging to the group that uses purely social approach to language testing is the terminology used. Basically, the Chalhoub-Deville's (2003) article presents a dichotomy between the cognitively based L2 construct and the interactional perspective on L2 construct. The interactional perspective on L2 construct is socially based, but the social aspect is shown as its subcategory rather than as its definition.

There is also a distinction made between interactional competence and social interactional perspective. The interactional competence is a "single

interacting structure” (Chalhoub-Deville, 2003, p. 373) including: “ability-in language user-in context”. The social interactional perspective, for its part, is viewed as dealing with two fundamental challenges. First, language use in communicative situations is a dynamic discourse co-constructed among participants. And second, language ability is local. In my opinion, rather than distinguishing between interactional competence and social interactional perspective, these two should be considered together or as cited in the conclusion part of this article: ‘*consider context*’ (emphasis in original) (Anastasi, 1986 cited in Chalhoub-Deville, 2003).

To summarize the distinction between the cognitive and social approach in language testing, one possible advantage of the cognitive approach is that it allows making generalizations across different contexts as it focuses on test-takers’ internal abilities. The social approach, on its part, promotes as stated in (Chalhoub-Deville, 2003):

the complex interaction of linguistics and non-linguistic knowledge, cognitive, affective, and conative attributes engaged in particular situations. (p. 380).

The desired outcome of social-interactionally based research is first, to better theories used in testing and the practical application of these theories, and second to contribute to the development of more detailed theories regarding the L2 construct in terms of “ability-in language user-in context”.

Swain’s (2001) article on oral interaction through using dialogue in small groups, also examines interfaces between SLL¹² and language testing. However,

¹² In the referred works SLA and SLL are used interchangeably, the terms are not subjected to the difference made by Krashen (e.g. 1985, 1994).

Swain's (2001) article can be considered to be, at least, one step further towards the social approach to language testing compared to the previously discussed works. Swain (2001) looks at dialogues in small groups as opportunities for language learning and thus as ways of joint construction of knowledge. The author addresses her ideas as socio-cultural theory of mind, drawing her theoretical perspective from Vygotsky (1978, cited in Swain, 2001). She views dialogue with others and with the self as serving both communicative and cognitive functions. In Swain's understanding the origins of the higher mental i.e. cognitive processes such as attending, predicting, planning, monitoring, and inferencing are social. This means that these processes are mediated by the interaction between individuals. The other main point in Swain's (2001) theoretical framework is that "knowledge is constructed through dialogue" and that "through dialogue participants co-construct knowledge" (p. 280) and that this leads to learning of the second language. Learning is considered to be a continuous process happening through learners' participation in "situated joint activities" (Swain, 2001, p. 281). However, in Swain's (2001) work there is still cognitive presence, the author addresses the discussed ideas about SLL as "neo-Vygotskian socio-cognitive perspective" (p.281), language learning and test performance for that matter are essentially cognitive phenomena mediated by social interactions.

As far as the application of the social approach to language testing is concerned, the works of McNamara (1995, 1997, 2001) and that of Fox (2003) are currently the ones that have completely adopted it.

In his 1995 and 1997 articles McNamara examines performance assessment. In the 1995 article he reviews different models of performance. McNamara's overall conclusion is that language performance as it involves a number of interactants such as a test-taker and interlocutor (other test-taker, a trained native speaker or rater) can be influenced by various variables¹³. In the author's opinion understanding the effect of these variables on test-takers' scores is of major importance and is a very fruitful research area. McNamara (1995) highlights that an adequate model of the factors influencing test-takers' performance that includes non-linguistics factors both cognitive and non-cognitive needs to be modeled.

In the author's opinion the available models of performance have a major common drawback that is their extensive focus on the individual test-taker rather than on the test-taker in interaction, given the interactive nature of performance assessment. McNamara (1995) goes further in explaining that the discussed models of performance "are loosely psychological in orientation" (p. 174), and include social interaction in a static way where strategic competence examines interaction from a cognitive viewpoint. What he argues for is describing performance as involving social interaction.

In the 1997 article McNamara focuses on the nature of interaction in performance, which he claims to be intrinsically social. He distinguishes between two definitions of interaction in language testing research. The first one is

¹³ For example, age, gender, educational level, language proficiency, native/non-native speaker status, and personal qualities.

“loosely psychological” (p. 447) dealing with mental/cognitive activities and emphasizing on a single individual. The second one is social /behavioral where joint construction of performance is based on joint behaviour between individuals. What McNamara (1997) suggests is that language testing should adopt exclusively social perspective on interaction rather than the predominant one, i.e. the cognitive. One of his contributions to the establishment of the social approach in language testing consists in the view that eliciting and interpreting test performance is a social activity. In fact, both in this and especially in a later article McNamara (2001) states that there is a growing understanding of the social character of the constructs of language testing. He adds that language use is seen as a social activity, and that language assessment and procedures are social facts. Most importantly he emphasizes that the test-taker is not “an isolated figure” (McNamara, 1997, p. 459), entirely responsible for his/her test performance.

When discussing Bachman’s (1990) interpretation of the term interaction in language testing, for instance, McNamara (1997) points out that even when referring to social interaction, Bachman means cognitive activity performed by the test-taker. The author goes further in explaining the danger of defining interaction, test-takers’ abilities, and performance in predominantly cognitive terms. Performance cannot be regarded as an exclusive projection of test-takers’ abilities, as it is inherently social act involving interlocutors, test developers, raters, and elicitation and interpretation of test scores.

McNamara’s (2001) article presents a more recent update of the view of language testing and all the procedures related to it as inherently social in their

nature. A significant and worthwhile aspect of this article is the discussion of Messick's (1989) theory of test validity and more specifically its social dimension. McNamara (2001) points out the extreme importance of the social consequences of test use cell in the validity matrix developed by Messick (1989). The presence of social consequences cell in relation to test validity once again confirms the social character of the entire process of language testing.

In this regard, Fox's (2003) article on bias detection is currently the one that draws on the social approach to language testing to its full extent. Fox (2003) applies what she refers to as an ecological perspective building upon Bronfenbrenner (1997, cited in Fox, 2003). She suggests the inseparable relationship between individuals and their social contexts is applicable within the environment of language testing in terms of the interactions between test-takers, test-raters and the other stakeholders who are involved in the process of testing, and who are inter-connected by actions related to language tests. McNamara (1997) emphasizes the importance of the perceptions of the stakeholders involved in language testing as those perceptions provide an insight into the nature of the testing process. Fox (2003) goes further in explaining that feedback from test takers and test raters can play an important role in the various practices of language testing such as test development, construction and validation.

As Fox (2003) notes sources of the ecological approach to language testing (Fox, 2002, 2003) can be traced to the Rhetorical Genre Studies (RGS) in North America (e.g. Freedman and Medway, 1994; Paré and Smart, 1994) as well as to the discursive psychology (see Edwards and Potter, 1992). Although a

detailed discussion of RGS and discursive psychology are beyond the scope of this thesis, it is important to provide some background on them.

RGS (e.g. Freedman and Medway, 1994; Miller, 1994) regard language, and spoken and written genres as social action. This view of genre is central to composition and language studies in North America. Within this theoretical framework knowledge is regarded as inherent to the social context. Writing is considered to involve not only the final product i.e. a text and a writer but also the social context writing occurs in (Freedman and Medway, 1994). The social context is typically viewed to include needs, goals, perceptions, and attitudes of participants in the writing process.

This understanding of writing is accounted for by genre theory, which considers different genres as responses to recurring social situations (Freedman and Medway, 1994). Giltrow (2002) represents genre as a combination of form and social situation, the use of genre is influenced by the characteristics of the social situation. Within this view of genre, formal features of a text “derive from and relate to the writer’s social motive in responding to a recurrent social situation of a certain type” (Freedman and Medway, 1994).

Thus, any piece of writing is not abstracted from its social context but is rather seen as a generic way of responding to particular social situations. Further these social situations and responses are seen as resulting from shared social interpretations and are part of a socially constructed reality (Miller, 1984), rather than being objective, independent realities (Freedman and Medway, 1994). According to Bakhtin (1986), whose ideas are fundamental for the social

constructionist view of RGS, genre is a reply to typical situations of speech communication, typical themes under certain typical circumstances.

As mentioned by Norton and Stein (1998) different genre analysis has been used in a wide variety of disciplines but only scarcely and recently in the area of language testing. Genre as it is used here does not mean a “text type” such as a play, a poem or a novel (Norton and Stein, 1998). It is a reply, as already mentioned, to typical situations (Bakhtin, 1986). Genres are combinations of form and situation (Giltrow, 2002) and their use are influenced by the characteristics of the social context.

Discursive psychology, the other area which can account for the nature of the social and more specifically the ecological approach to language testing (Fox, 2002, 2003) deals with people’s practices such as communication, interaction, argument as well as the reorganization of these practices in different kinds of settings (Edwards and Potter, 1992). It is concerned with how psychological entities and processes are constituted in discursive acts, and in this respect discursive psychology cannot be related to individual psychology. Even when an individual speaker produces an utterance, this person is not in complete control of their talk/writing.

The focus of discursive psychology is the action direction in writing and talk. The primary issue for researchers and participants in this area is social actions done through discourse. The discursive action model (DAM), which is based on discursive psychology, is concerned with the nature of knowledge, cognition and reality. It is designed as a scheme for conceptualizing phenomena.

Events, factual reports and cognitive states are defined as discursive topics.

According to DAM these discursive topics are topicalized or used by people for orientation in discourse.

Contrary to approaches in cognitive psychology, discursive psychology does not view these discursive topics or constructions as expressions of speakers' underlying cognitive states or as mental constructs, representations or processes but rather as situated in the context of the social actions in which they occur. Discursive psychology combines discourse studies on concepts such as attitude and belief (e.g. see Billig, 1991, cited in Edwards and Potter, 1992), categorization and schematization (e.g. see Edwards, 1991; Potter, 1988, cited in Edwards and Potter, 1992), social representations (e.g. see Billig, 1988; Potter and Wetherell, 1987, cited in Edwards and Potter, 1992), and conversational remembering (e.g. see Edwards and Middleton, 1988, cited in Edwards and Potter, 1992). Basically, within the framework of DAM, remembering is regarded as "situated production of versions of past events" (Edwards and Potter, 1992, p. 3). Attributions are defined as inferences occurring from these versions, which are viewed by participants as implied.

In order to study natural interactions, researchers need to examine discourse in terms of actions performed in speaking, in texts and/or other writing formats. In everyday settings, discursive actions do not occur in isolation but as a part of activity sequences, which are considered to be a primary substance of human life and to give sense to the individual discursive acts which activity sequences are composed of.

As far as the issue of treatment of language is concerned, and more specifically the grounds on which underlying knowledge and reasoning are based, discursive psychology claims that if there are underlying mental structures involved then they are only those which allow people to perform 1.) social actions in talk, 2.) the “situated and occasion version production” (Edwards and Potter, 1992, p. 158), and 3.) rhetoric. In discursive psychology’s point of view these underlying mental structures should not be used for speculation of what people may think. The study of situated discourse places language (that of cognition or that of reality) as representation in a subordinate position to language as action.

In fact, DAM can be used to provide detailed specifications about reports (rememberings) and attributions (inferences) (Edwards and Potter, 1992). Reports of actions in DAM are designed rhetorically and have two main features. The first one is that reports occur in contexts of speech acts such as dialogues, disputes and conflicts. The second feature is that reports are designed for their accuracy so that they can undermine alternative versions and resist being claimed as false, incomplete or biased.

According to DAM, memory and attribution connect in the area of rhetoric and this way they are part of rhetorically designed descriptions of how world operates. Attitudes, defined in psychology as representations of underlying personal stands on different issues, are also organized rhetorically to resist undermining by alternative versions.

Discursive psychology focuses particularly on discourse and the organization of discourse into activity sequences. Since participants’ practices cut

across different subfields in psychology, it is considered that their practices can be linked to social sciences and humanities. Moreover, within the framework of DAM some of the major concepts in social and cognitive psychology such as categorization, personality/self and attitudes can be defined in discursive terms.

Categorization is a major source of inferences. This is done through assigning categories, commonly understood properties and attributes given to persons, events and/or objects. In DAM categories operate in the area of fact construction, it is assumed that certain features of social categories belong to peoples' knowledge, experience and/or skills. As far as personality/self are concerned, in DAM they are not regarded as psychological objects but as discursive resources that people use when they are engaged in interactional work. In natural interaction, attitude is a rather complex and multifaceted concept. It can be used to indicate psychological involvement. For instance, attitude can enable participants to diminish a claim or an argument.

Discursive psychology and DAM interpret people's practices as social actions and classify them into different social settings. Moreover, people's social practices such as talk and/or writing are not viewed as individual acts but as influenced by the surrounding social discourse. Categorizations, as defined within DAM, are viewed as social categories according to which people's knowledge, experience and skills are organized. Attitudes allow revealing and understanding participants' accounts of events. Through categorizations in DAM, people's reports (rememberings) and attributions (inferences) can be organized into social categories and certain features can be distinguished.

In this relation, both discursive psychology (Edwards and Potter, 1992) and RGS (e.g. Freedman and Medway, 1994; Paré and Smart, 1994) identify typified and recurring patterns of response which can account for participants'/students'/test-takers' performance on a test or a task (Fox, 2003). Given that reports, accounts and attributions in discursive psychology can be used to study the mind as social and contextualized (Fox, 2003) and that typified rhetorical responses in RGS account for recurrent social situations, these two approaches can be applied to testing situations. Therefore, investigation of the discursive acts, occurring in the testing-taking process, and of test-takers' reports (rememberings) and attributions (inferences) (Edwards and Potter, 1992) of their test-taking expectations and experience, can provide more insight into the "subjective ecology" (Zeidner and Bensoussan, 1988, p. 100) of the entire testing process. It can probably help detect possible areas of the testing process that has been neglected that are problematic or need to be improved.

The ecological approach to language testing (Fox, 2002, 2003) looks at the complex relationship between test-takers, social context in which testing takes place and the interaction between test-takers, tests and testing situations. What is accounted for within the framework of this approach, is recurrent patterns and themes in test-takers accounts of their test-related experience. Moreover, applying the ecological approach to language testing (Fox, 2002, 2003) test-takers' perceptions may hold the key to: 1.) predicting performance on tests (A. Brown, 1993; Elder et al., 2002; Fox, 2003; and Hill, 1998), 2.) providing valuable insight in with regard to the validity and reliability of tests (Bachman, 1990; Davies et al.,

1999; Messick, 1989), and 3.) providing information for future test development (Alderson, 1988; A. Brown, 1993; and Stansfield, 1991). When examining test-takers' perceptions of a language test, the recurring patterns in test-takers' responses may be a valuable source for predicting their performance. Therefore, it seems reasonable to examine further the nature of test-takers' perceptions and the factors that account for them in relation to test-takers' overall performance on a test.

Using this approach to language testing (Fox, 2002, 2003), the research questions asked in this thesis are the following:

- 1.) How do test-takers' perceive a high stakes test?
- 2.) What characteristics (e.g. test-takers' language proficiency, test-takers' perceptions of their language proficiency, test difficulty, fairness, the testing situation itself) seem to be related to test-takers' perceptions?
- 3.) What is the relationship, if any, between test-takers' proficiency and their perceptions of the test?
- 4.) Is it possible to use test-taking strategies in order to compensate for insufficient language proficiency and knowledge?

The next chapter presents the methodology of the study and includes the analysis section.

CHAPTER THREE: METHODOLOGY

The introductory chapter of this thesis presented an overview of SLA and SLL research on affective factors in language learning, and definitions of perception from different perspectives. Chapter Two contrasted the cognitive and social approaches in research. It also examined testing research on test-takers' feedback and reactions to tests as well as research on difficulty factors in language testing, and test-takers' perceptions of task/ test difficulty. At the end of Chapter Two, the research questions regarding test-takers' perceptions of tests were introduced. The present chapter examines the methodology of a study designed in order to investigate those research questions. It also includes the analysis section.

3.1. Method

The main reason for undertaking this study was to compare test-takers' perceptions of a high-stakes test (the CAEL Assessment) before and after taking the test.

Phase One of the study was an analysis of CAEL test-taker responses to post-test questionnaires (Appendix A), which are routinely filled out by test-takers after writing a CAEL test. Phase Two of the study was comprised of two parts. In Phase Two, (Part One) I utilized semi-structured interviews (Appendix B) in order to explore three test-takers' perceptions before and after taking the CAEL test. In Phase Two, (Part Two) I analyzed test-takers' journals in order to

triangulate the findings from Phase One and Part One of Phase Two. In order to collect the journal data, prior to writing the test, the participants were asked to record their test-related experiences and in particular their expectations, and feelings about the test. Data obtained through the three research instruments: the post-test questionnaires (Appendix A), the semi-structured interviews with the three participants (Appendix B), and the journals, were compared for the study. Table 1 below provides an overview of the structure of the study.

Table 1: Structure of the study

Phase	Research Instrument
Phase One	Post-test questionnaire (Appendix A)
Phase Two, Part One	Semi-structured interviews (Appendix B)
Phase Two, Part Two	Journals

3.2. Participants:

Phase One

A randomly drawn sample of 200 test-takers across three CAEL

Assessment versions responded to the post-test questionnaire (Appendix A). The sample consisted of 60.3% male and 39.7% female test-takers. The information regarding their first language is presented in Table 2.

Table 2 - Test-takers' five most commonly reported first languages

First Language	Percentage
Mandarin	69.9%
Arabic	7.7%
Cantonese	5.3%
Farsi	3.3%
Vietnamese	2.9%

The information regarding the test-takers' intended area of study is presented in Table 3.

Table 3 - Test-takers' five most commonly reported intended area of study

Area of study	Percentage
Business	24.4%
Economics	22%
Engineering	14.8%
International Affairs	9.10%
Sciences	8.6%

The sample of 200 test-takers differs from the typical CAEL test-taker population previously reported in the CAEL Assessment: Test Score and Users' Guide (Fox, 2002a). Due to the influx of Asian test-takers, however, the sample is representative of the current CAEL test-taker population (Fox, 2004).

Phase Two

Three participants planning to take the CAEL were recruited for Phase Two¹⁴.

¹⁴ Prior to the study, Carleton University Research Ethics Committee conducted an ethics review of these procedures and granted approval.

All of the participants needed to take the CAEL test in order to be admitted to a Canadian university. The participants were given pseudonyms in order to protect their anonymity: Wendy, Dave and Chloe.

The participants had all taken the CAEL test on the 31st of January 2004 before they volunteered to participate in the research. Participant Two, Dave, had taken the CAEL test more than once. The three participants were native speakers of different languages, had different educational and English language learning backgrounds, and levels of English language proficiency. A case description of each participant is provided in the following chapter.

The Educational Setting

This study was situated within a particular educational setting. The English as a Second Language (ESL) courses at this university are divided into Intensive ESL courses and credit ESL/EAP (English for Academic Purposes) courses.

The Intensive ESL courses cover all skills from beginning to advanced levels (IC 10, IC 15, IC 20, IC 25, and IC 30), all within the framework of general introductory skills and strategies. Students attending Intensive ESL courses lack sufficient English language proficiency required in order to take credit ESL courses or to be regular students. Many of these students aim to get into the credit ESL courses by scoring at least 40 on the CAEL test.

The credit ESL/EAP (English for Academic Purposes) courses (ESLA 1300, ESLA 1500, ESLA 1900 or ESLA 1950) are offered to students who are admitted to a degree program, and whose language skills require improvement. While taking a credit ESL course, students are allowed to take regular classes in their degree program. The credits obtained from the credit ESL classes are considered towards students' degrees. The number of regular classes that students are able to enroll into depends on their credit ESL level. Students' initial level is determined by a CAEL Assessment placement and then students can progress through the remaining levels by satisfying a number of grade or CAEL requirements. The ESL / EAP credit courses are highly intensive and help students to deal with academic strategies in writing, reading, oral presentations, research and other skills students need to successfully complete their university program. Each ESL / EAP course normally is long one term, with the exception of ESLA 1905, which is offered to engineering students only and lasts an entire academic year.

3.3. Instruments:

a.) Canadian Academic English Language (CAEL) Assessment

The Canadian Academic English Language (CAEL) Assessment (Fox, 1995, 2001, 2003) is a criterion-referenced, fully integrated topic-based test of English for academic purposes (EAP). It is used for admission purposes at a number of Canadian universities and was created in response to a high number of

non-native speakers of English being admitted with TOEFL (paper-based) scores of about 560-580¹⁵, and then not being able to satisfy the academic demands they faced in their university studies. The CAEL test requires test-takers to perform tasks they will be faced with in their everyday university studies, for example: listening to academic lectures, taking notes, and writing academic essays.

The CAEL Assessment is divided into four skill areas: reading, writing, listening, and speaking. It is comprised of four sections: a lecture, two readings, writing an academic essay all on one topic and an oral test. The Oral Language Test (OLT) is a task-based, tape-mediated test and like the other sections of the test the OLT tests English for academic purposes (EAP) (Fox, 2000). However it is not related to the topic of the other sections of the test. The information contained in it is related to completing tasks that test-takers will be faced with in an everyday university setting.

All test topics are academically orientated and based on first-year, introductory courses offered in different faculties and departments at Carleton University during the first two weeks of classes. For test development purposes, courses offered at Carleton University are studied, and then suitable topics within the range of the courses offered are located.

¹⁵ The TOEFL scores were raised several times in an effort to recruit students with sufficient levels of proficiency for academic study.

At the beginning of the test, test-takers are presented with a statement to which they should respond to at the end of the test by agreeing or disagreeing with it. The reading section of the CAEL test consists of two readings as academically genuine as possible. Test-takers are required to read the texts and answer questions based on them. The listening section of the test is an academic lecture on the same topic. While listening to the lecture test-takers take notes, answer questions, and transfer information into charts and tables. For the writing part of the test, test-takers are required to write an academic essay using the information provided in the reading and listening sections of the test. The scores of the test are divided into criteria-referenced bands from 10 to 90, both for each of the sections of the test as well as for the overall score. Due to the construct representation of the test, the writing section of the test has greatest value.

b.) Post-Test Questionnaire (*Phase One*):

The test-taker post-test questionnaire (Appendix A) is routinely completed by test-takers who take the CAEL test (Fox, 2003). Test-takers are asked to fill-out questionnaires at the end of the test. They are given as much time as they need to answer the questions in the questionnaire. Test-takers are informed that the feedback they provide will be used for improving the quality of the test for future test-takers. However test-takers are free not to fill-out the questionnaire if they choose to do so. Although the questionnaire includes a space for test-takers' name and test number, test-takers' confidentiality is guaranteed. They are assured that

their responses will not influence in any way their scores on the test. Most test-takers choose to provide their names and test number.

The reasoning behind having the space for test-takers' name and test number has to do with research on test-takers' feedback regarding test administration, test characteristics such as time limits, test topics, sample test availability and usefulness, and test-takers' characteristics such as gender, language background and language proficiency.

The questionnaire itself is comprised of 6 questions, questions #1, 2, 4 and 6 are open-ended questions, questions # 3 and 5 are Likert-type scaled questions. Questions # 1 and 2 ask test-takers what they liked and disliked about the test. Question #4 asks test-takers for suggestions for improving the test. Question #6 asks test-takers if they liked the topic of the test. In Question #3 test-takers are asked to *agree*, *disagree* or choose *no opinion* in regards to statements about the test: its fairness, if it was good experience, if it would reflect their true knowledge of English and if the test topic was important for them. Question #5 ask test-takers to rank different statements from 1 to 5 where 1 is most important and 5 is least important. These statements ask test-takers about the physical comfort of the room, the amount of time allotted for the test, the topic of the test, the sound quality of the tape recorder, and the sample practice test before the actual test.

However, it should be noted that the data obtained through the questionnaire might not represent the test-takers' true opinion about the test. This might be due to the fact that questions in the questionnaire limit the test-takers'

focus to specific issues. They may not have identified these as issues if they freely discussed the test. There also might be a difference between what the test-takers wrote about the test and what they actually thought about it.

c.) Semi-Structured Interview Questions (*Phase Two*):

There are two sets of semi-structured interview questions (Appendix B), the first set (titled *before*) was administered before the participants took the test and the second set (titled *after*) was administered after the participants took the test. Because of the semi-structured nature of the interview questions, the participants were asked additional questions resulting from the course of the conversation during each interview.

Both sets of interview questions asked the participants for their opinion regarding:

- the testing experience in general
- what they think about the test
- their emotional state before, during, and after the writing

the test

- their expectations about the test and their test scores
- their opinion about the difficulty of test sections
- recommendations for its improvement

Nevertheless, it should be highlighted that the data obtained through the semi-structured interviews might not represent the participants' true opinion about the test. This could be due to the fact that there might be a difference between

what the participants in Phase Two said about the test and what they thought about it.

d.) Journals

The journals were designed for the participants to record their test-related feelings and experience before writing the test. The participants were not required to write essay-type entries, they could write down sentences, jot-notes phrases and/or words. The participants were asked to focus on their expectations about the test, their perceptions of test difficulty and feelings related to the testing process such as motivation, anxiety, etc. They were asked to write at least one entry a day during the time prior to the test. The length of each entry was not specified in order to allow the participants' to respond freely, without restrictions. It was hoped that their accounts would thus be more genuine. The test-takers' journals were used to triangulate (Nachmias and Nachmias, 1987) the data obtained from the CAEL test-taker post-test questionnaires and interviews. It is important to note that the participants did not keep their journals regularly. Dave did not keep a journal at all. Instead he sent the researcher an email explaining that he did not think about the test in his free time, did not feel comfortable keeping a journal and did not feel nervous or excited about the test.

It should be pointed out that the data obtained through the journals might not represent the participants' true experience prior writing the test. There might be a difference between what the participants wrote in their journals and what their actual experiences regarding the test were before writing it.

At this point it is important to explain that any findings and/or conclusions about the test-takers'/ participants' perceptions of the test are based on the test-takers'/ participants' accounts of their perceptions regarding the test.

3.4. Procedures:

The main focus of the study is on test-takers' perceptions¹⁶ of a high-stakes test and more specifically the relationship between test-takers' proficiency and their perceptions of the test. The test-takers' perceptions that were taken into special consideration were test difficulty, fairness, the entire testing experience, etc.

Phase One:

Phase One of the study included analysis of 200 CAEL test-takers' responses to the post-test questionnaires (Appendix A). The sample of test-takers was randomly drawn across test-takers writing three test versions. The purpose of Phase One was to identify recurrent patterns in test-takers' responses to the questions included in questionnaire. An important part of the analysis of the questionnaires was to establish a relationship, if any, between test-takers' scores/ proficiency and their perceptions of the test.

¹⁶ When I refer to perceptions using Edwards and Potter (1992), I acknowledge that these are actually test-takers' *accounts* of perceptions.

Phase Two:

Phase Two of the research included interviews with three students planning to take the CAEL test about their perceptions before and after taking the test. These students are referred to in the following parts of the thesis as participants. As a compensation for their participation in the interviews and keeping journals, the participants were offered a CAEL test free of charge.

The participants were assured that their anonymity would remain confidential. The participants' interviews and journals when referred to in the analysis, discussion and conclusion parts of the thesis are addressed by using the pseudonym assigned to each participant. The participants were assured, in writing and orally, that they would be free to leave at any time during the interviews and/or to decline answering any question(s) that they felt uncomfortable with.

The students who volunteered to participate in the research were contacted through email by the researcher and meetings were arranged. Each participant was informed about the purposes of the research and the interviews before taking the test were conducted. Each interview lasted between 40 and 45 minutes and took place at a mutually agreed to a place on campus. The participants' responses to the interview questions were taped. After conducting the interviews before taking the test, the participants were asked to start keeping journals about their expectations about the test e.g. test difficulty, scores as well as their feelings, motivation, nervousness, etc. about it. There was a period of about ten days between the interviews before taking the test and the time when the participants wrote the CAEL test. Wendy wrote the CAEL test prior to the other participants.

Dave and Chloe wrote the test at one and the same time. The interviews after taking the test were conducted about 12 days after the participants wrote the test. Before conducting the interview after taking the test, Dave chose to register and take the CAEL test once again. The relatively long period of time between writing the test and conducting the interviews after taking the test was due to the unavailability of participants to be interviewed earlier. All the participants were informed about their CAEL test scores before conducting the interviews after taking the test. For the purposes of test security, test-takers were not allowed to have the test reading and question booklets for reference when answering the interview questions. Therefore, they were not asked to give feedback for specific questions.

It is important to note that when writing the CAEL test for the purposes of this research, the participants did not take the Oral Language Test (OLT) again. Instead their OLT scores from their previous CAEL written on the 31st of January¹⁷ were considered. After conducting the before and after interviews, the tapes were transcribed and the data were coded and analyzed.

The data obtained through the three research instruments was compared and final conclusions were made.

Once again it is important to mention that there might be a difference between what the test-takers (Phase One) and the participants (Phase Two) reported about the test and their actual opinion about it. The findings and/or

¹⁷ The participants were not required to complete the OLT again because of the weighted average of the test subsections. Given the weight of the OLT for the overall test bandscore, the participants' bandscores would not have been affected if they had taken the OLT again.

conclusions presented later in this thesis are based on the test-takers/ participants accounts of their perceptions of the test.

3.5. Analysis:

Different analytic approaches (such as frequency, crosstabulation with chi-square analyses, and a modified grounded theory approach) were applied to the two Phases of the study.

3.5.1. Phase One: Analysis of Post-test questionnaire:

The randomly drawn sample of 200 test-takers that responded to the post-test questionnaire (Appendix A) wrote one of three CAEL Assessment versions. Frequency and chi-square statistical analyses ($\alpha < .05$) were applied to the data to investigate meaningful trends and significant relationships.

Responses to the open-ended questions in the post-test questionnaire (questions # 1, 2, 4 and 6) were coded using techniques proposed by Maxwell and Miller (1992, cited in Fox, 2003, p. 32). The responses to open-ended questions (questions #3 and 5) were organized into categories (see Fox, 2003, p. 32) using a modified grounded-theory approach (Creswell, 1998)¹⁸.

With regard to the research questions of this thesis, crosstabulation with chi-square analyses examined meaningful trends and tested for significance in

¹⁸The traditional grounded theory approach used multiple samples of data (see Glazer and Strauss, 1967). The modified grounded theory approach, used in this study, draws on ethnographic data. It applies an open coding approach to the research data with a focus on identifying repeating themes and patterns.

relation to the following categories: time limits, student comfort¹⁹, test version/topic²⁰, and test-takers' overall bandscores, which emerged as key in the analysis of the post-test questionnaires.

1.) Time limits

First, I used crosstabulation with chi-square analyses to examine meaningful trends and significant relationships between test-takers' stated perceptions of whether they had enough time to take the test and whether the test:

- was fair
- was good experience
- topic was important
- topic was interesting
- reflected true English knowledge

Because test-takers' perceptions of the test (e.g. test was fair, etc.) initially consisted of responses to a six-point Likert scale (1=strongly disagree, 6=strongly agree), these were analyzed for frequency of response. For analyses with crosstabulation and chi-square²¹ they were then recoded into different variables with 1 representing strongly disagree, somehow disagree, and disagree; and with 2 representing agree, somehow agree, and strongly agree.

In addition, two key categories: 1.) test version (i.e. test topic) and 2.) test-takers' overall bandscores were examined in relation to test-takers' perceptions of

¹⁹ In the initial coding of the data from the post-test questionnaire, the category *student comfort* was used to represent the questions regarding the sound quality of tape and the physical conditions in the testing room.

²⁰ In this case, the terms *topic* and *test version* are interchangeable.

²¹ Chi-square analysis uses nominal data.

the test. Response rates to individual questions in the post-test questionnaire as well as the average response rate are presented in Appendix C.

2.) Student Comfort

Crosstabulation with chi-square analyses was used to examine meaningful trends and significant relationships between test-takers' accounts of their perceptions of their comfort during the test and test-takers' accounts of their perceptions of whether the:

- test was fair
- test was good experience
- test reflected true English knowledge
- time was enough to take the test

Analyses of test-takers' stated perceptions of their comfort during the test and the test version they wrote were also performed.

Test-takers' indications of their perceptions of the test (e.g. test was fair, sound quality of tape was good, etc.) initially consisted of responses to a six-point

Likert scale (1=strongly disagree, 6= strongly agree). These were analyzed for frequency of response. For crosstabulation and chi-square analyses with the test-takers' stated perceptions were then recoded into different variables where 1 stood for strongly disagree, somehow disagree, and disagree; and 2 stood for agree, somehow agree, and strongly agree.

3.) Test version (i.e. test topic):

The crosstabulation with chi-square analyses also investigated meaningful trends and significant relationships between test version and test-takers' accounts of their perceptions of whether the:

- test was fair
- test was good experience
- test reflected true English knowledge
- test topic was important
- time was enough to take the test
- test topic was interesting
- knowledge of test topic was helpful

Test versions were also examined in relation to the test-takers' overall bandscores.

Test-takers' stated perceptions of the test (e.g. test was fair, etc.) initially consisted of responses to a six-point Likert scale (1=strongly disagree, 6=strongly agree). These were analyzed for frequency of response. For crosstabulation and chi-square analyses with the test-takers' indications of their perceptions were then recoded into different variables where 1 stood for strongly disagree, somehow disagree, and disagree; and 2 stood for agree, somehow agree, and strongly agree.

Test-takers' overall bandscores were also recoded in three different ways. First, I coded them into 1 standing for low proficiency for bandscores from 10 to 30; 2 standing for medium proficiency for bandscores from 40 to 60, and 3

standing for high proficiency for bandscores from 70 to 90. Second, I identified 1-low proficiency with bandscores from 10 to 40, and 2- high proficiency with bandscores from 50 to 90. Third, 1 stood for low proficiency with bandscores from 10 to 30, and 2 stood for high proficiency with bandscores from 40 to 90.

4.) **Test-takers' overall bandscores:**

Using crosstabulation with chi-square analyses I investigated meaningful trends and significant relationships between test version and test-takers' accounts of their perceptions of whether the:

- test was fair
- test was good experience
- test reflected true English knowledge
- test topic was important
- time was enough to take the test
- test topic was interesting
- knowledge of test topic was helpful

Test-takers' overall bandscores were recoded in three different ways. First, I coded them into 1 representing low proficiency for bandscores from 10 to 30; 2 representing medium proficiency for bandscores from 40 to 60, and 3 representing high proficiency for bandscores from 70 to 90. Second, I identified 1-low proficiency with bandscores from 10 to 40, and 2- high proficiency with bandscores from 50 to 90. Third, 1 stood for low proficiency with bandscores from 10 to 30, and 2 stood for high proficiency with bandscores from 40 to 90.

3.5.2. Phase Two: Analysis of Semi-Structured Interviews and Journals:

Phase Two was comprised of two parts: semi-structured interviews with three participants (Appendix B) and these participants' journals kept prior to taking the test. The semi-structured interviews (Appendix B), and the journals were analyzed again using a modified grounded theory approach (Creswell, 1998). An open coding approach looking for repeating themes and recurrent patterns in the participants' accounts of their test-related experience was applied. As previously indicated, the results obtained from the participants' journals were mainly used for triangulating (Nachmias and Nachmias, 1987) the results obtained from the two other research instruments: the post-test questionnaires (Appendix A) and the semi-structured interviews (Appendix B).

The next chapter includes the results obtained from the data analysis of Phases One and Two of the study. The Chapter starts with an overview of the key findings of the analysis of the post-test questionnaire.

CHAPTER FOUR: RESULTS

This Chapter discusses the results obtained from the analyses of Phases One and Two of the study. 1.) Phase One analysed the responses of a randomly drawn sample of 200 test-takers to the post-test questionnaires (Appendix A). A modified ground theory approach (Creswell, 1998) was used in coding the responses to the open-ended questions in the post-test questionnaire. Frequency and crosstabulation with chi-square statistical analyses ($\alpha < .05$) were applied to the data from the post-test questionnaire. 2.) Phase Two analysed the results obtained from the interviews (Appendix B) with the three participants and their journals. The results obtained from the latter were used to triangulate (Nachmias and Nachmias, 1987) the results from the post-test questionnaires (Appendix A) and the interviews (Appendix B).

4.1. Phase One: Post-Test Questionnaire

4.1.1. Frequency analyses:

The overview of the test-takers' responses to the open-ended questions on the post-test questionnaire²² revealed that the most frequently mentioned categories were:

²² It should be taken into account that the completion of the post-test questionnaires is completely voluntarily. Test-takers are assured that their responses will not influence in any way their scores on the test for future test-takers.

- *time*
- *student comfort*²³
- and *topic* (i.e. test version)²⁴

The section below initially discusses results of the analysis of other categories that test-takers commented upon. All the questionnaires had some responses, not all test-takers had responded to all questions. The overall RR was 85%. For each question analysed in this chapter the RRs are reported on the basis of answers to individual questions.

With regards to the research questions of this thesis the categories *time* and *topic* are discussed below. The category *student comfort* was also examined since it happened to be the second most frequently mentioned. The results from the analysis of the test-takers' overall bandscores are also presented.

²³ In the initial coding of the data from the post-test questionnaire, the category *student comfort* was used to present answers to two separate questions, that regarding the sound quality of tape and that regarding the testing room comfort.

²⁴ In this case, the terms *topic* and *test version* coincide.

As far as the entire testing experience was concerned, 86.7% (n=169; RR=97.5%) agreed that the test was good experience; 72.5% (n=142; RR=98%) of the test-takers considered that the test measured true English knowledge, and 71.4% (n=139; RR=97.5%) thought that the test was fair. In answer to the question what they liked about the test, test-takers' most frequent response with 21.4% (n=27; RR=63%) was that the test was good practice/challenge.

1.) Time

As far as the category *time* was concerned, 61.1% (n=119; RR=97.5%) of the test-takers indicated that time was insufficient for the test. In response to the question what test-takers did not like about the test, the category *time* also appeared with 22.6% (n=28; RR=62%).

It is important to point out that the category *time* got uniformly negative responses whereas other categories such as *reading*, *topic*, and *listening* elicited mixed responses. These appeared as categories both in test-takers' responses to the question of what they liked about the test and what they did not like about the test.

The results from the analysis of the post-test questionnaires regarding most frequently mentioned categories coincide with the results of the CAEL post-test questionnaires reported by Fox (2003). However, in Fox's (2003) study the second and third most frequently mentioned categories are topic and test fairness respectively whereas in this study the second and third most frequently mentioned categories are student comfort and topic respectively.

2.) Student Comfort

As already explained, in the initial coding of test-takers' responses to the post-test questionnaire the category *student comfort* stood for the questions referring to the sound quality of the tape used for the listening section, and the physical comfort of the testing room. 81.6 % (n=160; RR=98%) of the test-takers agreed that the sound quality of tape was good, and slightly fewer of them: 75.1% (n=145; RR=96.5%) said that the testing room was comfortable. The answers to both questions indicate that large number of the test-takers were content with the sound quality of the tape and the physical comfort of the testing room.

3.) Topic

Test-takers' responses with regard to the category *topic* did not seem to be as clear cut as those regarding time. 80.6% (n=154; RR=95.5%) of the test-takers said that the test topic was important to them; and 63.4% (n=123; RR=97%) indicated that test topic was interesting to them. *Topic* emerged as a category both in responses to the questions of what test-takers liked and did not like about the test. 18.3% (n=23; RR=63%) of the test-takers said that they liked the topic of the test while 7.3% (n=9; RR=62%) said that they did not like it. Overall the category *topic* seemed to be positively viewed by test-takers. It ranked as the second most frequently mentioned category in response to the question of what test-takers liked about the test and as the fourth most frequently mentioned category in response to the question what test-takers did not like about the test.

As far as previous knowledge about the test *topic* is concerned, 63.1 % (n=111; RR=88%) of the test-takers indicated that they did not have previous knowledge about it, and 57.5% (n=100; RR=87%) of the test-takers did not think that knowledge about the topic was helpful. For those test-takers who thought that knowledge about topic was helpful, it was mainly helpful for improving their understanding of the test topic, and content with 32.6% (n=15; RR=23%), and for writing a better essay with 26.1 % (n=12; RR=23%).

The most interesting trend as far as test topic was concerned was that although 57.5% (n=100) of the test-takers did not find knowledge about the test topic was helpful, 80.6% (n=154) of the test-takers said that topic was important to them. Therefore, it may be that the importance of topic was more related to whether test-takers found the test topic interesting (63.4%, n=123) than if they had previous knowledge about it.

4.) Overall Bandscores

The overall CAEL bandscores of the 200 test-takers who responded to the post-test questionnaire are presented in Table 4. Although drawn randomly across three CAEL versions the overall score distribution of this sample of test-takers is representative of the current CAEL test-taker population (Fox, 2004). The overall scores distribution of the current CAEL test-taker population differs from the one reported in the CAEL Assessment: Test Score and Users' Guide (Fox, 2002a) due

to the influx of test-takers with different English language learning background and language proficiency²⁵ (Fox, 2004).

Table 4: Overall test-takers' CAEL bandscores

Bandscores	Number of test-takers	Percent %
20.00	26	13.0 %
30.00	79	39.5 %
40.00	71	35.5 %
50.00	17	8.5 %
60.00	5	2.5 %
70.00	2	1.0 %
Total	200	100.0 %

4.1.2. Crosstabulation with Chi-square analyses:

1.) Time limits

There were significant relationships based on the chi-square analyses, of test-takers' perceptions of time limits on the test (i.e. "speededness"²⁶) and whether the test-takers perceived that the test:

- was fair
- topic was important
- topic was interesting
- reflected true English knowledge

²⁵ The current CAEL test-taker population consists mainly of Asian students whose English language learning background and preparation differs from the test-taker population used when the CAEL Assessment Test Score and Users' Guide (Fox, 2002a) was written.

²⁶ The amount of time allocated for completing a test or a test task is viewed to have a partial influence on test-takers' performance. The input that test-takers receive maybe perceived as speeded "if it is presented at a rate that calls test-takers' attention to speed as a factor in performance" (Bachman, 1990, p. 128). In fact, speededness is considered to belong to test-takers' perception of the importance of speed for the completion of test tasks. Therefore, in all probability, test-takers with low language abilities completing tests at even moderate speed would regard these tests as difficult. As a result, test-takers' perceptions of test difficulty and speededness would lead to guessing or compensatory test taking strategies (Bachman, 1990).

(see Table 5)

Table 5: Crosstabulation of test-takers' perceptions of time in relation to test fairness and topic.

Test-takers' perceptions	Value χ^2	df	Asymp. Sig. (2-sided) p
test was fair	10.32	1	.001
test topic was important	5.17	1	.023
topic was interesting	14.77	1	.000
test reflected true English knowledge	11.348	1	.001

Table 5 suggests that test-takers' perceptions of speededness are related to their perceptions of test fairness, topic importance, interest, and whether the test reflected their true knowledge of English.

The only test-takers' perception, which does not seem to be related to their perception of speededness, is their perception of the overall testing experience (i.e. test was good experience). Although the analysis did not reveal any significant differences ($p < .07$, $\chi^2 = 3.22$, $df = 1$) the relationship between these two perceptions suggests a meaningful trend (with $p < .07$ it approaches significance). However, as illustrated in Table 6, high proficiency test-takers (92%) tend to have slightly more favourable attitudes to the entire testing experience than low proficiency test-takers (82.9%). This finding is of particular importance as the response rate (RR) was 96%.

Table 6: Crosstabulation of test-takers' perceptions of having enough time and that of the overall testing experience

		Test was good experience		Total
		No	Yes	
Time was enough	No	20 (17.1%)	97 (82.9%)	117
	Yes	6 (8%)	69 (92%)	75
Total		26	166	192

These results show that test-takers' perceptions of speededness (i.e. whether they had enough time to take the test) is strongly related to their perceptions of test fairness, interest in test topic and the importance that they attribute to test topic. It can be speculated that insufficient time, from test-takers' point of view, could make a test topic seem less interesting and also less fair to them and therefore more important because they might experience interference with their performance. For the purposes of test face validity it is important to note that speededness does not seem to be related to test-takers' perceptions of the overall testing experience (i.e. test was good experience).

2.) Student Comfort

The category *student comfort* stood for the questions regarding the sound quality of the tape used for the listening section, and the physical comfort of the testing room. The analyses of test-takers' perceptions of whether the testing room was comfortable, and the sound quality of tape was good revealed significant relationships. Test-takers' perceptions of whether the testing room was comfortable appeared to be related to test-takers' perceptions of whether the:

- test was good experience

- test was fair
- test reflected true English knowledge
- time was enough to take the test

Test-takers' perceptions also seemed to be related to the test version they wrote. The results are summarized in Table 7.

Table 7: Chi-square analyses of test-takers' perceptions of room comfort and the test

Test-takers' perceptions	Value χ^2	df	Asymp. Sig. (2-sided) p
test was fair	6.384	2	.041
test was good experience	18.398	1	.000
test reflected true English knowledge	13.126	1	.000
time was enough to take the test	4.127	1	.042
test version	6.384	2	.041

Test-takers' perceptions of whether the sound quality of tape was good did not seem to be as strongly related to their perceptions of the test as it was with their perceptions of the comfort of the testing room. However, test-takers' perceptions of the sound quality of the tape appeared to affect their perceptions of whether:

- test was good experience
- test was fair

Test-takers' perceptions of sound quality were also affected by the test version that they wrote (see Table 8).

Table 8: Chi-square analyses of test-takers' perceptions of sound quality of tape and their perceptions of the test

Test-takers' perceptions	Value χ^2	df	Asymp. Sig. (2-sided) p
test was fair	5.788	1	.016
test was good experience	8.625	1	.003
test version	8.397	2	.015

There were not significant relationships between test-takers' perceptions of the sound quality of tape and whether they perceived that the test reflected their true knowledge of English ($p < .127$, $\chi^2 = 2.327$, $df = 1$), and that the time was enough to take the test ($p < .257$, $\chi^2 = 1.285$, $df = 1$).

3.) Test Version (i.e. Topic)

The analyses of test-takers' responses regarding test versions did not reveal any significant relationships between test versions and test-takers' perceptions. The results are summarized in Table 9.

Table 9: Chi-square analyses of test versions and test-takers' perceptions of the test

Test-takers' perceptions	Value χ^2	df	Asymp. Sig. (2-sided) p
test was fair	.204	2	.903
test was good experience	.090	2	.956
test reflected true English knowledge	.505	2	.777
time was enough to take the test	.501	2	.778
test topic was important	3.57	2	.167
topic was interesting	.462	2	.794
knowledge of test topic was helpful	.091	2	.956

However, the analysis of the test versions²⁷ and test-takers' overall bandscores revealed a meaningful trend (Table 10). The majority of test-takers writing version Cultural Geography (47, $n = 83$) scored in bandscores from 40 to 60, whereas the majority of test-takers writing versions Business (29; $n = 50$) and Environmental Policy (40, $n = 67$) scored in bandscores from 10 to 30²⁸.

²⁷ The titles/topics of the test versions are disciplinary, the real ones were changed for security reasons.

²⁸ Actual data analyses provided by the Testing Unit supports the level of proficiency defined across those groups. This conclusion is supported by evidence from a cloze test written by all of these test-takers (for details see Fox, 2001), and the Oral Language Test – administered as separate tests of the written CAEL that differences in these groups relate to proficiencies and not test versions.

Table 10: Crosstabulation of test version and test-takers' overall bandscores

Version	Bandscores			Total
	10-30	40-60	70-90	
1 Cultural Geography	36	47		83
2 Business	29	20	1	50
3 Environmental Policy	40	26	1	67
Total	105	93	2	200

This meaningful trend could be attributed to the overall proficiency of the group writing the test rather than to a particular test version. This finding seemed to be supported by the test-takers' perception that the test was fair. It is probable that if there was a relationship between test version and test-takers' test performance (overall bandscores), the low proficiency test-takers, especially, would have perceived the test as completely unfair. In addition to this, there is no significant relationship between test version and test-takers' overall bandscores ($p < .14$; $\chi^2 = 6.89$; $df = 4$). Therefore test version can be excluded as a category related to the test-takers' perceptions of the test.

4.) Test-takers' overall bandscores:

The analyses of test-takers' overall bandscores and their perceptions of the test revealed some very interesting results. No significant relationships and/or meaningful trends were established between test-takers' overall bandscores and their perceptions of whether the:

- test was good experience
- test reflected true English knowledge
- test topic was interesting
- time was enough to take the test

- knowledge of test topic was helpful

However, in some instances the cell count was less than five, indicating that the results must be considered only indicative, and would need to be re-investigated with a larger sample of test-takers.

However, the chi square analyses showed significant relationships between test-takers' overall bandscores and their perceptions of whether the test was fair with the first and the third approach to coding of the overall test bandscores. However, the results that need to be taken into consideration are those obtained with the third approach to coding. There was a cell count of less than five with the analyses using the first approach to coding (see Table 11).

Table 11: Crosstabulation of test-takers' overall bandscores and perceptions of test fairness

Approach to coding ²⁹	Value χ^2	df	Asymp. Sig. (2-sided) p
1: 10-30; 40-60; 70-90;	8.48	2	.014
2: 10-40; 50-90;	.003	1	.959
3: 10-30; 40-90;	8.11	1	.004

The fact that test-takers' overall bandscores are related to their perceptions of test fairness when the margin between low and high proficiency is set between bandscores 30 and 40 suggests that test-takers with lower proficiencies are more likely to perceive a test as unfair. This outcome happens to coincide with the CAEL minimum threshold level for engagement in academic activities set at bandscore 40 (Fox, 2004).

²⁹ For details on the approach to coding used see Chapter Three, page 91.

The chi-square analysis showed significant relationships between test-takers' bandscores and the perception that test topic was important when the second approach to coding was used (see Table 12).

Table 12: Crosstabulation of test-takers' overall bandscores and perceptions of topic importance

Approach to coding	Value χ^2	df	Asymp. Sig. (2-sided) p
1: 10-30; 40-60; 70-90;	2.01	2	.356
2: 10-40; 50-90;	9.72	1	.002
3: 10-30; 40-90;	1.02	1	.311

In this case, perceptions of topic (as opposed to fairness) suggest a threshold in test-takers' proficiency levels between bandscores 40 and 50. However it should be noted that for this analysis the cell count was less than five. Thus the analysis would need to be re-run with a larger sample. Nevertheless, the crosstabulation analysis (see Table 13) clearly showed that low proficiency test-takers' (with 83.9%) attribute more importance to test topic than high proficiency test-takers (with 56.5%) (when the second approach to coding was used). The response rate (RR) was 95.5%.

Table 13: Crosstabulation of overall bandscores and perceptions of topic importance

Overall bandscores	Topic was important		Total
	No	Yes	
low proficiency: 10-40	27 (16.1%)	141 (83.9%)	168
high proficiency: 50-90	10 (43.5%)	13 (56.5%)	23
Total	37	154	191

In conclusion, most of the test-takers did not consider that the provided test time was enough. In addition, the category time got uniformly negative

responses. As far as the category student comfort was concerned, test-takers' perceptions of the comfort of the testing room seemed to be related more strongly to the rest of their perceptions of the test than their perceptions of the sound quality of tape. The analyses of the other key category - test topic (i.e. version) showed that there was no relationship between the test-takers' perceptions of and performance on the test. The analyses of test-takers' overall bandscores and their perceptions of the test, showed that test-takers' proficiency seemed to be related to their perceptions of test fairness when the threshold of low and high proficiency was set between bandscores 30 and 40. Significant relationship between test-takers' overall bandscores and perceptions of topic importance was observed when the threshold of low and high proficiency was set between bandscores 40 and 50.

4.2. Phase Two: Semi-Structured Interviews and Journals

This section of the Chapter presents the results from the analysis of the semi-structured interviews conducted with the participants in the study: Wendy, Dave, and Chloe. The section begins with description of each participant, then the results from the analysis of the interviews before taking the test³⁰, the participants' journals (those of Wendy and Chloe), and the interviews after taking the test are presented. At the end of this chapter distinctive profiles of Wendy, Dave, and Chloe are presented as they emerged in the interviews.

Prior to presenting the descriptions of each participant, it important to note that on the test that they wrote for the purposes of this study all of them scored overall in the same bandscore. However, all of the three participants scored higher on the reading section.

4.2.1. Participant One: Wendy

Wendy was from Mongolia. She had studied English for three years. She started studying English in her home country by attending a three-month optional course in her high school. After graduating from high school in her home country, she attended for one academic year, a regular high school in the United States. While in the United States, Wendy took the following courses: mathematics, American literature, American history, algebra and planning. After graduating from the high school in the United States, she took a one-month TOEFL course,

³⁰ Although the participants were asked to talk about their expectations about the test they were going to write, they also referred to their previous experiences taking the test.

three times a week for three hours in her home country. At the time of the interview for this study she had been in Canada for two months from the tenth of January 2004. She arrived in Canada for the purposes of taking an ESL course at a Canadian university. At the time of this study she was enrolled in level 20 of the Intensive Program³¹. Wendy was planning to leave Canada after the end of the ESL course in April 2004. It was her hope that she would be accepted in the Bachelor Program in International Business at the university, starting Fall 2004. Her motivation for studying in a Canadian university was finding good employment in her home country after graduation. Wendy had written the paper-based version of the TOEFL test before writing the CAEL test. Her score was 395 out of 677. She also wrote the Test of Written English (TWE) but she did not remember her score. Prior to participating in the study, she took the CAEL test on the 31st of January 2004. Her bandscores on this CAEL test, along with the scores of the one written for the purposes of this research are presented in Table 14.

Table 14: Wendy – CAEL bandscores

CAEL Test	Reading section	Listening section	Writing section	Oral Language Test	Overall bandscore
31 st of January	30	30	30	40	30
Test for the study	40	30	20	40	30

³¹ The Intensive program (IC) at this university has 5 levels from 10 to 30. For more details see Chapter Three, page 179.

Her preferred way of studying English was listening to music, watching TV in English, and movies with subtitles. Wendy also thought that it was very important to learn new vocabulary and take notes in English classes.

4.2.2. Participant Two: Dave

Dave was from China, he had studied English for seven and a half years. He started taking English classes in middle school for three years, and then he continued through high school again for three years. At the time of the interview Dave had been in Canada for one and a half years. He came to study English in Canada in order to “quit Chinese in English classes”. He thought that it was difficult to study English in China because teachers used Chinese to teach it. Dave attended ESL classes in the Intensive ESL Program at the university for 4 academic terms. He attended levels 15, 20 (twice) and 25. Dave took level 20 twice as he thought that the method of instruction of the class teacher was very helpful for his English. At the time of the interview, he was not taking ESL classes, as he did not consider them helpful for writing and passing the CAEL test. In addition, he found trying to use only English in the ESL classes difficult because of the predominately Chinese student population there.

Apart from that, he hoped to be accepted to study International Business in the university although he had not applied. After graduating from high school in his home country he was accepted at university there. However, he thought that a degree from a Canadian university would give him better chances for professional success.

Dave took the paper-based version of the TOEFL test once in China two years ago. His score was 510 out of 667. He had also written the CAEL test five times in total (including the test on the 31st of January), that is three times prior to the interview, once for the purposes of this research and once again after that. His bandscores on the CAEL test on the 31st of January as well as on one written for the purposes of this research are shown in Table 15.

Table 15: Dave – CAEL bandscores

CAEL Test	Reading section	Listening section	Writing section	Oral Language Test	Overall bandscore
31 st of January	40	30	30	40	30
Test for the study	50	30	30	40	30

Dave thought that the best way to study English was to find a Canadian friend and spend time talking in English. In his opinion, the focus of formal English language instruction should depend on student's needs.

4.2.3. Participant Three: Chloe

Chloe was from Romania. She had studied English for 16 years in her home country. She started studying English in fifth grade in secondary school and continued through high school once or twice a week for two hours. She held a Bachelor degree in Physics and Chemistry with an emphasis in Physics from her home country. During her university studies, she took an English course for one term, when she was in her second or third year. While an undergraduate student, she was translating academic articles in Chemistry, Biology and Physics from her first language to English and vice versa to be published in international journals.

At the time of the interview she had been in Canada for almost three years. She arrived in Canada in May 2001. Chloe had not taken any ESL courses in Canada. Chloe's aim was to be accepted in a Master's Program in Physics or Chemistry.

At the time of the study she was contracted by the Immigration and Refugees Board of Canada on a call basis as an interpreter and translator of documents from her first language to English and vice versa. For that purpose she needed to take an English language test administered by the Immigration and Refugees Board of Canada. The test consisted of a grammar section, a listening and answering questions section, a speaking section (reading aloud), a written translation section (from English to her first language), and an interpreting section (from her first language to English and vice versa).

Chloe took the CAEL test on the 31st of January 2004. Her bandscores on this CAEL test, as well as on one written for the purposes of this research, are presented in Table 16.

Table 16: Chloe – CAEL bandscores

CAEL Test	Reading section	Listening section	Writing section	Oral Language Test	Overall bandscore
31 st of January	60	50	80	80	70
Test for the study	70	50	80	80	70

Her preferred means of studying English was on her own by referring to grammar books and reading different types of English texts, especially specialized ones.

4.2.4. The Interviews before the taking test

The results of the interview responses before the test are provided below:

Question #1 - What is the first thing that comes to your mind when the CAEL test is mentioned?

The three participants: Wendy, Dave, and Chloe had different associations about the CAEL test depending on their image and knowledge of the test. Wendy said it was like the TOEFL test and that students who want to study in Canadian universities take it. Dave's and Chloe's answers were more related to the content of the test itself. Dave associated the test with the topics and the writing section, which in his opinion is crucial for the final test score. Chloe added that the CAEL is a test that is not easy to pass. She emphasized that the listening section was the most difficult one because it required listening, reading and answering to questions simultaneously³². In her opinion the speaking section³³ was the easiest one.

Question #2 - Do you think that the test would be difficult?

Wendy, Dave, and Chloe said that the test was difficult. Wendy clarified that for her it was more confusing than difficult. She explained that she related confusing to the need to invest a lot of effort in completion of a task. Dave noted that given his familiarity with the test it did not seem to him as difficult as it did

³² The three participants made similar comments about the difficulty of the listening section.

³³ The part of the test addressed in this part of the thesis as speaking section is the Oral Language Test.

initially. Chloe identified the test difficulty with its time limit, which made the test:

- more interesting as it required quick reactions and concentration
- but also made her feel nervous and pressured and subsequently interfered with her productivity and ability to think clearly.

Question #3 - What do you think about the test?

The three participants agreed that the test was good experience and that they generally liked it. Wendy specified that it gave good opportunity for students to enter Canadian universities. Dave said that the CAEL test was better than the TOEFL test explaining that the CAEL test showed test-takers' proficiency in all skills: reading, listening, speaking, and writing, whereas the TOEFL offered more limited assessment of test-takers' knowledge of English. He added that the CAEL test gave test-takers the opportunity to demonstrate their English language knowledge in different ways whereas with the TOEFL there was just one option. In Dave's opinion the TOEFL test was harder vocabulary-wise whereas the CAEL was harder as it involved the four skills: reading, listening, speaking, and writing. He said that the CAEL test was more pleasant as it was shorter compared to the TOEFL test. Chloe added that despite the issue with the time limit the test was well-structured and organized.

Question #4 - What is your opinion of the test according to what you have heard about it?

The three participants had heard that the test was difficult/hard. Wendy said that the time limit was pointed as the main reason for that. Dave said that he was warned about the listening section because of its speed. Both Dave and Chloe said that they did not think the test was as difficult as they had been told. Dave attributed this to his experience with the test and Chloe to her high English language proficiency.

Question #5 - Do you feel nervous or anxious about the test?

Wendy said that she was nervous about taking the test because of the big number of students who were trying to enter university, and the short time, in her opinion, allotted for the listening section. Both Dave and Chloe said that they were not very nervous. Dave once again attributed it to his familiarity with the test, although he admitted that he felt anxious about passing the CAEL. Chloe expressed her concern that because of the time limit she would not be able to show her real abilities especially when the listening, and the writing sections were concerned. She also referred to her past experience taking the test saying that during the test she did not feel nervous but rather pressured to complete the test on time.

Question #6 - Would you be less or more motivated to study English if you did not have to take the test?

The three participants indicated that that they would be motivated to study English even if they did not need to take the CAEL test. Obviously, for Wendy passing the CAEL test, taking ESL and entering Canadian university are in a way intertwined, and she was not able to tell them apart. Dave said that he needed to know how to write an academic essay and how to deal with other academic tasks in English before starting university studies without English language support. However, he admitted that if given a chance he would try to take university courses and see if he could do it without taking ESL. He added that if he had problems he would seek help with his English. Chloe explained that she loved English and she used to study by herself in the past, and was still doing it. In her opinion, there were always ways to improve your knowledge.

Questions #7 and #8 - What are your strengths in English? What are your weaknesses in English?

Wendy said that her strongest skill both in general English, and on the CAEL was speaking, followed by listening, reading, and writing. She explained that speaking, and listening were easy for her as she had a lot of exposure and had not had serious problems with them. As for her reading, and writing skills Wendy said that she was not at ease with them due to limited chances to practice them.

Dave's strongest skill both in general English, and on the CAEL was reading followed by speaking, listening, and writing. He said that reading was

traditionally a strong skill for Chinese students whereas writing was traditionally a weak one. The reason for that Chinese students just transferred their thoughts from Chinese into English, using the “Chinese way of thinking”.

Chloe was the only one to distinguish between her skills in general English, and on the CAEL. She noted that she was comfortable with all skills. In general English she was best at reading and listening followed by writing, and speaking. She identified her speaking as the weakest because sometimes she needed to look for her words while speaking. She emphasized that she would not classify any of her English language skills as weak. On the CAEL her strongest skill was listening, followed by reading, speaking, and writing. She also pointed out that she would have liked to have a bit more time for the essay in order to write one at academic level. In Chloe’s understanding an academic essay was difficult to write and involved using high register, and therefore required more time.

Question #9 - What do you feel most comfortable doing in English?

(speaking, listening, writing, reading)

Both Wendy and Dave indicated that the skill they felt most comfortable with in English was speaking as they did it very often. For Chloe it was reading, as she loved reading all types of texts in English. She explained that as far as the CAEL readings were concerned more time would allow her the chance to read the texts more carefully and make sure that her answers were correct. She also said that when reading specialized texts in chemistry or biology sometimes she would

read some passages more than once in order to make sure that she understood everything.

**Question #10 - What do you feel least comfortable doing in English?
(speaking, listening, writing, reading)**

The three participants were the least comfortable writing in English because of its high demands in terms of grammar, vocabulary, and structure. For Dave writing was also the most important part of the test due to its weight for the final test score.

Question #11 - What do you expect to be most difficult part for you on test?

Initially, Wendy indicated that the entire CAEL was difficult, and then she specified that the reading section would be most difficult for her on the test because of her limited vocabulary followed immediately by the writing section due to grammatical problems and difficulties with the structure of an academic essay. Dave said that the writing section, which was also his weakest skill, would be the most difficult for him on the test. He added that the writing section showed students' thinking and understanding of the topic of the essay, and thus student's overall "value". He also explained that he viewed the listening section second in difficulty because he had fewer opportunities to practice it. Chloe explained that the most difficult section for her would be the listening. She explained that she was not comfortable with it, that she did not like it at all and that she felt very

pressured because of the time limit and the need to perform a lot of tasks simultaneously (i.e. to multitask). Chloe did not agree that students were asked to perform the same tasks in a regular university lecture. She was also curious to see how much a professor would score on this part. She said that sometimes wondering about the correct spelling of a word added to the difficulty of this section of the test. However, later on in the interview Chloe said that the writing section was the least difficult for her (not the easiest) because she had the freedom to put her thoughts and ideas about what she read and listened to in her own words without being obliged to use the vocabulary from the test. The only problem for her with the section was the time limit. She noted that an extra 15 to 20 minutes would have allowed her to better organize her ideas.

Question #12 - What do you expect to be easiest part for you on the test?

Wendy thought that the easiest part for her on the test would be the listening section especially compared to the writing one as it was easy for her to understand. Both Dave and Chloe identified the speaking section as the easiest. Dave considered the speaking section easy as it did not have much weight for the final test score and because it was easy for him to perform the necessary tasks. Although Chloe identified the speaking section as the easiest, she pointed out that she was not pleased with it at all due to the short time allotted for it, and the fact that she was able to hear the other test-takers' answering the questions. She said

that the combination of limited time left for answering the questions and the background noise made her very distracted.

Question #13 - Do you think that your test result will show your true abilities in English?

Both Wendy and Dave agreed that the test would reflect their true abilities in English to a certain extent. While Dave was convinced that the CAEL test evaluated all test-takers' English language skills, Wendy expressed such hope but was not completely sure. Chloe on her part was not completely convinced that the test would reflect her true abilities in English due to a number of factors. The ones that she pointed out were the limited time allotted for some sections of the test, and the background noise during the Oral Language Test. She also explained that it was not possible to have perfect English assessment as it was difficult to evaluate test-takers' language proficiency in two or three hours. She suggested that testing over a longer period of time might be more suitable. In Chloe's point of view the above-mentioned factors are especially crucial for test-takers' with upper-intermediate/ advanced English language proficiency and not as much for test-takers with low or very high English language proficiency.

Question #14 - Do you feel prepared for the test?

Wendy was the only one of the three participants to say that she did not feel completely prepared to take the test. However, if given a chance to improve

her English before taking the test she would still have preferred to write the test and then work on her English.

Dave and Chloe, on the other hand, that they felt prepared to take the test. Dave counted on his self-preparation using the books about CAEL from China, which showed how to write good and “hard” sentences. Chloe attributed her readiness to take the test on her long history of studying English and translations that she did. She also pointed out the CAEL preparation workshop as a significant part of her preparation for the test. Chloe commented upon the online sample test, she was the only one of the participants to take it. In her opinion that test was very simple compared to the real one and did not give a clear idea about the true difficulty and time limits of the actual test.

4.2.4.1. Emerging themes:

Test topics

The issue of test topics emerged as predominant especially in the interviews with Dave and Chloe. For Wendy test topics were not of significant importance. Dave was the participant who was most concerned about the test topics. In his opinion knowledge about the test topics was helpful for passing the test. Because of this he went to the library to read books about those of the CAEL test topics which were known to him. Another source of information about test topics was classmates and friends of his who had written the test. For Dave an easy test topic was one that he had previous knowledge about and had done some practice on, and difficult was one that he did not have previous knowledge about

and had not practiced. Familiarity with the test topics and information about them made Dave feel more comfortable during the test and less nervous about it. For him knowledge about test topic was more important than the fact if he liked the topic or not. Liking the test topic was secondary of importance to him. When asked what else he could do in order to perform better on the test, Dave said that this would be having more information about test topics.

Chloe, on her part, said that she liked the fact that the test was thematically organized into topics. She explained that her performance might depend to a certain extent on the test topics because for her it was easy to understand if she was interested in and liked the topic. In her opinion, being interested in a topic meant paying more attention while learning about it.

Time limits

The participants felt that the allotted time for the listening and writing sections was not enough. Chloe also said that the time allotted for the Oral Language Test was very limited.

Listening section

The three participants mentioned that the listening section was difficult because they needed to combine listening, reading, and writing.

*Test-takers' individual characteristics*³⁴

Depending on their individual characteristics test-takers' perceptions of what is interesting, easy, and difficult might differ. Wendy pointed out a difference between difficult and confusing tasks. For her confusing was one requiring a lot of effort. For Dave the notions of difficulty, importance, his weaknesses, and most weight for the final test score were intertwined. It is important to note that his personal preferences (likes) did not matter much in attributing importance to test sections. For him being comfortable was related to the frequency of exposure: the more frequently he practiced a skill the more comfortable he felt about it on the test.

For Chloe the more difficult section on the test was the more interesting one. Another interesting characteristics about her was the fact that when working under significant pressure (such as time limits) she felt pressured and "blocked" regardless of the relative ease/difficulty of the task she was performing. She also viewed as a positive characteristic of the CAEL test the fact that it was a "tough" test on a high academic level.

³⁴ The term individual characteristics distinguishes between "high road" and "low road" students as described in Bereiter and Scardamalia, 1983. Basically, "high road" students always try to give their best performance, whereas "low road" students simply strive to get the job done with minimum effort.

4.2.4.2. Recommendations:

Chloe was the participant who gave a lot of recommendations for improving the test. She recommended including a section with grammatical items and a section with vocabulary exercises such as synonyms. As far as the listening section was concerned, Chloe suggested some restructuring it so that test-takers first listen to the tape, take their own free notes, and then answer the questions.

Chloe made a number of recommendations directed at the administrative aspect of the test such as providing the test-takers with safe place where they can leave their valuable belongings, making sure that the CAEL test scores are accepted in bigger number of Canadian universities, and making sure that test-takers cannot hear each other during the Oral Language Test. She also said that she found acceptable the price of the test.

4.2.5. The Journals:

The results of the journals are provided below:

Wendy's Journal

Wendy felt nervous about the test due to the big number of students wanting to enter university. She saw the test as good opportunity to do that herself. She was afraid of the questions that she would be unable to answer. It was apparent that Wendy realized that her English was not good enough and needed improvement. She identified the reading as hard because she had problems understanding the readings and the Oral Language Test as complicated as she

could hear the other test-takers talking. She expressed a hope to be lucky and to get a good score.

Wendy also wrote a journal entry after taking the test. For her the test topic was interesting but she was not familiar with it, she was comfortable with the entire test-taking experience. Wendy identified the listening section as the most difficult. In her opinion she completed the writing section well.

Wendy was a bit scared, impatient and anxious about her test score because she perceived it as very important for her future life. She wrote that she hoped that her dream to be accepted in a certain Canadian university would come true.

Dave's Journal

Dave did not write a journal. In an email he explained that when thinking about the CAEL he felt nervous and fidgety. Overall he did not feel nervous or excited about the test.

After writing the test he said that the test was good practice.

Chloe's Journal

Chloe was the only participant who mentioned her strategies to fight her nervousness about the test. She found that thinking about her previous experience with English (translating texts), talking to friends, watching TV, and exercising made her feel less nervous. She also noted that the sample CAEL test available online was not of big help as it was very easy compared to the real one. On the contrary, the CAEL preparation workshop that she took was very helpful

especially in terms of providing her with reading strategies. Chloe also mentioned that she needed to prepare for the test. She even prepared in her mind answers for the Oral Language Test.

The main things on the test that she was worried about were the time limit and more specifically getting a headache if she felt pressed for time. The test topic and the listening section were the other sources of her worries. She also found herself thinking about her previous testing experience.

Chloe wrote a journal entry after taking the test. She was not pleased with her overall performance as she did not manage her time better and because she was not as motivated as she was on the first time she wrote the test. She wished that she had more time especially because it took her some time to understand the test questions. For her the test topic was not difficult, she said that she was interested in it and enjoyed it and tried to understand it. The difficult section on the test was the listening one and also figuring out how to extract the most important ideas of the readings and use them in her essay. Chloe also mentioned being disturbed by noises from the other test-takers.

She also wished to be able to see the right answers. In her opinion talking about her performance on the test helped her realize her weaknesses in English and the need to work on them.

4.2.6. The Interviews after the taking test

The results of the interview responses after the test are provided below:

Question #1 - Was the test difficult?

The three participants agreed that the test was difficult. In fact, Wendy identified a difference between test difficulty and topic difficulty. For her the topic was difficult. She mentioned that if she knew something about the test topic it would have been easier for her to write the test. Wendy also said that the listening section was very difficult because of its speededness. She said that she understood what was said but she could not catch the answers to the questions. She implied that this happened because she had to listen to the tape, read and answer questions in the same time.

In Dave's opinion the test was not that difficult, he mentioned that he had some general knowledge about the topic, which helped him a bit. He also said that if he had more time he could have done better on the reading and the writing sections.

Chloe said that the test was difficult. For her the listening section was the most difficult. She said that she was expecting that she would have lowest scores on this part of the test. She attributed the difficulty of this section to the fact that she needed to listen to the tape, read and answer the questions simultaneously. The also noted that having more time for the writing and especially the reading sections would have helped her perform better. Chloe explained that some of the questions in the reading section required some time to be answered.

Question #2 - Was the test what you thought? How? And how not?

None of the participants agreed that the test was what they expected it to be. Wendy was surprised by the test topic, which she regarded as difficult. For Dave the different in this testing experience was the fact that there were only two test-takers taking the test. As a result he could not speak quietly while reading because he was bothering the other test-taker. He thought that if he could speak quietly he had better chances of getting a higher score. For Chloe it was her attitude towards the test this time that was different. She was very relaxed and laid-back which reduced her perception of her productivity.

Question #3 - Do you think that your result will reflect your true knowledge of English?

Only Wendy agreed that the test score would show her true knowledge of English. She explained that the CAEL test showed test-takers true knowledge because test-takers are writing by themselves unlike the TOEFL test where there was the possibility of guessing. However, Dave and Wendy did not agree that the test score would show their true knowledge of English. Dave said that he could not speak quietly while reading and that he was also bothered by the sunlight in the testing room. It was his beliefs that if he was able to speak quietly, was not bothered by the sunlight in the testing room, and had more time, he would have performed better on the reading section. He also mentioned that if he had more knowledge about the test topic he would have scored more on the writing section, and if he was able to spell some of the words from the lecture he would gotten

higher score on the listening section. However, Dave said that he liked the reading section best. Chloe felt that she was too relaxed and not sufficiently worried about test scores. She liked the writing section best, and thought that it would reflect her true knowledge of English.

Question #4 - Did you feel nervous or anxious during the test?

None of the participants felt particularly nervous during the test. Wendy admitted that she was wondering about the test topic prior to the test. Dave explained that he was not completely comfortable but definitely did not feel nervous. He noted that he could not feel nervous as he was taking the test for fourth time and besides that he had prepared well for it. Chloe, as already mentioned, was very relaxed and laid-back.

Question #5 - Did you feel more nervous before writing the test than during the test itself?

Dave admitted that the only time when he felt uneasy was when he was not able to find an answer to a question in the readings quickly. Chloe, on her part, said that she became a bit anxious during the test when she realized that she was almost running out of time. She said that this happened because she was too relaxed initially. She also explained that she was disturbed by the noises made by the other test-taker and the fact that he seemed to be ahead of her on the test.

Question #6 - What was the most difficult part for you on test? Is this what you are weak at in English?

The three participants said that the most difficult part was the listening section of the test. Wendy said she understood what was said in the lecture but that she could not catch the answers to the questions because the speaker was speaking too fast. Dave became really angry, as he could not spell some of the words that he heard and whose meaning he knew. He also mentioned that his vocabulary was poor. Chloe said that it was too much for her to listen to the tape, read and answer questions. She added that sometimes the lecture would go off topic for a significant period of time which was confusing because she was not sure if she lost track of the questions. As a result she missed a couple of them. Wendy and Dave said that their weakest skills in general English were other different from those on the CAEL. For Wendy it was her writing skills although initially she said that everything in English was difficult. Moreover, she thought that this was normal as English was not her first language. For Dave writing was also his weakest skill in general English.

Question #7 - What was the easiest part for you on test? Is this what you are best at in English?

Both Wendy and Chloe pointed out the writing section as the easiest one on the test. Wendy said that it was easy for her as she had topics to write about and also she could refer to the readings, look at words in them and then change some of the words. Chloe saw the writing section as the easiest one on the test as she

had the opportunity to express her thoughts and ideas freely. For Dave the reading section was easiest. He identified reading as his strongest skill in English. He also pointed out that he had done a lot of practice on it.

Both Wendy and Chloe identified listening was the easiest skill for them in English. Chloe especially saw it as requiring minimal effort compared to the other ones, and also was giving her least trouble. She explained that when she read she needed to go over some sentences again, that when she spoke it happened that she could not think of some words, and that when she wrote she had occasional spelling problems. For Dave speaking was the easiest skill in English, as he did it daily and it was something common for him.

For all the participants when their skills in general English were concerned, the respective skill that they identified as easiest was particularly easy when performed in informal settings.

Question #8 - Where do you expect to get highest scores?

Wendy expected to get highest scores on the reading section. She explained that the writing one was easiest for her but she still had problems with grammar and structure. Dave thought that he would get highest scores on the reading and the writing sections. He said that he expected high score on the reading part because he had general knowledge of the topic, and he was strong at it anyway. As for the writing section, he said that in addition to having a general knowledge of the topic, he tried to write a "good essay". For Chloe, the writing

section was the one where she expected to score most. In addition to being able to express herself freely she said that she liked the writing part best.

Question #9 - Where do you expect to get lowest scores?

The three participants said that they expected to get lowest scores on the listening section of the test. Dave made an interesting comment saying that for him the most important part of the test was the writing section, because it mattered the most for the final test score. For him this section was “the truth”.

Question #10 - What recommendations do you have for improving the test?

Wendy and Dave did not have any recommendations for improving the test in terms of test development. She said that the test was ok, and the time allotted for it was enough. Wendy mentioned that test-takers needed to study hard, do grammar practice, and listen to radio and TV in English as well as to native speakers talking. In her opinion all that was very important for test-takers' future. Dave said that writing an essay every day was the best thing to do. He added that it was not easy to do that kind of practice, as he did not have access to listening and reading texts similar to those on the CAEL.

Chloe said that she had nothing new to add to the recommendations about the test that she gave in the previous interview. She emphasized once again that the structure of the listening section should be changed.

4.2.6.1. Additional comments

Test fairness

The three participants agreed that the test was fair to a certain extent.

Wendy said that she guessed and hoped it was so but she was not sure. Dave was curious why every time he took the test only about 20 or 30 students out of 100 passed the CAEL³⁵. Chloe expressed doubt that the listening part was following the structure of a usual university lecture. She was not convinced that a lecturer would lecture and ask students to answer questions in the same time. She also said that a little bit more time would have been better.

Overall evaluation of the test

Wendy repeated that the test was doing a great job. She said that the topic was uninteresting, the test was difficult, and she felt fine although a bit nervous. Dave said many times that the test was a good evaluation of all test-takers' English language skills. For him the topic was not interesting, the test was difficult, not fun, and challenging. Most importantly, in his opinion that test that he wrote gave him good practice on test topics. Chloe said she generally trusted the test and was happy with it, she was said that it was a "tough" test. In her opinion, the test was difficult, challenging, fun, good experience, a great challenge, and the topic was interesting.

³⁵ In his opinion, every test-taker who scored bandscore 40 and above had passed the test.

Could have been done better on the test

Both Dave and Chloe agreed that that they could have performed better on the test if certain circumstances were different. Dave said that he could have gotten higher scores on the reading and writing sections if he had more time, better vocabulary, had knowledge about the test topic, and was able to speak quietly while doing the reading section. In Chloe's opinion it was always possible to do better.

If the CAEL test reflects tasks normally performed in everyday university setting

Although none of the participants attended regular university classes at a Canadian university they somehow agreed that the test reflected tasks normally performed in everyday university setting. Chloe's only doubt was that during real university lectures, lecturers talk and ask students questions in one and the same time.

Was the test more difficult than the one of the 31st of January?

For Wendy the topic of the test for the purposes of this study was more difficult. She said that the reason was understanding the idea expressed in the reading, and the listening sections not the vocabulary used. She also noted that this time the listening section was harder whereas the reading was easier.

Dave explained that each test was different because of the reading section. He also emphasized that the difference did not resided in test topic or vocabulary

used but in the fact that he needed to express his own opinion when answering some of the questions. For this reason the test on the 31st of January was more difficult for him than this one. Dave also found the listening section of the test for the purposes of this study more difficult than the one of the test on 31st of January.

For Chloe the CAEL test on the 31st of January was harder than the one she wrote for the purposes of this study. The reason resided in the way she felt about the entire testing experience. On the test on 31st of January she was very nervous and stressed about the final test score, this made her perceive the test as more difficult. She also added that she was more relaxed the second time she wrote the CAEL test, as she was already familiar with it.

Instructions on CD

The instructions on the test for the purposes of this study were played on a CD instead being spoken by test administrator as it normally happened. The three participants said they preferred the instructions to be played on CD. Wendy said that they were clearer, Dave noted that it helped him adapt quickly to the voice of the speaker for the listening section, and Chloe said that it helped her pay more attention to the lecture and the instructions.

Test preparation

Dave and Chloe had done some test preparation. Dave used the CAEL preparation books from China, and also practiced writing “good” and “hard” sentences and essays. Chloe completed the sample CAEL test available online and

also considered attending the CAEL workshop as a significant part of her preparation. Wendy did not anything else but attend her ESL classes.

Test-taking behaviour and strategies

The participants followed the instructions provided in the test. They first read the questions and then the texts in the reading section. They all returned to unanswered questions in the time they had left in the writing section. Wendy did not circle or underline anything in the reading texts and/or questions. Both Dave and Chloe underlined/circled keywords (such as numbers-Chloe). Dave also marked unanswered questions. Chloe admitted to devoting too much time to the first series of questions in the first reading. She also tried to look for answers to two or three of questions at a time. She also said that she tried to read questions in advance in the listening section. Chloe also said that that she found helpful when writing the essay to browse through the texts in the reading section and the questions in the listening section at the beginning of the writing section. She also explained that she went back to answer omitted questions in the listening section while still the lecture was playing as she realized some of the answers. Chloe added that although there was not much time to think about the essay while doing the other sections of the test, it was a very good idea that the test topic was provided at the beginning of the test.

4.2.6.2. Emerging themes:

What was difficult?

Wendy made a difference between test difficulty, and topic difficulty. She attributed topic difficulty to her difficulty in understanding the idea expressed and not so much to vocabulary or familiarity with the topic. Dave made a difference between topic difficulty and reading difficulty. For him reading difficulty consisted not so much in having unfamiliar vocabulary but in the fact that he had to answer questions requiring him to express his own opinion and/or to put author's opinion in his own words.

Test topics

Wendy mentioned that the topic of the test that she wrote for the purposes of this study surprised her and in a way changed her perception of the entire testing experience. Chloe once again mentioned that the test topic did not matter much for her. She explained that liking the test topic to some extent helps to understand it but previous knowledge about the test topic did not make a test easier. For Dave previous knowledge about the test topic has a direct relation to the final test scores.

Time limits

Dave and Chloe were the participants who mentioned on many occasions in the interviews that if they had more time on the reading and writing sections they would have performed better.

4.2.7. Outcomes: Distinctive Participants' Profiles

The section of the chapter presents distinctive profiles of each participant as they emerged in the interviews. The section finishes with similar comments made by the three participants.

As it can be seen from the participants' descriptions earlier in this chapter, they have different levels of English language proficiency, educational background, and reasons for taking the CAEL test. Although they all were CAEL test-takers because of their proficiency levels and reasons for taking the test, they should not be considered as belonging to one and the same category of test-takers. This is to say that the participants' perceptions and feedback of the test seemed to be to a large extent shaped by their proficiency levels, reasons for taking the test, and individual characteristics. They provided insight into the CAEL test from different perspectives showing what test-takers' with different levels of English language proficiency think, experience, like and do not like about the CAEL test.

In fact, Wendy's, Dave's, and Chloe's different levels of English language proficiency and individual characteristics seemed to lead them to have different awareness and knowledge about the CAEL test as well as different test-taking behaviour and strategies.

4.2.7.1. Wendy

Wendy had the lowest level proficiency of the three participants. In her opinion the entire test was hard and confusing. Initially, it was difficult for her to identify a section of test that was of a particular difficulty. She also had difficulty

providing detailed explanations in support of her opinions and ideas. This was surprising given the fact that initially she seemed to be a very fluent communicator. However, later on in the interviews it became obvious that her listening comprehension was not that fluent. Sometimes she would misunderstand a question and/or contradict herself. For this reason, her interviews did not provide as much data as the interviews with the other two participants.

Both interviews showed that Wendy saw the test as a means of being accepted into a certain Canadian university and consequently of having a successful future:

P: I will do my best because it's important.

R: Why is it so important?

P: Because I have, it's important for my future life.

However Wendy did not have a clear plan how to achieve her goal, in a way she was not aware of what her weaknesses in English were. Although she was attending an ESL class and she definitely realized that her English needed improvement she would rather work on her English *after* taking the test. Wendy viewed passing the CAEL test as a matter of luck, it was her hope, so that her "dream" to study in a certain Canadian university would come true. This participant lacked enough awareness of her English language proficiency to such an extent that she could not even take advantage of test-taking strategies. She was ready to do anything in order to improve her English, but *after* passing the CAEL test. In a way she viewed the entire situation as a deal: she passes the test and then she would continue studying English. This attitude of hers might be partially due

to the fact that in her mind English was difficult anyway because it was not her first language:

P: Everything is difficult, I mean, you know. It's // and what can I say.

R: Everything is difficult in English or everything is difficult on the CAEL?

P: Everything is difficult for me in English. Not only that CAEL test.

R: Can you explain?

P: // Of course we // our language // our English language is second, you know, second language. //

R: yeah so what you are trying to say is that because English is not your first language // whatever you do in English

P: Of course it will be hard. If you are in my, you know. // If you come to my country, you know, you can't speak in English, it just surprise you and you have to learn my language. Of course it will be hard, really, really, hard.

It was also difficult for Wendy to expand her awareness about the CAEL test and realize its distinctive characteristics, she though it was the same as TOEFL. Thus Wendy was not able to evaluate realistically her English language skills, and to take a course of action that would eventually help her achieve her goal. For the time being she was "stuck" in a stage where she suffered from her own unawareness caused by her low language proficiency. It was encouraging that during the interview, after writing the test, Wendy showed slightly increased awareness about the CAEL test and her English language skills. For instance, she was able to see a difference between the format of the TOEFL test and that of the CAEL test.

4.2.7.2. Dave

The three participants, especially Wendy and Dave did not have clearly internal motives for studying English except for passing the CAEL test. Although Dave and Wendy scored 30 on the CAEL test, Dave had higher scores on the reading section. Most importantly, the main difference between the two of them was his awareness of his English language proficiency, the nature of the CAEL test, and the way he approached the entire test-taking experience. What they had in common was that the CAEL test was in a way an “obstacle” or rather “means” for them to be admitted into a certain Canadian university. Although the CAEL test held seemingly equal stakes for both of them, their perception of it differed. Unlike Wendy, Dave had a more “positive” attitude to English. In everyday English nothing was particularly difficult for him.

His interviews provided rich data regarding test-taking, and what was easy/difficult. From test-takers point of view. He was an avid test-taker. This participant knew which section was the most important (i.e. had greater weight for the final test score), he read about known test topics, practiced writing “good and hard essays and sentences” and believed that if he:

- had more background knowledge on the test topic
- had more time for the reading and writing sections
- could speak quietly while doing the reading section on the

test for the purposes of this study he would have scored better.

Dave was the “pro” in CAEL in terms of test-taking strategies. He was the strategic test-taker. He was able to identify his weaknesses on the CAEL test and

in general English and work on them. It is another issue whether he was doing this in the best way. He was trying to compensate and/or at times substitute missing knowledge in English with test-taking strategies instead of regarding them as a facilitating device. He provided a definition of what passing the CAEL test was. In his opinion, every test-taker who scored bandscore 40 and above had passed the test. Dave also produced a definition of what a good academic essay was:

P: // I mean you should use the hard word, hard vocabulary

R: Hard vocabulary, what is this? Can you explain?

P: // Not too simple

R: So, long words, academic words?

P: Yeah, academic words. Yeah, not simple words

Dave was in a very interesting stage in building his awareness about his English language proficiency. He was able to see the advantages of the CAEL tests in terms that it measured all test-takers' skills, and gave test-takers the chance to express their knowledge in many ways unlike the TOEFL test:

P: Yeah// if // how to say that? /// just the TOEFL test is just like a straight line, and maybe the CAEL test like. We have // we go from A to B

R: Yeah

P: Maybe TOEFL test is just one way.

R: Ok

P: And the CAEL test shows many way we can go to B. Maybe // we have three ways to go to B.

Although Dave realized the need:

- to be able to write a good academic essay (different from a personal one)
- to be able to learn in English in order to be successful in his university studies

he admitted if given a chance he would:

- write a personal essay if this would help him enter a university
- try to take university courses without ESL support.

Dave was on his way in understanding more clearly what it took to study in Canadian university but he was not there yet. As he himself said he would write a personal essay but he did not deceive himself that this would help him succeed in university studies. This participant also realized that:

- his essay writing had improved as a result of attending an ESL course:

P: Did not they tell you about the way you should organize your essay?

P: Actually // sometimes I tell my friend how to organize their essay

R: Oh, you tell them?

P: Yeah // because I get // level 25 and my teacher already taught me how to write an essay. I tell them how to write the essay, how to organize the //

R: Ok, but basically why did you tell them how to organize it? Did they ask you or ?

P: Because you know also my friend, you know, when they take the CAEL test they use the // personal like I, they, our, but they cannot use this word, they will not pass. You cannot write personal essay. You

know, the first time I just like, I write down an essay personally. (laugh). So I get really poor score.

- and that his listening skills had improved with his stay in

Canada:

P: It's much better now before then I first get to come to Canada. I think it has been a long time, time is passing and your listening skill is much better.

Dave was simply caught in a dichotomy between what he knew was helpful for improving his English and his immediate pragmatic needs and goals, i.e. passing the CAEL test and being accepted into undergraduate university program. He thought that improving his English by taking ESL and only then writing the CAEL was the long way:

P: I just imagine maybe 10 years later, I get really good in writing. (laugh) And I think back about 10 years before, it's terrible. Maybe I will keep some writing from these years, and 10 years later I will see my essay wow it's childish, so childish. (laugh)

Instead Dave had decided to try the shortcut and "study for the CAEL".

For him, knowing about the test topic was much more important than if he liked it or not. Although he disliked writing academic essays and it was difficult for him this was the most important part of his test preparation. However, he cannot be entirely classified as the typical Chinese test-taker only trying to "crack the code of the CAEL". Dave was subjected to the same pressures (i.e. time, money, family) but as he himself said he knew

- that learning essays by heart was not helpful
- and where to stop and go back to studying English.

Using Smith's (2003) grounded theory³⁶ which explains the stages through which students in the Carleton University Intensive ESL program go through and/or get stuck, Dave was on the "Yellow Brick Road" midway through the stage called "Shooting Yourself in the Foot". According to Smith (2003) students at this stage are aware, to different extents, of the adjustments that they need to make in order to be successful in their university studies in Canada, but "they can 'talk the talk' but not yet 'walk the talk' (p. 42). Dave was most definitely 'talking the talk' and even making small steps into his attempt to 'walk the talk'.

4.2.7.3. Chloe

Chloe was the participant whom I would call "recommendations available" as she provided a lot of recommendations for improving the test and the overall testing experience. Wendy and Dave seemed more focused on passing the CAEL test and entering a certain Canadian university whereas Chloe operated from a broader perspective which was probably partially due to her significantly higher English language proficiency. For her the test also was a means of entering a university, but at the same time it also was a challenge, a way to prove herself. For instance, she perceives test difficulty and test time limits (to a certain extent) as interesting, challenging and fun. What was interesting about Chloe was that although she did some test preparation, had general knowledge about the structure

³⁶ Smith (2003) uses "tongue-in cheek" approach (p. 38) to describe the stages through which students in the Carleton University Intensive ESL program go through and/or get stuck. She uses images from the famous children story *The Wizard of Oz* (1990) by Baum cited in Smith (2003). The main character Dorothy travels down the Yellow Brick Road, far away from home, passing from stages of total confusion and helplessness to understanding, acceptance, and control.

of the test, and admitted that the CAEL workshop provided her with strategies for each section of the test, she did not seem interested in “cracking its code” or learning in details about it (e.g. test topics). Moreover, Chloe was generally confident about her English language proficiency and as a result about her performance on the test:

P: I agree that it's difficult yes. // I mean I do believe that my level of English is // is / pretty high. I mean I use to do translations at home for my professors at the university. // I mean, work, papers and different university documents, and then the reports, and the // all kind of specialty magazines that had articles

P: /// Well, with my 16 years of English in my country and // all the translations that I did, I think I am // ok to take the test. Yeah, I am feeling very // I mean I am fully prepared.

For her time limit was an issue affecting the way she felt during the test and the way she perceived her test performance:

P: ... but on the other hand // it does make you feel nervous and it's, it is a // you feel pressure and then // it's just // it could get // pretty not annoying but / stressful. Because I mean you get headache and if you are stressed, me // I do get headaches. Then I cannot think clearly

P: But still I was not pleased with the oral part of the test, I am telling you. I was not pleased at all. I could not concentrate and then // the timing was too short, like one minute there was no way that I can // you know// you block, you get blocked when you have only 1 minute to think about and //. It's just //

Chloe thought that given more time she could have scored better on the reading and writing sections of the test as it was her belief that reading specialized and academic texts and writing academic essays required more time:

P: // For the CAEL essay I did not have time to write anything on an academic level so what I put down on the paper there was just the normal level. I did not feel that I wrote an academic essay. That's for sure, I did

not. I just // you know, told them with my own words what I // was able to understand from this // as I read, I mean. So // the difference was not that big but if I had the time you know to put into writing an essay I would make it better. And of course // it's more difficult to write an academic essay than to write // to a friend obviously.

R: What about the CAEL readings, did you find them difficult? Were you comfortable with them?

P: /// I was comfortable but not very comfortable because I did not have the time to look // I could not take the time to read it and to understand everything, so I'd have liked to // you know to have done that.

R: Ok

P: Then when the time is pressing you just go over it and you can't really understand what they are trying to say at the first sight

P: Oh yeah // yeah /// well, the reading It is // again difficult because it contains, I mean some of the questions, some of the questions do require a lot of time in order to for these questions to be answered correctly.

Chloe insisted on using "professional" high-level English not only in formal and/or academic setting but constantly:

P: Of course when you are talking to a professor you have to be more careful, you have to choose your words. You have to at least try to speak at a higher level. // Like academic level. I am using this like word, I started to get it from everybody. I have to get it out of my // vocabulary.

R: // So, let's say you are going to a store, talking to the shop-attendant, how would you rate this? Would you feel more comfortable talking to the shop-attendant than to a // professor?

P: Yeah, usually //when you talk to somebody in a store you do not think it you just words just come out. But when you are talking to the professor you want to make sure that you are understood and you // talk to // that certain level. I mean you talk in a professional way. I do try to talk professionally // when I ma in the store. I do not like to talk, you know, I mean I do not like this street talk. Just //

It is also important to note that Chloe did not study English in order to enter a Canadian university, she appeared to be interested in English as a language:

P: // Yeah, pretty comfortable // What I am trying to do is // when you want to // I am trying to keep in mind that I need to improve my English all the time so // you know, even when go to a store to buy something there is your chance to improve your English. Although the level is not an academic one but you can still stay at a high level but and try and get better all the time.

P: At the moment, no. // I love English, and at home I studied English for 12 years in my country. Actually 16, yeah, during university as well. And I just at times studied it by myself, I loved English at the extent that I just take the grammar I read it like a novel. I liked it, and I read a lot, to take more English, sure. I agree that nobody can be perfect, you know. I mean, in one field. There is also things to learn, you know there are things that you forget if you do not // go over it from time to time.

4.2.7.4. Common comments:

Listening section

Wendy, Dave, and Chloe said, especially in the interviews after taking the test, that the listening section was most difficult and they expected lowest scores on this section as they were worried about:

- the spelling of some of the more specialized/ complex words,
- the speed of the lecture,
- and the need to combine listening, reading, and writing.

Besides that the participants said that this section made them feel nervous, confused, and lost.

The effect of retesting:

As observed by Scott and Madsen's (1983) retesting seems to reduce test-takers' frustration and nervousness both with the high and low proficiency groups. This conclusion was confirmed by the accounts of the participants in this study.

Wendy, Dave, and Chloe said that they felt generally less nervous when they took the test for the purposes of this study compared to their testing experience on the 31st of January. The reasons for that were varied:

- familiarly with test format
- awareness of knowing what comes next
- certainty that they have prepared for the test (Dave and Chloe)
- not being worried about the final test score (Chloe)

4.2.8. Results Summary

In conclusion, the key findings from Phase Two were the following:

- 1.) The participants' proficiency levels seemed to shape their perceptions of and feedback on the test.
- 2.) The participants' opinions whether previous knowledge about the test topic was helpful once again seemed to depend on their proficiency.
- 3.) It seems that the participants' believed that if given more time they would have performed better, especially on the writing and the reading sections of the test.
- 4.) Finally, the participants' individual characteristics seemed to shape their perceptions of what was interesting, easy, and difficult on the test.

This Chapter presented the results of Phases One and Two of the study conducted for the purposes of this thesis. The following Chapter includes the discussion of these results.

CHAPTER FIVE: DISCUSSION

This Chapter presents the discussion of the results from Phases One and Two of the study. It starts with a comparison of the results from Phase One and Phase Two. Then answers to the research questions asked in the thesis are presented as well as findings that emerged from the data.

5.1. Phase One and Phase Two – A Comparison

Time, student comfort, and topic were the most frequently mentioned categories by the test-takers in responses to the open-ended questions on the post-test questionnaire in Phase One of the study. Phase One also included analysis of the test-takers' bandscores and their perceptions of the test. In this Chapter conclusions based on these categories are compared with the findings from Phase Two.

*Time limits*³⁷

61.1% (n=119, RR=97.5%) of the test-takers indicated that time was insufficient for the test. The participants in Phase Two indicated that time limits were directly related to test scores. Dave and Chloe believed that if given more time they would have scored better on the reading and writing sections.

³⁷ Although 61.1% (n=119, RR=97.5%) of the test-takers from Phase One said that time was insufficient of the test, the fact that the category time got uniformly negative responses has to be taken into consideration.

Time was also seen to be related to test-takers' emotional states during the test. For instance, Chloe reported feeling nervous, pressured and unable to concentrate due to the limited time allotted for certain test sections. However, the participants in Phase Two did not as negatively rate the category time as it was by the test-takers in Phase One.

In Phase One, categories such as reading, topic, and listening elicited mixed responses compared to the category time. In Phase Two the listening section of the test, especially after the participants wrote the test, was extremely negatively evaluated. It was considered to be the most difficult one, as it required combining listening, reading and writing. Although the reading section was not always rated as the easiest one, it was evaluated by the participants in Phase Two more positively than by the test-takers in Phase One. Dave mentioned that he disliked the writing section, as it was his weakness.

The chi-square analysis in Phase One established significant relationships between test-takers' perceptions of the test time limits and whether the test:

- was fair
- topic was important
- topic was interesting

It was suggested that that test-takers' perceptions of speededness are related to their perceptions of test fairness, topic importance and interest. Further, in Phase One, it was concluded that test-takers' perceptions of speededness (i.e. whether they had enough time to take the test) was a strong characteristic that was related to their perceptions of test fairness, interest in test topic and the

importance that they attributed to test topic. It was also speculated that insufficient time, from test-takers' point of view, could make a test topic seem less interesting and also less fair to them and therefore more important because they might experience interference with their performance. In Phase Two only Chloe said that she would have perceived the test as more fair if the time allotted for some section was increased. The findings from Phase Two did not coincide with the rest on the conclusions made in Phase One.

Both in Phases One and Two, test-takers' / participants' perceptions of the overall testing experience was not affected by their perception of time limits. However, Phase One suggests that high proficiency test-takers tend to have slightly more favorable attitudes to the entire testing experience than low proficiency test-takers. On the basis of Phase Two it can be said that test-takers with higher proficiency are more capable of expressing what they liked and did not like about the testing experience.

Student Comfort

The issue of student comfort also emerged in Phase Two. However, the aspects of student comfort discussed in Phase Two differed from those in Phase One. Dave mentioned the importance of the physical comfort of the room for his test performance. Chloe made a comment about the background noise in the testing room. She also raised issues different and/or more specific from those included in the post-test questionnaire, they included:

- providing a safe place for storing test-takers' belongings

- reducing/ eliminating background noise made by the other test-takers especially during the Oral Language Test
- ensuring that the CAEL test score are accepted by a bigger number of universities throughout Canada

Chloe noted that the test price was accessible, an issue which was of importance to her.

The chi-square analysis in Phase One suggested that test-takers' perceptions of whether the testing room was comfortable seemed to be related to test-takers' perceptions of whether the test:

- was good experience
- was fair
- reflected true English knowledge
- was enough to take the test
- version

In Phase Two, the participants' perception of the comfort of the testing room seemed to be related only to their perception of whether test reflected their true knowledge of English.

The finding in Phase One that test-takers' perceptions of room comfort were affected by the test version they wrote might be attributed to characteristics different from test version/ topic such as background noise in the testing room (as suggested by Chloe).

The issue of sound quality of the tape examined in Phase One was not mentioned by the participants in Phase Two. For the purposes of Phase Two the

test instructions were played on a CD instead of being spoken by a test administrator. The three participants approved of the former method.

Test Topic

Test topic was broadly discussed both in Phases One and Two of the study. 18.3% (n=23, RR=63%) of the test-takers in Phase One said that they liked the topic of the test while 7.3% (n=9, RR=62%) said that they did not like it. However, in Phase Two test topic was mostly mentioned as being easy or difficult, and not as much as being liked or disliked. Only Chloe mentioned that liking the test topic could help her understand it better. The issue of the test topic being interesting, important, and having previous knowledge about it also emerged in Phase Two.

57.5% (n=100, RR=87%) of the test-takers in Phase One did not think that previous knowledge about the test topic was helpful. The findings from Phase Two, in a way, also coincided with this conclusion. Dave was the only participant for whom the test topic was directly related to his test scores. He thought that having previous knowledge about the test topic would be helpful for the reading and particularly the writing section. In fact, Dave's opinion was shared by 26.1 % (n=12, RR=23%) of the test-takers in Phase One who initially had said that previous knowledge about the test topic was helpful. For Dave easy test topic was one that he had previous knowledge about. Chloe thought that having previous knowledge about the test topic did not make the test easier.

The conclusion made in Phase One that topic importance might be more related to whether test-takers found the test topic interesting than if they had previous knowledge about it did not coincided with the findings from Phase Two.

In Phase One there were not any significant relationships established between test versions and test-takers' perceptions of the test. In Phase Two, Wendy said that her perception of the testing experience changed because of the test topic but she could not explain in what way. Dave attributed great importance to test topics, he saw a relationship between test topic and:

- previous knowledge about it
- the final test score

Overall, based on the findings in Phases One and Two test version can be excluded as a category influencing test-takers perceptions of the test.

Test-takers' bandscores

In Phase One no significant relationships and/or meaningful trends were established between test-takers' overall bandscores and their perceptions of whether the:

- test was good experience
- test reflected true English knowledge
- test topic was interesting
- time was enough to take the test
- knowledge of test topic was helpful

In Phase Two, there is an evidence of a relationship between the participants' bandscores (i.e. proficiency) and certain perceptions of the test.

Dave and Wendy did not find the test topic interesting while Chloe was positive that the test topic was interesting. Perhaps it can be speculated that test-takers with higher proficiencies are more likely to show interest in test topics while those with lower proficiencies are more occupied with completing the test. Test-takers' individual characteristics and attitudes in this case should also be taken into serious consideration. Chloe was the only participant in Phase Two who demonstrated intrinsic interest in English and viewed the testing experience as challenge and as a way to prove herself.

As far as the participants' perceptions of time limits were concerned, both Dave and Chloe declared that the allotted time was not enough. However Wendy, the participant with lowest proficiency of the three, did not seem to share this opinion. It might be speculated that there is a threshold of proficiency under which test-takers are unaware of the test characteristics (e.g. time) that might affect or seem to affect their performance.

Dave, unlike Chloe, attributed a lot of importance to having previous knowledge of the test topic. In his case this might be a direct consequence of his proficiency level. Wendy, who scored in the same bandscore, did not share his view. As already mentioned, a possible explanation for this might be the fact that she lacked awareness of test characteristics (e.g. time) that might affect her performance and/or more detailed knowledge of the specifics of the CAEL test.

The findings in Phase Two did not coincide with the indications in Phase One that the test-takers' bandscores were related to their perceptions of test fairness when the margin between low and high proficiency was set between bandscores 30 and 40. Said in other words, test-takers with lower proficiencies are more likely to perceive a test as unfair. The three participants in Phase Two had different reasons for questioning, to some extent, the test fairness. However, overall they agreed that the test was fair.

The findings from Phase Two, which coincided with the ones made in Phase One, were the following:

- test-takers' bandscores impact their perceptions of topic importance when the margin between low and high proficiency was set between bandscores 40 and 50
- that low proficiency test-takers' attribute more importance to test topic than high proficiency test-takers.

Dave attributed a lot of importance to the test topic whereas Chloe did not share his point of view. Wendy, who had scored in the same bandscore as Dave, did not share his opinion either. However, in her case, the reasons might be different as already discussed.

5.2. Discussion of the Research Questions

This section of the Chapter presents answers to the research questions asked in the thesis.

Test-takers' perceptions of high stake tests seem to depend on test-takers' reasons for taking the test. As demonstrated by the results from Phase Two of this study, even a seemingly homogenous group of test-takers can have different perceptions of the test depending on their individual characteristics, language proficiency, perceptions of how they performed on a certain test, and motivation for that particular test. Test-takers with high proficiency levels seem to be more aware and/or capable of expressing what they think (e.g. liked and did not like) about the testing experience. In addition, as it was suggested in Phase One high proficiency test-takers tend to have slightly more favorable attitudes to the entire testing experience than low proficiency test-takers. It can be speculated that there is a threshold of proficiency under which test-takers lack awareness regarding which test characteristics (e.g. test time) might affect their performance. Other test-takers' perceptions of the test and the entire testing experience also seem to be affected by their proficiency (e.g. the ability of the test-takers' to identify what was difficult for them in the test, what was problematic for them in the test, etc.).

Test difficulty

Test-takers with low proficiency perceived as difficult, test sections requiring of them to express their opinion and/or to express others' opinions in their own words. They viewed as easier test sections providing more scaffolding.

On the other hand, test-takers with higher proficiency seemed to consider as easier the opportunity of being able to express their thoughts and ideas freely as this gave them a sense of control over the testing experience. In fact, test-takers' higher perceived sense of control over the testing situations has proved to be beneficial (Averill, 1973, cited in Zeidner, 1990). However, when considering this finding, the individual characteristics of test-takers should not be ignored. Some test-takers enjoy having more liberty while others prefer to follow certain structure.

With regard to test topics, they were considered difficult if test-takers did not have previous knowledge about them. However, it should be noted that the test-taker with low proficiency (Dave) who was above the previously discussed threshold of proficiency, considered having previous knowledge about a test topic as crucial. For the test-taker with high proficiency levels (Chloe) previous knowledge about the test topic did not seem to matter as much with regard to test difficulty.

“Cracking the code” of the test

From the analysis of Phase Two it became clear that test-takers with certain proficiency levels were most likely to be interested in “cracking the code”

of a test. Although the three participants (Phase Two) used test-taking strategies (to different extents) Dave was the only one interested somehow “in cracking the code” of the test. (e.g. He was reading about test topics he knew, using preparation books about the CAEL test, etc.). It can be speculated that test-takers with such proficiency would be more interested in “cracking the code” of the test than those with higher or lower levels. Test-takers with higher proficiency levels are more likely to rely on their language proficiency and to use test-taking strategies as a facilitation device. Test-takers with lower proficiency levels can be considered to be under the threshold. This does not allow them to develop enough awareness of their own language proficiency and/or the test characteristics in order to utilize test-taking strategies.

Test topic importance and interest

Test-takers’ proficiency seems to be to some extent related to their perceptions of test topic interest and importance as well. Perhaps it can be speculated that test-takers with higher proficiencies are more likely to show interest in test topics while those with lower proficiencies are more occupied with completing the test. Test-takers’ individual characteristics and attitudes in this case should also be taken into serious consideration. Chloe was the only participant in Phase Two who demonstrated intrinsic interest in English and viewed the testing experience as challenge and as a way to prove herself. However, the finding that test-takers’ proficiency is related to their perceptions of

test topic interest and importance did not coincide with the results from the statistical analysis in Phase One.

It can be concluded that test-takers with lower proficiency levels tend to attribute more importance to test topics (e.g. Dave). Reasons for this can be the test-takers' belief that knowledge about the test topic might help them perform better. This finding did not coincide with the results from the statistical analysis when the margin between high and low proficiency was set between bandscores 40 and 50.

As far as the relationship between test-takers' proficiency and their test-taking strategies is concerned, Phase Two provided some evidence that there was a proficiency threshold beneath which test-takers could not make use of test-taking strategies (i.e. they could not relate certain test-taking strategies to specific parts of the test). The reason might be that these test-takers lack awareness and/ or knowledge about:

- their real language skills, and
- compensation techniques (used when test-takers' proficiency is not at the necessary level)

It is important to explain at this point that test-taking strategies can help test-takers' performance but cannot compensate for significant lack of knowledge/ proficiency. Dave was a very good illustration of this claim. Despite his broad supply of test-taking strategies and knowledge of the test specifics, he was not able to pass the test (i.e. to score at bandscore 40 or above in this context). However, it might be speculated that possessing test-taking characteristics and

knowledge of the test characteristics, in cases like his, might be beneficial for test-takers' motivation and positive attitude towards the target language.

On the other hand, the combination of relatively high language proficiency (e.g. CAEL bandscore 70) and test-taking strategies is clearly advantageous to most test-takers. Chloe, the participant who scored in bandscore 70, reported on the usefulness of some of the test-taking strategies she learned in the CAEL workshop. Compared to the strategies Chloe used, a part of Dave's test-taking strategies were based on certain conditions, such as having previous knowledge about a specific test topic, having more time, etc. Chloe also showed that test-takers with higher proficiency might be able to make use of more complex test-taking strategies.

Thus, perhaps test-takers with higher proficiency (e.g. CAEL bandscore 60-70) are better equipped to take advantage of test-taking strategies. Test-takers with low proficiency cannot compensate for their lack of language knowledge or "pass" the test using test-taking strategies alone.

It can be concluded, therefore, having more knowledge about their perceptions can help test-takers increase their awareness about their own language proficiency (e.g. areas of weakness that require improvement) and can equip them with useful learning/test-taking strategies to better their performance. However, as demonstrated in Phase Two test-taking strategies can not be a substitute for language proficiency. They can be utilized only as a facilitating device (e.g. despite his strong test-taking strategies, Dave could not pass the test).

When the link between test-takers' proficiency, their perceptions of test fairness and their overall judgment of the test is concerned, it appears that test-takers with high proficiency levels are not necessarily more objective than test-takers with low proficiency levels. As discussed earlier in this section, test-takers with higher proficiency are better equipped to communicate their perceptions and/or opinions about tests. Although Phase Two did not show test-takers' perception of test fairness to be related to their proficiency, Phase One points at such a likelihood. Perhaps this issue is worth exploring in a further study with a larger sample using both quantitative and qualitative methods.

Test version is another test characteristic that appeared to be related to test-takers' perceptions of test fairness and subsequently to their proficiency. Phase One suggested that test version was not related to test-takers' perceptions of the test. The test-takers/ participants in the two phases of the study agreed that the test was fair overall. The participants in Phase Two (who wrote parallel test versions) identified certain differences between their CAEL test-taking experiences:

- The reading section was perceived as the cause of differences between the different tests. Dave said that the tests were different depending if he needed to express his personal opinion and/or put others' opinions in his own words. He distinguished this perception of his, from his perception of test topic difficulty and/or vocabulary knowledge.
- Test-takers' emotional state (e.g. being nervous, relaxed, etc.) and/or immediate pragmatic needs (e.g. being worried about the final test

score) at the time of taking the test, as noted by Chloe, might also affect their perception of parallel test versions.

5.2.1. Findings

The final section of this Chapter presents the findings that emerged mainly from Phase Two of the study. Some of these findings are to some extent linked to issues already mentioned in the discussion of the research questions.

Perceptions of test fairness:

The reasons why test-takers might perceive a test as being unfair appeared to be rather varied. The most frequently recurring ones were test time limits, test-takers' proficiency, test-takers' perception of construct validity, and the number of test-takers "passing" the test. In both Phases of the study the test-takers said that test time was insufficient and thus it was viewed as a possible reason for their unsatisfactory test performance. When talking about test-takers' perception of construct validity, Chloe expressed doubt that the structure of the listening section resembled a real university lecture. It also can be concluded that if the test would have appeared to be more fair if greater numbers of test-takers "passed" it (The notion of "passing" is relative, depending on the situation of the test-taker).

Perceptions of good testing experience

Test-takers seemed to distinguish between a good testing experience, and their perceptions of other test characteristics such as:

- test time limits
- test topic
- test difficulty
- and test-takers' proficiency

In addition, Phase Two provided ample evidence that test-takers can differentiate between different emotional states such as, feeling nervous, uneasy, uncomfortable, worried, wondering, and/or pressured/stressed.

Overall, although test-takers seem able to differentiate between different emotional states, their perceptions of a good testing experience seem to be determined most by their general emotional state during the test (as opposed to certain parts of it) and their perceptions of the testing room comfort. As mentioned in Phase Two, the thematic organization of the test by test topic was seen as a positive trait of the test.

Perceptions of whether the test reflected test-takers' true knowledge of

English

The main test characteristics that seemed to be related to test-takers' perceptions of whether the test reflected their true knowledge of English were:

- time limits. Test-takers seemed to perceive that if given more time they would have demonstrated better their true knowledge of English. This finding coincided with the results from the statistical analysis in Phase One.

- background noise. Some test-takers seemed to think that background noise prevented them from showing their true knowledge of English due to loss of concentration
- testing over extended period of time. One of participants in Phase Two believed that it was not possible to obtain a real idea of test-taker knowledge in two or three hours. The option of a test extended two or three days seemed more reasonable.

Perceptions of the test as good practice/challenge

This perception was an interesting concept to investigate. If a test-taker perceives that certain aspects of the test (e.g. test topic, essay) are directly linked to their final test score then having exposure to more testing experience is thought of as good practice. Alternatively, a test can be regarded as a challenge (in the positive sense of the word) if a test-taker seeks to prove their language skills and personally favors challenges.

Perceptions of test difficulty

Test-takers perceptions of test difficulty with regard to:

- Test sections that had a certain degree of formality (e.g. writing academic essays, reading academic/ specialized texts, listening to academic lecture) were seen as more difficult and/or intimidating.

- The listening section of the test (that the participants wrote for the purposes of this study) was rated by the three of them as very difficult. One of the participants stated that it did not realistically represent a regular university lecture. While a part of the reason why this section was regarded as difficult might be its cognitive load, test-takers' perceptions seemed to play an important part. None of the three participants in Phase Two had attended university lectures in Canadian universities and/or was familiar with this experience. Thus their perception of what a typical university lecture was might have mislead them in their assessment of this section of the test.
- Test-takers' emotional states and their immediate pragmatic goals might also have been related to their perceptions of test difficulty

Test-takers' perceptions of test ease appeared be due to several characteristics:

- length. Apparently, shorter tests were perceived as easier and were seen as more pleasant.
- frequent exposure/ regarding a test task as common. Test tasks involving skills that test-takers used often and/or were frequently exposed to were seen as easier.

- weight of a certain test section towards the final test score.

There was some evidence that test-taker might have perceived as easier test sections that had lesser weight for the final test score.

It is also important to note that the test-takers did not necessarily associate test/ test task difficulty with the score they received. Although the listening section was rated as the most difficult one, the participants' scores (Phase Two) on it were not directly reflected in the scores of the listening section. All participants (Phase Two) received the same scores on this section in both tests. Only Chloe's scores were lower than her scores on the other sections. However, her listening score was the same on both tests. Wendy expected to have highest scores on the reading section but considered the writing section as the easiest. This is to suggest that test-takers' perceptions of test difficulty were not only restricted to their performance and/or the cognitive load of a test section. Other characteristics such as test-takers' individual characteristics, current emotional dispositions, attitudes, and expectations might be at play.

At this point it is important to note that high test scores did not guarantee test-takers' favorable attitude to a certain test section.

Perceptions of insufficient test time. What does it mean?

As shown in Phases One and Two, test-takers, regardless, of their proficiency perceived the test time, especially the one allotted for some test sections as not enough. Test time limits appeared to be related to test-takers' perceptions of whether the test:

- was fair
- topic was important (only Phase One)
- topic was interesting (only Phase One)
- reflected true English knowledge
- was good experience (Phase Two)

However, what do test-takers mean by this? The most recurring explanation provided by the participants in Phase Two was that if given more time they would have scored better, especially on the reading and writing sections of the test. The reasons for this perception could be the following:

- Test-takers believed/ perceived that producing formal writing such as academic essays, listening to academic lectures, and reading academic/specialized texts was more difficult and/or they felt intimidated and therefore it required more time
- Test-takers' perceptions of not having enough time might also be due to test-takers' insufficient language proficiency or in cases with high proficiency test-takers to their lack of certain reading / writing/ listening strategies (such as skimming, scanning when reading). In fact, what the test-takers referred to, as insufficient time, might be their speed of processing information in the target language. Language testers have argued that speed of processing is an essential part of language proficiency (Buck, 1991).

Retesting/test-takers' familiarity with tests and its effect on their perceptions

Research literature in language testing (e.g. Scott and Madsen, 1983; Averill, 1973, cited in Zeidner, 1990) states that retesting/test-takers' familiarity with tests could:

- help reduce their anxiety/ nervousness
- cause them to feel more favorable towards the test
- increase their expectations regarding their test performance (i.e. final test scores)

Phase Two, in fact, confirmed this finding, i.e. retesting/test-takers' familiarity with the test reduces test-takers' frustration and anxiety and make them feel more in control of the entire testing experience. However, being overly relaxed and laid-back during a test might prove to have an adverse effect. Chloe perceived her performance on the test as "lousy" as she was "too relaxed".

In general retesting/ test-takers' familiarity with test can cause test-takers to have more favorable attitude towards the entire testing experience. However, the latter does not mean that test-takers' criticism towards certain aspects/ sections of the test would be reduced/ neutralized.

As far as having increased expectations regarding their test performance, although not directly voiced by the participants in Phase Two, in general one of their main expectations, when retaking a test was to score better.

Test-takers' preferences:

Test-takers' familiarity with certain test formats and/or preferred ways of studying English seems to be related to their perceptions of tests. Test-takers seem to perceive more favorably test formats they are familiar with and/or resemble their preferred ways of studying the target language. This fact became obvious from the recommendations provided by the participants for improving the CAEL test and to future test-takers.

To what do test-takers seem to attribute importance in a test?

The answer to this question depends on a combination of characteristics such as test-takers':

- individual characteristics
- immediate pragmatic goals
- beliefs/ perceptions about the tests
- proficiency and awareness of the level of their own language skills.

Dave (Phase Two) considered test topic as being very important as he fostered the belief/perception that previous knowledge about it would help him score better. The statistical analysis from Phase One demonstrated that test topic was not related to test-takers' perceptions of the test.

Dave also rated the writing section as the most important one as it had most weight for the final test score (in his opinion the writing section was "the truth").

Due to her low proficiency and resulting lack of awareness of her own language skills, Wendy was not clearly able to identify aspects/characteristics of importance to her on the test.

The following Chapter presents the conclusion of the thesis. It also includes suggestions for the CAEL test and testing in general.

CHAPTER SIX: CONCLUSION

As Hamp-Lyons (2000) emphasizes language testing is associated with social, professional, and individual responsibilities. She points out that the test-takers are the ones who have the highest stakes of all stakeholders involved in the testing process. Testers and test-developers, on their part, are the ones who are the holders of the biggest social, professional, and individual responsibilities.

Despite the available research in test-takers' feedback and perceptions "test-takers' voices are those least heard" (Hamp-Lyons, 2000, p. 581). As concluded in this thesis and stated in Hamp-Lyons's (2000) article not all test-takers' feedback should be considered. It is both not practical and/or feasible. However, systematically eliciting test-takers' feedback and perceptions of test proves to be beneficial for test development and validation. For instance, test-takers' feedback and perceptions can provide insight into what a language test really tests. Another maybe even more important advantage of researching and considering research in test-takers' feedback and perceptions is that such an investigation could help test-takers become more aware of their own language proficiency (e.g. realize weak skills and the need to improve them). Thus test-takers could obtain a more realistic and comprehensive understanding of the most suitable course of action they need to adopt in order to achieve their goals.

Testing has social and personal consequences (Hamp-Lyons, 2000) mainly affecting test-takers, their futures and families. The participants from Phase Two Dave, Chloe, and Wendy, attested to this statement. In Dave's opinion the writing

section of the test showed “student’s value”. For Wendy “passing” the test was a “dream”, the most important thing in her life for which she was ready to do everything. It was a crucial step towards her successful future life. For Chloe the test was a way to prove herself and her knowledge of English. For all of them it was, to a different extent, a matter of self-esteem.

Asking test-takers questions about their perceptions could help avoid situations similar to the one described in Norton and Stein’s (1998) article *Why the “Monkeys Passage” Bombed: Tests, Genres, and Teaching*, discussed in Chapter Two. As pointed out by Hamp-Lyons (2000) sometimes test-takers feel uneasy, endangered, incompetent, and uncomfortable to raise objections and/or to question a test. They tend to perceive themselves as being in unequal power relationships in order to voice their concerns and/or as unequipped with the necessary means to do it. Not all of the test-takers are “Chloes”. There are “Wendys”, and there are also “Daves”. There are ones like myself, who have been all of the three: Wendy, Dave, and Chloe while advancing slowly along the road of language learning and testing.

This thesis adopted an ecological approach to language testing (Fox 2002, 2003) which takes into account various characteristics that could possibly affect test-takers’ performance and perceptions. This approach does not distinguish between cognitive and affective factors, it does not consider test-takers’ perceptions as a mere cognitive process. Most importantly as Chalhoub-Deville (2003) urges, citing Anastasi (1986 cited in Chalhoub-Deville, 2003) ‘*consider context*’ (emphasis in original). I would like to add consider test-takers’

personalities and perceptions as inseparable parts of the entire testing process, context and experience.

The study conducted for the purposes of this thesis draws essentially on an ecological approach (Fox 2002, 2003), as it looked at how the participants in Phase Two and their perceptions of the test were shaped by their immediate social context. Due to her very low proficiency and limited experience taking ESL classes, Wendy was totally lost as to how to approach the test and what to do about her insufficient knowledge of English. Lacking familiarity with “the ecology” of the context of university studies in Canada, she was just hoping for a “miracle” in order to pass the test.

Dave’s perceptions of the test were also strongly related to the ecology of the social setting within which he was embedded. He had a lot of experience with taking ESL classes and recognized their usefulness especially for his essay writing. However, Dave could not see the connection between taking more ESL classes and preparing himself for the CAEL test. Because of that, he tried other ways to pass the test, some of which (e.g. reading about test topics) were also determined by another aspect of his immediate social surroundings - the high number of Chinese students in this educational institution exchanging strategies for passing the test. That Dave’s perceptions of the test were so strongly shaped by the social context he was in, was best proven by his definition of “passing” the CAEL test which was only valid in the educational institution within which he had been taking ESL classes and wanted to enter.

Chloe's responses to the test seemed to relate to her awareness of her high English language proficiency and her experience translating specialized texts from English to her native language and vice versa. As a result Chloe could not see the difference between her expertise in translating and the fact that she lacked certain experience and knowledge about the type of academic English used in university setting (especially because she did not attend ESL classes in Canada). Another important feature in Chloe's accounts of her test-taking experience was her belief that if given more time she would have performed better on the test. This once again clearly shows her unfamiliarity with the university setting in Canada and the time constraints students face. As a result Chloe's perceptions of her language proficiency did not correspond to her real proficiency in academic English.

The above-mentioned case studies demonstrate that different test-takers can provide different insights about a test or a testing situation. The suggestions about the CAEL test and testing in general that follow show what can be learned from feedback and perceptions provided by the stakeholders who hold the highest stakes— the test-takers. In fact, test-takers' feedback and perceptions can offer valuable suggestions with regard to test validation, reliability, and development.

6.1. Limitations of the thesis

1.) It is very important to mention that the data obtained through the three research instruments (post-test questionnaires-Appendix A, Interview questions - Appendix B, and journals) might not represent the test-takers/ participants' true

opinions about the test. There might be a difference between what the test-takers/ participants reported about the test and what they actually thought about it.

2.) From an ecological perspective to language testing (Fox, 2002, 2003) my own interpretations and perceptions have become a part of the research. However, this is an inseparable part of any research study. Although I mention this as a limitation it is an inherent to the ecological approach to language testing (Fox, 2002, 2003).

3.) Likewise, it is important to explain that any findings and/or conclusions about test-takers'/ participants' perceptions of the test are based on the test-takers'/ participants' accounts of their perceptions regarding the test.

4.) As far as Phase Two is concerned, the time gap between writing the test for the purposes of this study and the interviews should also be considered in relation to the accuracy of the participants' accounts of the test and the testing experience. The effect of the time that passed between writing the test and being interviewed about it could be addressed in a future study.

5.) Given the limited number of participants in Phase Two, it would be worthwhile to replicate the study with a larger sample in order to further investigate these results. More specifically, the present study could be replicated examining how test-takers define/perceive:

- test difficulty
- test fairness
- what is important in a test

6.) Phase One was used to set the framework for the issues to be researched and discussed in Phase Two. Phase Two on its part, provided insight, where applicable, into the data obtained from Phase One. However, if the data in Phase Two were collected and analyzed prior to the data in Phase One, the study might have taken a different course. As discussed earlier the questionnaire focused the attention of both test-takers and researcher on specific issues. An open interview might have not lead to the same issues.

7.) As far as the relationship between the research literature discussed in Chapter Two and the study conducted for the purposes of this thesis, it should be noted that the former deals with mostly cognitive studies. However, the approach adopted in this thesis is ecological (Fox, 2002, 2003). As a result, some of the issues discussed in Chapter Two could not directly relate to the results obtained from this study. However, there are a number of important links between Chapter Two, and the design and results from this study:

- Hill (1998) suggests that test-takers' feedback can be used to support statistical analyses performed for identifying problematic areas of tests and testing situations. As it was demonstrated in this study test-takers' feedback/ perceptions

can offer some insight into the results obtained from statistical analyses.

- The data from Phase Two also identified testing conditions that might be perceived as problematic, negative and/or disadvantaging by the test-takers (e.g. background noise during the OLT, lack of secure place for storing test-takers' belongings, etc.) The testing literature agrees that identifying such issues is one of the main applications of test-takers' feedback (Hill, 1998; Zeidner and Bensoussan, 1988; Zeidner, 1988).

6.2. Suggestions for the CAEL post-test questionnaire:

1.) Based on the analysis in this thesis, there are a number of new aspects of student comfort that should be investigated in the post-test questionnaire:

- the need for a safe place for storing test-takers' belongings during the test
- background noise made by the other test-takers, especially during the Oral Language Test
- CAEL test score is acceptance by a larger number of universities throughout Canada
- test price

2.) The coding of data for Phase One and the results from Phase Two of the study suggested that along with the open-ended questions, the test-takers and especially those with lower proficiency levels need more scaffolding (i.e. more prompts in the free-writing section) in order to give more specific and detailed feedback.

Another important consideration for obtaining test-takers' feedback is giving the test-takers opportunity to choose different means to do so. Test-takers' perceptions of their strengths and weaknesses in the target language come into play when providing feedback. As it happened when conducting the study for this thesis, one of the participants (Dave) did not keep a journal because he disliked writing and perceived it as his weakest skill in English. However, during the interview, this participant provided very valuable data (he thought speaking was very easy). In such cases taping test-takers' feedback might be an option. Given practical considerations this could happen immediately after the Oral Language Test, especially for the test-takers who take it after writing the other part of the test.

6.3. Suggestions for the CAEL Assessment:

1.) One of the participants (Chloe) from Phase Two said that bigger spaces/gaps between the questions in the lecture (the listening section) were confusing and as a result she missed several questions. Although having such spaces/gaps in real lectures is normal, in order to provide test-takers with optimal

conditions to give their best performance, a more balanced spacing between questions on the lectures should be given some consideration.

2.) The three participants in Phase Two favored having the test instructions played on CD. They thought it made the instructions clearer, helped them pay attention to them, and helped them adapt faster for the listening section. Once again in order to provide test-takers with optimal conditions to give their best performance, playing the test instructions on CD at all times should be considered. In addition, such practice could guarantee consistency in the test administration, and in training test administrators especially in test centers outside Canada.

3.) As mentioned earlier, one of the participants (Chloe) from Phase Two doubted that the listening section resembled a regular university lecture. Perhaps test-takers who are not familiar with the nature of studying in Canadian universities, the regular workload in and outside of classes, and the time provided on examinations, should be given a brief overview of this. This way test-takers would have more realistic expectations and could better evaluate their own language proficiency. This suggestion could also have implications for test-takers' perceptions/attitudes to the listening section. As already demonstrated in this thesis, test-takers' perceptions/attitudes are related to their performance on tests.

6.4. Suggestions for testing in general

1.) As already said in this chapter, different methods of obtaining test-takers' feedback should be considered. Test-takers' strengths and weaknesses in the target language and their perceptions of them matter because instruments for eliciting test-takers' feedback can also serve as means for obtaining a more accurate picture of test-takers' actual language proficiency.

2.) When considering test-takers' feedback and its elicitation, test-takers' proficiency levels should be taken into consideration. Test-takers with different proficiency levels have different types of feedback to offer. As already demonstrated in the thesis, test-takers with different proficiency levels have different levels of insight, knowledge and awareness of the testing experience, test characteristics, and their own language proficiency. Test-takers' proficiency could be used as a principle when eliciting feedback.

3.) In order to ensure that the voices of test-takers, who undoubtedly have the highest stake of all participants in the testing process, are heard, tests like TOEFL, and SAT I, should consider developing and using post-test questionnaires. In order to "bias for the best" (Swain cited in Fox, 2004a) as well as to establish consistency in eliciting test-takers' feedback, post-test questionnaires should be used on general basis not only when piloting new tests.

6.5. Directions for Future Research

The findings and suggestions discussed in this thesis raise further issues.

Some of the findings of the study were inconclusive due to the limited number of test-takers/ participants who provided the research data. Exploring further the questions listed below could prove worthwhile:

- 1.) How much do test-takers' individual characteristics (i.e. personality) contribute to test-takers' performance?
- 2.) Which test-taking strategies are really helpful and which ones only seem helpful?
- 3.) Is it easier to write topic-based tests (e.g. CAEL) or tests with discrete item types (e.g. TOEFL)? Which one do test-takers prefer more?
Which one do test-takers find more helpful? How do test-takers perceive them?

Some of the answers to these important questions might be provided by further research in test-takers' accounts of their perceptions and feedback.

References:

- Alderson, J.C. (1988). New procedures for validating proficiency tests of ESP? Theory and practice. *Language Testing*, 5, 220–232.
- Arnold, J. (Ed.). (1999). *Affect in Language Teaching*. Cambridge: Cambridge University Press.
- Bachman, L.F. (1990). *Fundamental considerations in language testing*. Oxford and New York: Oxford University Press.
- Bachman, L.F. (2002). Some reflections on task-based language performance assessment. *Language Testing*, 19 (4), 453-476.
- Bachman, L.F., Davidson, F., and Milanovic, M. (1996). The use of test methods in the content analysis and design of EFL proficiency tests. *Language Testing*, 13, 125-150.
- Bachman, L.F., and Cohen, A.D. (1998). Language testing – SLA interfaces: An update. In L.F Bachman. and A.D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp.1-31). Cambridge: Cambridge University Press.
- Bakhtin, M. (1981). *The Dialogic Imagination. Four essays by M. M. Bakhtin* (C. Emerson and M. Holquist, Trans.; M. Holquist, Ed.). Austin: University Of Texas Press.
- Bakhtin, M. (1986). The problem of speech genres. In C. Emerson and M. Holquist, (Eds.), *Speech genres and other late essays*. (Vern W. McGee, Trans) (pp. 60-102). Austin, TX: University of Texas Press.

- Bereiter, C., and Scardamalia, M. (1983). Does learning to write have to be so difficult? In A. Freedman, I. Pringle, and J. Yalden. (Eds.), *Learning to Write: First Language/Second Language* (pp. 1-24). London: Longman.
- Bialystok, E. (1994). Representation and ways of knowing: three issues in second language acquisition. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 549-569). London: Academic Press.
- Bradshaw, J. (1990). Test-Takers' Reactions to a Placement Test. *Language Testing* 7 (1) 13-30.
- Brindley, G., and Slatyer, H. (2002). Exploring task difficulty in ESL listening assessment. *Language Testing*, 19 (4), 369-394.
- Brown, A. (1993). The role of test-taker feedback in the test development process: test-takers' reactions to a tape-mediated test of proficiency in spoken Japanese. *Language Testing*, 10, 277-303.
- Brown A., and Iwashita, N. (1996). The role of language background in the validation of a computer-adaptive test. *System*, 24 (2), 199-206.
- Brown, J.D., Hudson, T., Norris, J.M., and Bonk, W. (2002). An investigation of second-language task-based performance assessments (Technical Report #24). Honolulu: University of Hawai'i, Second Language Teaching & Curriculum Center.
- Buck, G. (1999). The testing of listening comprehension: an introspective study. *Language Testing*, 8 (1), 67-91.
- Casado, M. (2001). Foreign language anxiety of university students. Northern Arizona University Press.

- Chalhoub-Deville, M. (2003). Second Language interaction: current perspectives and future trends. *Language Testing*, 20 (4), 369-383.
- Chapelle, C. (1998). Construct definition and validity inquiry in SLA research. In L.F Bachman. and A.D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 33-70). Cambridge: Cambridge University Press.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge: MIT.
- Clapham, C. (1998). The effect of language proficiency and background knowledge on EAP students' reading comprehension. In A.J. Kunnan (Ed.), *Validation in language assessment* (pp. 141-167). Mahwah, NJ: Lawrence Erlbaum.
- Cohen, A. D. (1984). On taking language tests: What students report. *Language Testing*, 1, 70-81.
- Covington, M.V. (1986). Anatomy of failure-induced anxiety: the role of mediators. In R. Schwarzer (Ed.), *Self-related cognitions in anxiety and motivation* (pp. 247-263). Hillsdale, N.J.: L. Erlbaum Associates.
- Creswell, J. (1998). *Qualitative inquiry and research design: choosing among five traditions*. Thousand Oaks, CA: Sage Publications.
- Davies A., Brown, A., Elder, C., Hill, K., Lumley, T., and McNamara, T. (1999). *Dictionary of language testing*. Cambridge, UK: Cambridge University Press.
- Dornyei, Z. (2001). *Teaching and Researching Motivation*. Harlow, UK: Pearson Education.

- Dulay, H., Burt, M., and Krashen, S. (1982). *Language Two*. Oxford: Oxford University Press.
- Edwards, D., and Potter, J. (1992). *Discursive Psychology*. London: Sage Publications.
- Elder C., Iwashita, N., and McNamara, T. (2002). Estimating the difficulty of oral proficiency tasks: what does the test-taker have to offer? *Language Testing*, 19 (4), 347-368.
- Ellis, N. (Ed.). (1994). Introduction: implicit and explicit language learning – an overview. *Implicit and explicit learning of languages* (pp. 1-9). London: Academic Press.
- Ellis, N. (1998). Emergentism, Connectionism, and Language Learning. *Language Learning*, 48 (4), 631-664.
- Ellis, N. (1999). Cognitive Approaches to SLA. *Annual Review of Applied Linguistics*, 19, 22-42.
- Ellis, N. (2002). Frequency effects in language processing. A review with implications of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24, 143-188.
- Ellis, R. (1994a) A theory of instructed second language acquisition. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 79-114). London: Academic Press.
- Freedman, A., and Medway, P. (Eds.). (1994). *Genre and the new rhetoric*. London: Taylor & Francis Ltd.

- Foster, P., and Skehan, P. (1996). The influence of planning time and task type on second language performance. *Studies in Second Language Acquisition*, 18, 299-323.
- Fox, J. (1995). *Initial inquiry in validity*. Unpublished M.A. thesis, Carleton University.
- Fox, J. (2001). *It's all about meaning: L2 test validation in and through the landscape of an evolving construct*. Unpublished Ph.D. thesis, McGill University.
- Fox, J. (2002). Test takers' and raters' accounts of three L2 writing tests. Paper presented at the Language Testing Research Colloquium, Hong Kong, China, December.
- Fox, J. (2002a). CAEL Assessment: Test Score and Users' Guide. Language Assessment and Testing Research Unit. Ottawa, ON: Carleton University.
- Fox, J. (2003). From products to process: An ecological approach to bias detection. *International Journal of Language Testing*, 3 (1), 21-48.
- Fox, J. (2004). Test decisions over time: tracking validity. *Language Testing*, 3, (in press).
- Fox, J. (2004a). "Biasing for the best": An interview with Merrill Swain. *Language Assessment Quarterly*, (in press).
- Fulcher, G., and Reiter, R. M. (2003). Task difficulty in speaking tests. *Language Testing*, 20 (3), 321-344.
- Hamp-Lyons, L. (2000). Social, professional and individual responsibility in language testing. *System*, 28, 579-591.

- Gardner R. (2001). *Language Learning Motivation: The Student, the Teacher, and the Researcher*. Retrieved November 12, 2002, from <http://publish.uwo.ca/~gardner/>
- Geen, R., and Shea, D.C. (1997). Social Motivation and Culture. In D. Munro, J. F. Schumaker and S. C. Carr (Eds.), *Motivation and Culture* (pp. 33-48). New York: Routledge.
- Giltrow, J. (2002). *Academic Writing: Writing and Reading in the Disciplines* (3rd. ed.). Peterborough, ON: Broadview Press.
- Ginther, A. and Stevens, J. (1998). Language background, ethnicity, and the internal construct validity of the Advanced Placement Spanish Language Examination. In A.J. Kunnan (Ed.) *Validation in language assessment* (pp. 169-193). Mahwah, NJ: Lawrence Erlbaum.
- Glazer, B., and Strauss, A. (1967). *The discovery of grounded theory; strategies for qualitative research*. Chicago: Aldine Pub. Co.
- Hill, K. (1998). The effect of test-taker characteristics on reactions to and performance on an oral English proficiency test. In A.J. Kunnan (Ed.) *Validation in language assessment* (pp. 299-323). Mahwah, NJ: Lawrence Erlbaum.
- Jennings, M., Fox, J., Graves, B. and Shohamy, A. (1999). The test-takers' choice: an investigation of the effect of topic on language-test performance. *Language Testing*, 16 (4), 426-456.
- Iwashita, N., McNamara, T., and Elder, C. (2001). Can We Predict Task Difficulty in an Oral Proficiency Test? Exploring the Potential of an

- Information-Processing Approach to Task Design. *Language Learning*, 51 (3), 401-436.
- Krashen, S. (1981). *Second Language Acquisition and Second Language Learning*, Oxford, New York: Pergamon Press.
- Krashen, S. (1985). *The Input Hypothesis: Issues and Implications*. London: Longman.
- Krashen, S. (1994). The input hypothesis and its rivals. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 45-77). London: Academic Press.
- Madsen, H.S. (1982). Determining the debilitating impact of test anxiety. *Language Learning*, 32, 133-143.
- McLeod, S. (1997). *Notes on the Heart: affective issues in the writing classroom*. Carbondale, Ill.: Southern Illinois University Press.
- McNamara, T. (1995). Modeling performance: opening Pandora's box. *Applied Linguistics*, 16 (2), 159-179.
- McNamara, T. (1997). 'Interaction' in second language performance assessment: Whose performance? *Applied Linguistics*, 18 (4), 446-466.
- McNamara, T. (2001). Language assessment as social practice: challenges for research. *Language Testing*, 18 (4), 333-349.
- Miller, C. (1984). Genre as Social Action. *Quarterly Journal of Speech*, 70, 151-167.
- Messick, S. (1989). Validity. In R.J. Linn (Ed.), *Educational Measurement* (3rd ed.). New York: American Council on Education/Macmillan.
- Nachmias, D., and Nachmias, Ch. (1987). *Research methods in the social sciences* (3rd ed.). New York: St. Martin's Press.

- Nevo, B. (1985). Face validity revisited. *Journal of Educational Measurement*, 22, 287-293.
- Nevo, B., and Stetz, J. (1985). Examinees' feedback questionnaires. *Assessment and Evaluation in Higher Education*, 10, 236-249.
- Norris, J.M., Brown, J.D., Hudson, T.D., and Bonk, W. (2002). Examinee abilities and task difficulty in task-based second language performance assessment. *Language Testing*, 19 (4), 395-418.
- Norton, B., and Stein, P. (1998). Why the "monkeys passage" bombed; tests, genres, and teaching. In A.J. Kunnan (Ed.), *Validation in language assessment* (pp. 231-249). Mahwah, NJ: Lawrence Erlbaum.
- Paré, A., and Smart, G. (1994). Observing Genres in Action: Towards a Research Methodology. In A. Freedman and P. Medway (Eds.), *Genre and the new rhetoric*. (pp. 146 – 154). London: Taylor & Francis.
- Purpura, J. (1998). The development and construct validation of an instrument designed to investigate selected cognitive background characteristics of test-takers. In A.J. Kunnan (Ed.), *Validation in language assessment* (pp. 111-139). Mahwah, NJ: Lawrence Erlbaum.
- Robinson, P. (2001). Test Complexity, Task Difficulty, and Task Production: Exploring Interactions in a Componential Framework. *Applied Linguistics*, 22 (1), 27-57.
- Scott, M. (1986). Student affective reactions to oral language tests. *Language Testing* 3, 99-118.

- Scott, M., and Madsen, S.H. (1983) The influence of retesting on test affect. In J. Oller (Ed.), *Issues in language testing research* (pp. 270-279). Rowley, Mass.: Newbury House.
- Skehan, P. (1996). A framework for the implementation of task-based instruction. *Applied Linguistics*, 17, 38-62.
- Skehan, P., and Foster, P. (1996). Task type and task processing conditions as influences on foreign language performance. *Language Teaching Research*, 1 (3), 185-211.
- Shohamy, E. (1982). Affective Considerations in Language Testing. *The Modern Language Journal*, 66 (1), 13-17.
- Shohamy, E. (2000). The relationship between language testing and second language acquisition, revisited. *System*, 28, 541-553.
- Smith, L. (2003). Shooting Yourself in the Foot: Language Testing and Cultural Differences. TESL Ottawa Conference, Ottawa, Ontario.
- Spolsky, B. (1995). *Measured Words: The Development of Objective Language Testing*. Oxford University Press, Oxford.
- Stansfield, C. (1991). A comparative analysis of simulated and direct oral proficiency interviews. In S. Anivan (Ed.), *Current developments in language testing* (pp. 199-209). Singapore: SEAMEO RELC.
- Stansfield, C.W., Kenyon, D.M., Paiva, R., Doyle, F., Ulsh, I., and Cowles, M.A. (1990, Sep.) The Development and Validation of the Portuguese Speaking Test. *Hispania*, 73 (3) 641-651.

- Swain, M. (2001). Examining dialogue: another approach to content specification and to validating inferences drawn from test scores. *Language Testing*, 18 (3), 275-302.
- Volpe, A. (1995). Genetic Classification and the historical method. In S. Puppel (Ed.), *The Biology of Language* (pp. 73-99). Amsterdam, Philadelphia: J. Benjamins Publishing Company.
- Wenden, A. (1998). Metacognitive Knowledge and Language Learning. *Applied Linguistics*, 19, 515-537.
- Woods, D. (2003). The social construction of beliefs in the language classroom. In P. Kajala, and A. Barcelos (Eds.), pp.201-229. *Beliefs about SLA. New Research Approaches*. Amsterdam: Kluwer Academic Publishers.
- Zeidner, M. (1987). Essay versus multiple-choice: type classroom exams: the students' perspective. *Journal of Educational Research*, 80, 352-358.
- Zeidner, M. (1988). Evaluation Studies. Classroom testing: the examinee's perspective. *Studies in Educational Evaluation*, 14, 215-233.
- Zeidner, M. (1990). College students' reactions. Towards key facets of classroom testing. *Assessment and Evaluation in Higher Education*, 15 (2), 151-169.
- Zeidner, M., and Bensoussan, M. (1988). College students' attitudes towards written versus oral tests of English as a Foreign Language. *Language Testing*, 5, 100-114.
- Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing*, 14 (1), 85-106.

- Wigglesworth, G. (1998). The effect of planning time on second language discourse. In A.J. Kunnan (Ed.), *Validation in language assessment* (pp. 91-109). Mahwah, NJ: Lawrence Erlbaum.
-

A P P E N D I X A

APPENDIX A
Test Takers' Questionnaire

Post-test Questionnaire

Directions: We would appreciate your feedback on the CAEL Assessment. What you write here will not affect your own test results in any way, but it may help us to improve the test for future test takers.

Your response is optional and confidential.

Name _____ Date _____

1. What did you like about the test?

2. What did you dislike?

3. Read each of the following statements. Do you agree, disagree or have no opinion?

Please circle only one response for each statement.

This test was fair	Agree	Disagree	Have no opinion
This test was a good experience.	Agree	Disagree	Have no opinion
This test reflects my true knowledge of English.	Agree	Disagree	Have no opinion
Choice of topic is important to me as a test taker.	Agree	Disagree	Have no opinion

4. What would make the test better for you?

5. Which of the following were the most important in your performance on this test? Write a number from 1 to 5 on the lines below to show which was the most important to you. The most important is (1), the second most important is (2), etc. the least important is (5). Use each number only once:

_____ the physical comfort of the room (heat, light, space, etc.) _____ the amount of time allowed for the test
_____ the topic of the test _____ the sound quality of the tape recorder _____ the sample test practice before the test

Page 2

6. Did you like the topic of this test?

^oThe questionnaire's format has been altered to fit this space.

A P P E N D I X B

Semi-structured interview questions

before taking the test

1. What is the first thing that comes to your mind when the CAEL test is mentioned?
2. Do you think that the test would be difficult?
3. What do you think about the test?
4. What is your opinion of the test according to what you have heard about it?
5. Do you feel nervous or anxious about the test?
6. Would you be less or more motivated to study English if you did not have to take the test?
7. What are your strengths in English?
8. What are your weaknesses in English?
9. What do you feel most comfortable doing in English? (speaking, listening, writing, reading)
10. What do you feel least comfortable doing in English? (speaking, listening, writing, reading)
11. What do you expect to be most difficult part for you on test?
12. What do you expect to be easiest part for you on test?
13. Do you think that your test result will show your true abilities in English?
14. Do you feel prepared for the test?

Semi-structured interview questions

after taking the test

1. Was the test difficult?
2. Was the test what you thought? How? And how not?
3. Do you think that your result will reflect your true knowledge of English?
4. Did you feel nervous or anxious during the test?
5. Did you feel more nervous before writing the test than during the test itself?
6. What was the most difficult part for you on test? Is this what you are weak at in English?
7. What was the easiest part for you on test? Is this what you are best at in English?
8. Where do you expect to get highest scores?
9. Where do you expect to get lowest scores?
10. What recommendations would you have for improving the test?

A P P E N D I X C

Table 17: Response rates (RR) of test-takers' perceptions analysed for frequency of response

Test-takers' perceptions	Total number of test-takers	Number of test-takers responded	Response Rate (RR)
Test reflected true English knowledge	200	196	98%
Sound quality of tape was good	200	196	98%
Time was enough	200	195	97.5%
Test was fair	200	195	97.5%
Test was a good experience	200	195	97.5%
Topic was interesting	200	194	97%
Room was comfortable	200	193	96.5%
Topic was important	200	191	95.5%
Did you know something about the test topic	200	176	88%
Was knowledge of test topic helpful	200	174	87%
Like about the test	200	126	63%
Did not like about the test	200	124	62%
How did knowledge help	200	46	23%
Average:	200	169	85 %

Table 18: Response rates (RR) of test-takers' bandscores and perceptions of the test

Test-takers' perceptions	Total number of test-takers	Number of test-takers responded	Response Rate (RR)
Test reflected true English knowledge	200	196	98%
Test was good experience	200	195	97.5%
Time was enough	200	195	97.5%
Test was fair	200	195	97.5%
Topic was interesting	200	194	97%
Topic was important	200	191	95.5%
Was knowledge of test topic helpful	200	174	87%

Table 19: Response rates (RR) of test-takers' perceptions of test time and the test

Test-takers' perceptions	Total number of test-takers	Number of test-takers responded	Response Rate (RR)
Time was enough & Test was fair	200	192	96%
Time was enough & Test was good experience	200	192	96%
Time was enough & Topic was interesting	200	191	95.5%
Time was enough & Topic was important	200	188	94%

Table 20: Response rates (RR) of test-takers' perceptions of test topic/version and the test

Test-takers' perceptions	Total number of test-takers	Number of test-takers responded	Response Rate (RR)
Test topic/version & Test-takers' bandscores	200	200	100%
Test topic/version & Test reflected true English knowledge	200	196	98%
Test topic/version & Test was fair	200	195	97.5%
Test topic/version & Test was good experience	200	195	97.5%
Test topic/version & Time was enough	200	195	97.5%
Test topic/version & Topic was important	200	191	95.5%
Test topic/version & Topic was interesting	200	194	97%
Test topic/version & Was knowledge of test topic helpful	200	174	87%

Table 21: Response rates (RR) of test-takers' perceptions of room comfort and the test

Test-takers' perceptions	Total number of test-takers	Number of test-takers responded	Response Rate (RR)
Room was comfortable & Test topic/version	200	193	96.5%
Room was comfortable & Test reflected true English knowledge	200	191	95.5%
Room was comfortable & Test was fair	200	190	95%
Room was comfortable & Test was good experience	200	190	95%
Room was comfortable & Time was enough	200	190	95%

Table 22: Response rates (RR) of test-takers' perceptions of sound quality of the tape and the test

Test-takers' perceptions	Total number of test-takers	Number of test-takers responded	Response Rate (RR)
Sound quality was good & Test topic/version	200	196	98%
Sound quality was good & Test reflected true English knowledge	200	194	97%
Sound quality was good & Test was fair	200	193	96.5%
Sound quality was good & Test was good experience	200	193	96.5%
Sound quality was good & Time was enough	200	193	96.5%

A P P E N D I X D

OFFICE OF RESEARCH SERVICES

Research Ethics Committee
Room 511A Tory Building
1125 Colonel By Drive
Ottawa, ON K1S 5B6

Telephone: 613-520-2517; Fax: 613-520-2521

17 March 2004

Ms. Elena Aminkova
M.A. Candidate
School of Linguistics and Applied Language Studies
Carleton University
1125 Colonel By Drive
Ottawa, Ontario
K1S 5B6

Re.: Ethics Clearance

Dear Ms. Aminkova,

The Carleton University Research Ethics Committee has reviewed your application for your M.A. thesis entitled *Test-Takers' Perceptions: Behind the Construct of a Test* and given it approval.

The guidelines governing ethics approval are as follows:

1. Ethics clearance is valid for one-year (from date of this letter). If you do not complete the research portion involving human participants indicated in your proposal within that time frame you must contact the Research Ethics Committee so the research project can be considered for an extension.
2. Should there be any significant changes to the proposal you are obligated to contact the Carleton University Research Ethics Committee with the changes.
3. Should any participant suffer negatively from the research please report this to the Research Ethics Committee and indicate what steps you have taken to resolve the situation.