ENTRANCE TESTING AT A PRIVATE ENGLISH LANGUAGE INSTITUTE: TOWARD MORE ACCURATE PLACEMENTS

by

JOEL RAYMOND MURRAY

B.A., University of British Columbia, 1981

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in

THE FACULTY OF GRADUATE STUDIES

(Department of Language Education)

We accept this thesis as conforming

to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

June, 1999

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of LANGUAGE EDUCATION

The University of British Columbia
Vancouver, Canada

Date JUNE 24, 1999.

ABSTRACT

At a large school for adult international ESL students, the placement test, testing process, and placement of students were perceived to be problematic. Three research questions were asked:

1. What are the historical and institutional influences and limitations that have shaped the institute's placement testing process?
2. On what basis do those concerned think that the test is flawed, and how widespread is the problem?
3. Based on the findings of the previous research question as well as test analysis and participant observation, where might problems with the institute's placement testing and the resulting placements lie?

A qualitative design known as evaluative case study was used to describe and analyze the placement test and testing process. To answer question #1, document analysis was used, along with participant observation. To answer question #2, a questionnaire was distributed to all instructors and two administrators, and both were interviewed informally and formally. To answer question #3, for Part One (grammar, vocabulary, reading) of the placement test, 571 answer sheets were analyzed, and two sets of data were correlated using the Pearson product moment correlation coefficient. For Part Two (guided writing), 121 answer sheets were analyzed for completion and trends that might emerge. For Part Three (interview), the

aforementioned questionnaire and interviews were used, and test-takers and teacher-interviewers were observed on five registration days.

Findings related to question #1 indicated that historical and institutional influences and limitations have shaped the placement testing process, with expansion of the school playing an important role. Those related to question #2 indicated that the faculty and administration felt that the test was flawed because of their belief that components of the placement test were of limited usefulness and that the entire placement process was only somewhat useful to them. Additionally, interview and questionnaire data revealed that the problem was widespread. Those related to question #3 indicated that there were problems with each of the three parts of the test. Specifically, the correlation between level as assigned by Part One and level as assigned by interviewer, while high ($r=0.84$), was nonetheless not perfect. Furthermore, the correlation could not take into account that the difference between levels at VELI were not equal. Part Two, the writing test, did not distinguish among levels in its first section, and in its second section, a disparity existed between the directions and a strip of cartoon-like drawings that caused confusion among the test-takers. For Part Three, teacher-interviewers indicated that both a lack of formal training in interviewing and varying competencies of interviewers might contribute to misplacements.

Implications included guidelines both for the institute to follow to contribute to a more accurate placement test and process, and for other language institutes or programs to use in the creation of useful, accurate placement tests.

TABLE OF CONTENTS

TABLE OF CONTENTS (cont'd)

TABLE OF CONTENTS (cont'd)

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

As is usual in an undertaking of this nature, there are many people to thank. First, I would like to thank Dr. Margaret Early and Dr. Bonny Norton; their patience, guidance, and encouragement have been unfailing. Second, my gratitude goes out to Robyn Inman and Boba Stefanovich, without whom so many years ago, none of this would have been possible. Third, I wish to show my appreciation to the administrators and to my colleagues at my place of employment: without your willing cooperation, this thesis would not have come to fruition. Fourth, I would like to thank my good friend Edward Francis, who has kept me sane throughout the years. Fifth, I owe a debt of gratitude to my family: my mother, my twin brother, Frank, and his family; their love and support in good times and in bad have been invaluable. Sixth, I deeply thank Kim and Peggy Mah for everything they have done (and for everything they continue to do) and for their unwavering love, support, and guidance. Finally, I am especially thankful to my wife, Mimi, and my son, Tyler, not only for their love and support, but also for their patience and understanding, especially at those times that I could not be an attentive, loving spouse and at those other times that Daddy could not come out to play or to read Dr. Suess books at night-time.

CHAPTER I. INTRODUCTION

BACKGROUND

To paraphrase Elson (1992), there has been much diverse work on issues related to the testing of English as a second language (ESL) or English as a foreign language (EFL) within the past 25 years. One issue that has received attention has been the topic of placement testing, the initial testing most ESL students undergo upon entering an ESL or EFL program. For the most part, placement testing has been approached from the perspective of issues surrounding the test itself. For example, some researchers have described the implementation of a placement test (Chandavimol, 1988; Malu, 1989). Others, noting problems with a particular placement test, have suggested either ways of increasing the accuracy of the test (Brown, 1989; Ilyin, 1970; Rich, 1993) or the use of alternative forms of testing (LeBlanc & Painchaud, 1985). Missing from these discussions, however, is both a detailed look at the historical and institutional influences and limitations that shape a placement test in a particular social context and an in-depth study of an existing placement test together with a discussion and a set of recommendations for ways in which it may be improved.

Placement testing is seen to be important. Brown (1989) states that "placement is an important element in most programs . . . for sorting students into relatively homogeneous language-ability groupings, sometimes within specific skill areas" (p. 65). According to Chandavimol (1988), placement testing serves two significant needs: the first is "to assign large groups of incoming students to language classes of varying levels of difficulty—from [beginner to] intermediate through upper-intermediate to advanced" (p. 1), while the second is "to determine the students' needs, and thus [provide] a greater degree of flexibility in catering to

those needs" (p. 1). O'Malley and Pierce (1996) note that "accurate and effective assessment of language minority students is essential to ensure that ELL [English language learning] students gain access to instructional programs that meet their needs" (p. 3).

While important and necessary, placement testing can be problematic. One problem with placement testing is that, according to LeBlanc and Painchaud (1985), the process of testing itself removes test-takers from active participation in their own assessment. During many placement tests, test-takers sit and answer test questions. At no time during the testing process are the test-takers asked to assess their own language skills, their own perceived level of second language competence, or their performance on the test. One reason, according to LeBlanc and Painchaud, that test-takers have been removed from participating in their own assessment is that "in practice learners have not often been involved in language testing mostly because it has been felt that they did not have a great deal to contribute" (p. 673) and that there have been questions as to whether students "know enough about their abilities in relation to the language they are learning to make a useful contribution to their evaluation" (p. 674). Regardless, LeBlanc and Painchaud demonstrate that self-assessment can be a valuable alternative to the type of traditional standardized tests most often used in placement testing, yet they recognize that self-assessment does not work under all conditions and thus is not "a panacea for all testing problems" (p. 686). Oscarson (1997), in fact, would appear to argue against self-assessment for placement purposes, noting that while it may be well-suited to mature learners for diagnostic reasons, "the use of self-assessment for grading, promotion, certification, or other 'high-stakes' purposes appears to be inappropriate" (p. 176).

Another problem with placement testing is that the testing process may be flawed in some way. The test may have been designed for another purpose or for another type of student and is

thus, as Brown states, "unrelated to the needs of the students in a particular language program or to the curriculum being taught there" (1989, p. 66). An example of this flaw is a situation in which a test used for the placement of students into a general ESL program designed for foreign students attending a short summer course is used for the placement of English for Academic Purposes students into a highly specific academic content course. As an aside, it is interesting to note that this problem continues to be of concern despite the fact that it was observed by Ilyin as far back as 1970; while trying to find a placement test for a non-academic adult ESL program in San Francisco, she found that the only tests available were those designed either to place first-language learners into grades or to place foreign born students into college-level courses.

A final problem with placement testing is that the resulting placement may be inaccurate for a variety of reasons, some of which are (a) one or a combination of the previously mentioned flaws, (b) inadequate training of examiners, (c) inferior test design, or (d) poor test administration. Inaccurate placement can cause a great amount of trouble at the student level, at the administration level, and at the program level. Gaffney and Mason (1983) point out that students, who incidentally sometimes do not even recognize that a wrong assignment has taken place, may become discouraged and give up on their studies unnecessarily if assigned to a class that they perceive to be too difficult. On the other hand, Gaffney and Mason offer further that students may become bored and cause disruption to the rest of the class if assigned to a class that is too easy. At the administration level, administrators may experience organizational problems resulting from inaccurate placement, such as their having difficulty finding places in other classes for misplaced students after recognizing and trying to solve the misplacement. At the program level, "no matter how dedicated a teaching staff may be or how impressive a program

may at first glance appear to be, inappropriate placement . . . will tend to undermine staff and program effectiveness" (Gaffney & Mason, 1983, p. 98).

THE PROBLEM

An illustration of the problems arising from inaccurate placement can be found at the Vancouver English Language Institute[1] (VELI), a large, privately-owned English language institute for "privileged" (Vandrick, 1995) adult foreign students in Vancouver, British Columbia, Canada. Inaccurate placement at VELI poses difficulties for students, teachers, and administrators alike. Because VELI is a private school, students bear the full cost of their education; that is, there is no subsidization, government or private, of their school fees. As a result, tuition is high: one month of full-time studies consisting of 20 hours of class time per week costs $1,000. Since misplacements are often not recognized until two to three days after the placement process, some students—especially those studying for only one month or two—upon learning that they have been inaccurately placed, feel that they have wasted both time and money by missing almost a full week of appropriate classes. In addition, some students have feelings of anger or frustration because they have difficulty understanding what is happening around them, and as a result, become disruptive—especially those who have been assigned to too high a class for their abilities. Others may become bored because they perceive little to no challenge in the class and thus become abusive or unruly, particularly those who have been assigned to too low a class.

Misplacements can pose problems for teachers at VELI. First, they may find it difficult

[1] A pseudonym.

to handle misplaced students. If teachers are not careful, disruptive, bored, or abusive students can derail concentration, subvert lesson plans, and negatively influence the rest of the class. Concentration may be diverted through some teachers' paying too much attention to misplaced students to the detriment of the rest of the class. Lesson plans may be subverted through teachers' adjusting large amounts of the lesson they are teaching or even discarding parts of the lesson in order to attempt to encourage or placate misplaced students. Other students may be negatively influenced through the misplaced students' actions or words—actions such as the obviously rolling of eyes, the sighing loudly, the slamming shut of textbooks; words such as the insistent complaining that the class is too easy. Second, some students may not recognize that they have been misplaced, and in that case, the teacher may find it difficult to persuade these students that they should move to a class more appropriate to their needs, especially those students who must move from a higher level class to a lower level.

Misplacements can pose problems for administrators at VELI. Once misplacements have been identified, administrators must deal with the misplaced students, and at this point, a number of questions arise: the misplaced students must be moved into an appropriate class, but is there an available seat in that class? What if there is not? Can the students to be moved be persuaded to take a different class? What if they cannot? In addition to these questions, administrators must deal with teachers: those with currently misplaced students must be reassured that administration is working on the problem efficiently and quickly, while those who have to receive students who have been moved as a result of misplacement must be placated if the teachers complain that students are entering their classes at too late a point in the first or second week of classes.

PURPOSE OF THE THESIS

At VELI, the placement testing process—that is, the test and its components used to assign students to the appropriate levels of instruction and the placements resulting from it—is perceived to be problematic, by students, instructors, and administration alike. The perceptions of some students and both faculty and staff are that the placement process is flawed, that serious misplacements occur as a result, and that a refinement of the placement process is urgently needed. Thus, this thesis will seek to describe and analyze the placement testing process (including the placement test itself) at VELI in order to answer the following research questions: First, what are the historical and institutional influences and limitations that have shaped the VELI placement testing process? Second, on what basis do those concerned think that the test is flawed, and how widespread is the problem? Third, based on the findings of the previous research question, where might problems with the institute's placement testing and the resulting placements lie?

This thesis will also examine two important implications of the findings, one for VELI in specific and the other for those involved in placement testing in general. The first implication, while specific to VELI and not to be generalized, can nonetheless serve as a stimulus to others working on placement testing. It is an examination of the insights to be gained, based on the findings of the three questions above and on the reflections of the researcher, as to what can be done in the future at VELI to refine the placement process or to contribute toward more accurate placements. In addition, it explores the constraints that might exist. The second implication, more generic in nature, can serve as a guideline for any institution working on placement testing. It is a review of steps and recommendations, based on the literature in the field, for other language institutes or programs to use in the creation of useful, accurate placement tests.

DEFINITION OF TERMS

Criterion-referenced test

A test in which the assessment of the amount of knowledge or material known by each student is compared with a level of achievement or set of criteria. Subjectively marked tests are often criterion-referenced (Alderson, Clapham, & Wall, 1995).

Norm-referenced test

A test in which the amount of knowledge or material known by each student is compared with that known by other students, with the aim to spread students out along a continuum of general abilities or proficiencies so that differences among them are reflected in the scores (Brown, 1995). Objectively marked tests are often norm-referenced (Alderson, et al., 1995).

Direct test

A test which requires the test-taker to perform the skill to be measured. "If we want to know how well candidates can write compositions, we get them to write compositions. If we want to know how well they pronounce a language, we get them to speak" (Hughes, 1989, p. 15).

Indirect test

A test which "measures the abilities which *underlie* the skills in which [one is] interested" (Hughes, 1989, p. 15).

EE

"Entrance Evaluation." An abbreviation which refers to the first part of the VELI placement test, focusing on grammar, vocabulary, and reading, and divided into seven sections timed at 30 minutes.

ESL

"English as a Second Language." An abbreviation which refers to English taught to non-native English language learners in a country or locale where English is the first, or major, language.

EFL

"English as a Foreign Language." An abbreviation which refers to English taught to non-native English language learners in a country or locale where English is not the native language.

Construct Validity

A term referring to how well a test measure the constructs, the key components in a theory, on which the test is based. "To measure the construct validity of a test a tester must articulate the theory underlying his or her test and then compare the results with that theory" (Alderson, et al., 1995, p. 287).

Face Validity

A term often dismissed as unscientific referring to how well the clients of a test receive it as a test. "Essentially face validity involves an intuitive judgement about the test's content, by people whose judgement is not necessarily 'expert.' Typically such people include 'lay' people—administrators, non-expert users and students" (Alderson, et al., 1995, p. 287).

Reliability

In simple terms, the extent to which test scores are consistent from one application of a test to the next.

Rubric

A scoring scale, also known as a rating scale, that assigns a numerical value to a test-taker's performance depending on the extent to which it meets pre-designated criteria (O'Malley & Pierce, 1996). This scale "may consist of numbers, letters, or other labels . . . which may be accompanied by statements of the kind of behaviour that each point on the scale refers to" (Alderson et al, 1995, p. 107). Rating scales can be either holistic, in which a single score is assigned based on an overall impression of the test-taker's performance, or analytic, in which a separate score for each of a number of aspects of a task is assigned (Hughes, 1989).

Placement Test

A test designed to assess test-takers' level of language ability so that they can be placed in the appropriate course or class (Alderson, et al., 1995).

Specifications for Placement Tests

Alderson, et al. (1995) define specifications as providing "the official statement about what the test tests and how it tests it. The specifications are the blueprint to be followed by test and item writers, and they are also essential in the establishment of the test's construct validity" (p. 9). As used in this thesis, specifications are more a blueprint in that they are a set of recommendations, based on literature in the field, that the creators of placement tests can follow in constructing placement tests applicable to their own contexts.

TI

"Teacher-interviewer." This abbreviation refers to a teacher at VELI who is also tasked with interviewing test-takers during placement testing on registration day.

VELI

"Vancouver English Language Institute." This abbreviation refers to the pseudonym for a large, privately-owned English language institute for adult foreign students in Vancouver, British Columbia, Canada.

VELIPT

"Vancouver English Language Institute Placement Test." This abbreviation refers to the name to be given to a proposed future revision of the VELI placement test.

SUMMARY AND PREVIEW

This thesis is organized into six chapters. Chapter I, "Introduction," provides a background of placement testing, introducing studies important to this thesis and describing the most common approach to the research into placement testing. In addition, problems with placement testing are explained, as well as the purpose of the study and the research questions. Also included are definitions of the specialized terminology used in this thesis and a general overview of what the reader can expect to find in each of the chapters.

Chapter II, "Review of the Literature," touches on a variety of issues concerning placement testing and is divided into two sections, "Testing and Assessment" and "Important Considerations." In the former section, the four types of test—achievement, proficiency, diagnostic, and placement—are discussed, and the term "placement test" is defined. The components found in many placement tests—tests of listening, speaking, grammar, vocabulary,

writing, and reading—are examined, as well as information concerning language test creation, methodology, and practice. Also in this section steps and recommendations in test creation are reviewed. In the latter section, the significance of accurate placement testing and the steps that were taken to improve placements in various environments are examined. In addition, the ethics and effects of testing are discussed.

Chapter III, "Methodology," first restates the three research questions. In order to provide necessary background for the rest of the discussion of methodology, it then describes the site and the VELI testing narrative, that is, a description of the VELI placement test and testing process. After that, the chapter outlines the design and methodology used for the purposes of this thesis, beginning with an explanation of the qualitative design known as evaluative case study, and continuing with a chronology of the various methods—or multimethods (Schumacher & McMillan, 1993)—used in the collection of data. The chapter continues with a description of the participants in the study, and ends with a summary in point form which outlines the research questions and the methodology used to examine them.

Chapter IV, "Findings," is divided into three sections. The first section, "Research Question #1," provides an answer to the first research question and is an exploration both of the history of VELI from the time of its founding up to the present and of the institutional influences and limitations that have affected placement testing at the institute. The second section, "Research Question #2," provides an answer to the second research question and is an examination of why those concerned think the test is flawed and an exploration of how widespread the problem is. The third section, "Research Question #3," provides an answer to the third research question and is a detailed dissection of the three parts of the current VELI placement test.

Chapter V, "Discussion," discusses the findings in relation to each of the first three research questions. Specifically, this chapter examines placement testing at VELI, perceptions of the placement test, and recognition of misplaced students. In addition, it discusses each of the three parts of the current VELI placement test

Chapter VI, "Recommendations and Conclusion," discusses, in relation to two sets of implications, both recommendations for a more accurate placement process at VELI and a set of steps and recommendations for accurate placement for other language institutes or programs doing in-house placement testing and placement test creation. In addition, this chapter concludes with a summary of the thesis, an exploration of other implications including some suggestions for future research which may be conducted in order to contribute to more useful and accurate placement tests, and a final note about placement testing.

CHAPTER II. REVIEW OF THE LITERATURE

INTRODUCTION

This literature review touches on several issues concerning placement testing and will be divided into two main sections: "Testing and Assessment" and "Important Considerations." In "Testing and Assessment," the four types of test—achievement, proficiency, diagnostic, and placement—will be discussed first in order (a) to establish that placement tests are distinct entities, different from other types of test, and (b) to define the term "placement test." The components of language tests—tests of listening, speaking, grammar, vocabulary, writing, and reading—will be examined next, as all or a combination of these components are found in placement tests. These components will be discussed in terms of important considerations and implications in reference to testing generally and placement testing specifically. Language test creation, methodology, and practice will be discussed next, for these points are crucial in the analysis of any test. Finally, steps and recommendations in test creation will be reviewed, as there already exist suggestions or advice that may be applicable to the creation of an accurate placement vehicle.

In "Important Considerations," the significance of accurate placement testing and the steps that were taken to improve placements in various environments will be examined. Accuracy in placement testing is important, and a variety of researchers have been faced with placement tests that have appeared inadequate to the task of placing test-takers into appropriate classes. The ethics and effects of testing will also be examined, as any type of testing can be considered to be powerful, having far-reaching effects on the lives of test-takers.

TESTING AND ASSESSMENT

What is a Placement Test?

Many authors (Alderson, et al., 1995; Harrison, 1983; Hughes, 1989; Underhill, 1991) agree that there are four types of test: achievement, proficiency, diagnostic, and placement. Achievement tests are those which are given in order to assess what the students in a particular course of study have learned. Some authors (Alderson, et al.; Hughes) recognize two kinds of achievement tests: progress, which are administered at various stages throughout a language course, and final, which are administered at the end of a course. Proficiency tests are those which are given in order to measure general language ability regardless of previous language training. They are different from achievement tests in that proficiency tests are not based on a particular language program. Instead, they are based on "a specification of what candidates have to be able to do in the language in order to be considered proficient" (Hughes, 1989, p. 9). Diagnostic tests are those which are given in order to determine the test-takers' strengths and weaknesses. Used to identify areas in which the test-taker needs help, few purely diagnostic tests exist "since it is difficult to diagnose precisely strengths and weaknesses in the complexities of language ability" (Alderson, et al, p. 12). For this reason, achievement and proficiency tests are often used for this purpose.

Placement tests are seen to be different from the other three types of tests. Harrison (1983) defines "placement test" as a test that is created so as "to sort new students into teaching groups, so that they can start a course at approximately the same level as the other students in the class" (p. 4). He adds that placement tests are concerned with the test-taker's present state of general language ability rather than with "specific points of learning" (p. 4), and as a result, "a

variety of tests is necessary because a range of different activities is more likely to give an accurate overall picture of a student's level than a single assessment" (p. 4). While accurate in a general sense, Harrison's definition may be somewhat limited, in that not every placement test may be one of *overall* language ability. By way of illustration, a language program that has a dominant oral/aural focus is unlikely to employ a placement test that places a great deal of emphasis on reading comprehension or writing. Hughes (1989) provides a definition similar to that of Harrison but different in that Hughes' is not as narrow. He states that placement tests "provide information which will help to place students at the stage (or in the part) of the teaching programme most appropriate to their abilities. Typically they are used to assign students to classes at different levels" (p. 14). Alderson, et al. (1995) provide one of the best definitions. They state that "placement tests are designed to assess students' level of language ability so that they can be placed in the appropriate course or class. Such tests may be based on aspects of the syllabus taught at the institution concerned, or may be based on unrelated material" (p. 11).

The Components of Second Language Placement Tests

An examination of the components of second language placement tests follows. These components are explored in terms not of what they are, what should be tested, and how, but of important considerations and implications in reference to testing generally and placement testing specifically.

Tests of Listening Comprehension

Tests of listening comprehension seek to "assess the ability to use knowledge of the language for the purpose of understanding spoken texts" (Buck, 1997, p. 71). In reviewing and

detailing the problems inherent in, history of, and practical advice about the testing of listening in a second language, Buck makes a number of important points. First, he asserts that testing listening comprehension is by necessity indirect; thus, "listening scores will always be influenced by other skills required for task completion" (p. 66). This point is important in an overall sense in that listening tests may not assess listening alone, and as a result, accurate assessments of listening comprehension may not be possible. Second, he points out that listeners' interpretation of any spoken text will be influenced to a large degree by their purposes for listening in the first place, their interests, and their background knowledge. There can be, therefore, a variety of valid interpretations of a text, some of which cannot be anticipated. Third, he states that "virtually all second language listening tests use non-interactive tasks, that is tasks in which the listener cannot interact with the speaker; interactive listening is usually only assessed as part of a spoken interview" (p. 66). Buck adds that "traditionally testers have not been interested in visual media for the presentation of listening texts" (p. 72), the result of which has been a removal of what Weir (1990) refers to as "the wealth of normal exophoric reference and paralinguistic information" (p. 54) found in the visual element.

On the topic of assessing listening comprehension through the use of visual media, Progosh (1996) reports on the value of using video in conjunction with listening assessment. He conducted a study to determine test-takers' opinions of a video-mediated listening test by using a random sample of the second year population of intermediate-level second language students at a tertiary institution in Tokyo, Japan. Using a questionnaire consisting mainly of questions answered on a seven-point Likert scale, Progosh found that "the sample think [sic] video in listening comprehension is a good idea, preferring video-mediated tests over audiocassette tests" (p. 40). He does warn, however, that the video in this case was used to assess learner

achievement and that it has "yet to be determined if such tests can be used for purposes of general language proficiency" (p. 40). Nonetheless, the use of video in conjunction with a listening assessment could be valuable in contributing to more accurate placement instruments, for as Progosh points out, "most people both hear and see in most communicative situations" (p. 35).

Tests of Speaking

Tests of speaking—or as Underhill (1991) puts it, "oral tests"—seek to assess the ability to communicate orally. Underhill states that oral tests are repeatable procedures "in which a learner speaks, and is assessed on the basis of what he says" (p. 7). This assessment is often viewed in terms of "providing information about a person which [is to] be used to predict success in communication in some future real-life situation" (Fulcher, 1997, p. 75). Fulcher, in his review of the problems inherent in, history of, and practical advice about the testing of speaking in a second language, brings up some significant issues. One is that of task; as Fulcher states, "it has been increasingly observed that task type has a systematic effect on speaking test scores" (p. 79). However, "neither the nature nor the degree of the effect of tasks on scores from tests of speaking are well understood" (p. 80). This point is notable in that importance of task type is either not mentioned in some of the literature focusing on testing spoken language (Underhill; Weir, 1990) or only mentioned in passing (Hughes, 1989), yet it may play a large role in how a test-taker's oral abilities are assessed.

On a more practical note, Underhill (1991), in his guide to oral testing, does not assume that the reader has any knowledge of language testing. Stating that the book "deplores the cult of the language testing expert" (p. 1), Underhill writes that it was written for language teachers,

sequenced in the order that a test program might be implemented: starting with questions about needs and resources, continuing with a choice of different oral test types and tasks, and discussing the marking system and evaluation. Underhill's text is useful and valuable in that it details various oral testing techniques and how to create, administer, mark, and evaluate tests of speaking. If there is to be a criticism of this text, it would be one of oversight—that is, Underhill neglects to mention that task type can be a factor on speaking test scores.

Tests of Grammar

Tests of grammar seek to assess "grammatical ability, or rather the lack of it, [for it] sets limits to what can be achieved in the way of skills performance" (Hughes, 1989, p. 142). Hughes asks whether separate grammar testing is justified in these days of communicative language testing. He states that "there is often good cause to include a grammar component in the achievement, placement and diagnostic tests of teaching institutions. It seems unlikely that there are many institutions, however 'communicative' their approach, that do not teach some grammar in some guise or other" (p. 142). As Hughes points out, "there appears to be room for a grammar component in at least some placement tests" (p. 142). Rea-Dickins adds that "the construct of grammar itself carries different meaning but is still considered by many to be an important aspect in the measurement of an individual's overall performance in a language" (p. 87).

Tests of Vocabulary

Tests of vocabulary seek to assess the knowledge of vocabulary, mostly in terms of either depth of word knowledge or size of lexicon. Read (1997) examines the assessment of

vocabulary and notes that while there is a revival of interest in the teaching and learning of vocabulary, "that has not yet led to a re-definition of the role of vocabulary within language testing or to the development of many innovative procedures for lexical assessment" (p. 99). Also on the same topic, Hughes (1989) asks whether separate vocabulary testing is justified. He concludes that "the arguments for a separate component in other kinds of test [other than proficiency tests] may not carry the same strength" (p. 147). For placement tests, Hughes suggests that "we would not normally require, or expect, a particular set of lexical items to be a prerequisite for a particular language class. All we would be looking for is some general indication of the adequacy of the student's vocabulary" (p. 147). The problem is, of course, how to go about doing so.

Tests of Reading

Tests of reading seek to assess the ability of a reader to "extract an agreed level of meaning under specified performance conditions" (Weir, 1997, p. 39). In exploring the testing of reading in a second language, Weir reviews early developments, test methods, test validation methods, work in progress regarding the construct of reading, and the problems inherent in testing reading. While noting that "a direct reading test should reflect as closely as possible the interaction that takes place between a reader and a text in the equivalent real life reading activity" (p. 39), Weir admits that "although full genuineness of text or authenticity of task is likely to be unattainable in the second language reading tests we develop, we still need to select appropriate texts, to be read for realistic purposes, and we expect the reader to extract an agreed level of meaning under specified performance conditions" (p. 39). Hughes (1989) has some very practical advice concerning the testing of reading comprehension. He suggests specifying, as

accurately and completely as possible, the abilities to be tested, and he discusses the selection of test content and the setting of criterial levels of performance. Both Weir and Hughes are important as they both address the advisability of authenticity (or something that approaches it) in a test of reading comprehension.

Courchêne (1995), also on a practical note, offers an alternative to the multiple-choice tests often used in testing reading comprehension: the summary cloze technique. A summary cloze is prepared by summarizing the content of a text so that the resulting new text is approximately one-third the length of the original. This shorter version is transformed into a cloze using a "rational deletion approach as opposed to deleting every nth word" (p. 52). Although he allows that the use of the cloze test has been challenged as a measure of language proficiency, Courchêne maintains that "the cloze procedure . . . can be used to measure reading comprehension if one selects texts of general interest to students, uses a rational as opposed to a random deletion of items, pretests them on both native and nonnative speakers, and uses them in a foreign language context" (p. 51).

In order to demonstrate that the summary cloze technique is at least as good a measure of reading comprehension as the multiple-choice format often used in testing reading comprehension, Courchêne tells of a study using 66 Chinese students at intermediate and advanced levels who were to come to Canada for academic and professional reasons. They were matched for language ability and then randomly assigned to one of two groups. These groups were given five reading passages which were controlled for length, difficulty, and reading level, and which were prepared with both multiple-choice questions and summary cloze formats. Each group did two texts in one format and two in the other, and both did the summary cloze and multiple-choice. The results were compared, and the summary cloze was correlated with other

measures of language proficiency. Courchêne found that the summary cloze technique "produces tests that tend to yield higher levels of reliability than their [multiple-choice] counterparts" (p. 56). Furthermore, he found that "the correlations of the task types with general measure of ESL proficiency do provide evidence that there are no substantial differences in the way the tasks behave, and in general the assumption holds that the two task types are both measures of reading comprehension" (p. 57). There is at least one problem with this new technique: the choice of text and its summarization may affect the test-takers ability to respond to it. Nevertheless, Courchêne's article is important in that he offers a viable alternative to the common multiple-choice reading comprehension sections of most placement tests. As Courchêne states, "initial use in the classroom and as a testing instrument have resulted in positive feedback from the students [who] see summary cloze as having face validity" (p. 57).

Tests of Writing

Tests of writing seek to assess the ability in writing "to make effective use of varied, complex aspects of language proficiency in a purposeful manner [while] providing, for the purposes of assessment, direct evidence of individual students' language performance" (Cumming, 1997, p. 51). Weir (1990) looks at the testing of writing and comments that two different approaches can be taken: the assessment of writing can be indirect and divided into discrete levels (such as grammar, vocabulary, spelling, and punctuation) and tested objectively, and it can be direct, through extended writings tasks, and tested subjectively. He suggests that writing be tested directly, and maintains that "the writing component of any test should concentrate on controlled writing tasks where features of audience, medium, setting and purpose can be more clearly specified" (p. 73). Weir also discusses the great importance of reliable

marking schemes; he compares two approaches to marking, analytical and general impression; looks at multiple marking; reviews holistic scoring; and considers factors which may contribute to the reliability of a writing test. Hughes (1989) agrees with Weir in stating that "the best way to test people's writing ability is to get them to write" (p. 75). He also notes the importance of reliable scoring and looks at the analytical and holistic approaches to marking. Weir and Hughes are important in that they both advocate the direct testing of writing and place a great amount of emphasis on the reliability of marking schemes.

Language Test Creation, Methodology, and Practice

Up to this point, the four types of test have been examined briefly, the placement test has been defined, and the components of a second language test have been surveyed. Language test creation, methodology, and practice will now be discussed, as these points are crucial in the analysis of any language test.

Weir (1990) looks at the implications of the communicative approach in terms of language testing and examines discrete point, integrative, and communicative approaches to language testing. He details such terms important to any discussion of testing as reliability, validity, and efficiency, and makes a point about the concept of face validity, an important concept in the discussion of any type of test. Weir's text is important for two reasons. First, he covers the design, development, operation, and monitoring of tests, all of which are important in the analysis of testing and in the creation of specifications for accurate placement tests. Second, he makes a strong case for communicative testing by reviewing the deficiencies of discrete-point and integrated testing.

Hughes (1989) looks at testing from a language teacher's perspective. Like Weir (1990), he reviews terms and offers suggestions on the design and development of tests. Although looking at testing from the view of a language teacher would seem at first glance to be incompatible with the purpose of this thesis, Hughes does make some valid observations. First, he notes that "very often, [tests] fail to measure accurately whatever it is that they are intended to measure. Teachers know this. Students' true abilities are not always reflected in the test scores that they obtain" (p. 2). Second, he identifies test content and testing techniques to be sources of inaccuracy. Third, he reviews test techniques for testing overall ability, which placement testing seeks to assess. Finally, Hughes offers some practical advice on how a placement test might be designed. Unfortunately, although his advice contains ideas that are of value to anyone endeavoring to create a language test (his stages of test construction are particularly useful), his example of general procedures for the construction of a placement test "for a commercial English language teaching institution" (p. 55) contains recommendations for format (cloze tests and partial dictations) that alone could hardly contribute to accurate placement.

Steps and Recommendations in Placement Test Creation

Steps and recommendations in placement test creation will now be discussed, as there already exist suggestions or advice that may be applicable to the creation of an accurate placement vehicle.

Hughes (1989) looks at test creation from the teacher's perspective and discusses important concepts in testing, such as the kinds of tests and testing, marking, and validity and reliability. Hughes' text is particularly valuable in that he offers practical suggestions on testing writing, oral ability, reading, listening, and grammar and vocabulary, and couples these with

useful examples. Harrison (1983) also looks at test creation from the teacher's point of view and explains basic principles and concepts in testing, different types of tests, marking, and procedures to help interpret scores and the efficiency of the test itself. While his book is now a little dated and more a practical survey of different types of tests (placement, diagnostic, achievement, proficiency) and things related to them (marking, statistics, and so on), it is useful in that it offers a valuable list of specifications for a placement test, and a commentary on them. This list, which contains information on objectives of the placement test, the skills to be tested, the content of the test, its format, rubrics to be used, materials, and marking, could be used as a guide in the construction of a list of recommendations for other language institutes or programs to follow.

Carroll (1980) aims to "outline principles and techniques for specifying the communicative needs of a language learner and for assessing his language performance in terms of those needs" (p. 5). In order to do so, he gives suggestions on the design and development of communicative tests and their operation, suggestions which are useful because they can then be worked into a set of steps and recommendations that one might use in creating a more accurate placement test. His text is useful for another reason; he suggests, in the design phase, describing the test-takers and analyzing their communicative needs, and he offers advice on how to go about doing so. This description and analysis of needs should prove to be helpful in the creation of an accurate placement instrument.

OTHER CONSIDERATIONS

Issues Related to Accurate Placement

Accuracy in placement testing is important, and a variety of researchers have been faced with placement tests that have appeared inadequate to the task of placing test-takers appropriately. For example, Chandavimol (1988) describes the implementation of a placement test at Chulalongkorn University in Thailand and in doing so, makes some important points. First, the author emphasizes that "accurate placement is essential" (p. 3), as the result of what the author refers to as "misclassification" (p. 3) can be student-related (e.g., misplacements and their consequences) or program-related (e.g., inadequate numbers of instructors). Second, Chandavimol looks at the testing instrument in terms of its ease of administration. The author states that "the efficiency, or the lack of it, of the test greatly depends on administrability, which depends, in turn, on a number of factors" (p. 4), one such example being the clarity of the test's instructions: "the directions must be presented in simple, uncomplicated and unequivocal language that all examinees can understand" (p. 4). In other words, miscomprehension or lack of comprehension of directions on behalf of the test-takers could lead to their not performing to the best of their abilities on the placement test, perhaps resulting in misplacements. Finally, Chandavimol points out that the quality of assessments arising from placement tests "directly depends on the time and the human resources an institution can commit to the task of grading and double-checking" (p. 4), and, although not mentioned by the author but equally as important, to the creation and implementation of an accurate placement instrument.

Further on the issue of the importance of accurate placement of students in class levels, Brown (1989) tells of his noticing that students who were placed into existing classes were

different in level from those who had been promoted from lower level courses. Noting that the type of test most often used in placement testing, the norm-referenced test (NRT), "may not necessarily measure what is being taught and learned in the courses" (p. 73), Brown outlines what he calls "a completely new strategy for constructing language placement tests" (p. 73): the combination of the useful characteristics of a criterion-referenced test (CRT) with those of a NRT "to create placement tests that not only spread students out along a continuum of language abilities (NRT), but do so on the basis of items that are demonstrably related to what the students learn while in the program (CRT)" (p. 73). After reviewing the development of a reading placement test that was based on the new strategy and intended to replace the existing test, Brown examines the item and descriptive statistics of the two forms of the test and seeks to discover how reliable the two tests were and to what degree they were valid as tests of ESL reading comprehension. He does so through the use of a pretest-posttest study involving two groups of foreign students: the first group comprising 194 incoming students required to take the initial placement test, the second group, a subset of those, comprising 61 students who were placed into the reading course. Brown found that the item statistics indicated that a revision was possible and practical, that the revised version was effective as a norm-referenced reading placement test, and that the revised version was valid in terms of its construct validity. While a criticism of this paper is that a better assessment of the revised version of the test may come from an analysis of more than one administration of the test, and while Brown acknowledges that the revision "is just a beginning" (p. 79), the study nonetheless highlights the issue of accurate placement through Brown's devising a test that is closely related to what is taught in the program.

Another study which focuses on the issue of the importance of accurate placement of students in class levels is Rich (1993), who describes the addition of a writing sample to entry-level testing. Rich tells of a college in Florida, where it was felt that the multiple-choice format of the entry-level placement test was not accurate in properly placing students into highly important college-preparatory English courses. Consequently, research was conducted to discover whether the addition of a writing sample would improve placements, how the addition could be accomplished in terms of practical matters (time, location, scoring, and so on), and whether the test-taker's score on a reading subtest might be useful for more accurate placement if the writing sample did not improve placement. In order to answer these questions, all students in what seems to be three levels of college-preparatory English classes were required to produce a writing sample, which was then sent to the College Board to be scored. In answer to the first question, the addition of a writing sample was found not to improve placements: "approximately 85% of all the writing samples scores indicated that students were properly placed. Very few students (29 of 1,399, or 2.1%) in the group received a judgment that they should placed in ESL . . . . Only 7% of the students received scores indicating a lower English course was needed" (p. 13). In answer to the second question, it was found that logistics were a problem: writing demands a place to write, a desk or table on which to write, and so on. In addition, writing creates essays to be marked, and it was determined that there was no way to mark the thousands of resultant essays; also, there was no determination as to who should mark them, and it was impossible to mark them without a week's turn-around time—an amount felt to be much too long. In answer to the third question, the scores on the reading and writing subtests were combined to test the effect of adding the reading placement score to the regression equation

which already contained the writing placement score. The combination was found not to be entirely useful:

> "In searching for a combination of reading and writing scores which could be used in placing students, one situation arose which made the combinations ineffective. In some cases where reading scores made significant contributions, the writing scores tended to cluster around the cut scores. This clustering was not consistently present. Thus before reading scores are placed in combination with writing scores, the cut scores for writing need to be reevaluated" (p. 16)

This study has its problems; for instance, the author mentions that at the college, "there are two levels of college preparatory course work in English, ENC0002 and ENC0020" (p. 2), yet she states that the writing test was administered to all students in ENC0002, ENC0020, *and* ENC1100, which is presumably either a third level of college preparatory course or a college course—whichever it is simply remains unexplained. Furthermore, Rich does not indicate how the writing topics were chosen nor who chose them, and she states that "the set of topics varied from campus to campus" (p. 4). This variation is a problem, as the topics may also vary in degree of difficulty. In spite of its problems, Rich's study nevertheless illustrates the issue of accurate placement through looking at crucial placement decisions at a college and the attempt to improve the placement vehicle through the addition of another type of test. Rich's study also demonstrates that what may seem to be a placement problem alone may in fact be a result of problems with the system that need to be addressed first.

Ilyin (1970) also notes the importance of accurate placement of students in class levels. She tells of the large numbers of students of varying abilities in each class of an adult ESL program in San Francisco and states that she "was surprised at the utter chaos that existed in our classes" (p. 1). After discovering that the program would have to develop its own placement test, she describes the development of a standardized placement test to place the students into the first three levels of language classes and discusses work done on an experimental test to place students in the last three levels of language classes. In addition, she reports on a subsequent study to set norms and to investigate gains, but unfortunately, because the paper is outdated, the results are really of little use—after all, advances have been made in testing in the past three decades, and the use of discrete-point testing alone, as is suggested in Ilyin's paper, has been supplanted by other forms of testing (Weir, 1990). In spite of that, the paper underlines the importance of accurate placement. In Ilyin's words,

> we have one placement test for our lower levels that has been standardized and which has a high reliability . . . that place our students in classes better than previous methods. We still have to move a student or two, but not the large numbers of students we did before. The morale in the school is better. In short, both students and teachers are happier when placements are made more accurately." (p. 14)

Malu (1989), in her outline of ESL entrance testing and course placement procedures for the ESL program at the United Nations International School for grades 6-12, approaches the importance of accurate placement of students in class levels from a different perspective. Before doing so, however, she explains the procedure for determining whether a student needs to be

tested for ESL; explains testing procedures, test, and interviews; and describes course placement and procedures. The difference is that instead of focusing on the placement test alone and its accuracy, Malu pinpoints the initial heavy investment of time taken during the placement test (as much as three hours in some cases) as having the beneficial result of "minimal class switching because of misplacement" (p. 211). In terms of the thoroughness with which the program conducts its testing, placement decisions are not made on the basis of test results alone; rather, many other factors play a role, such as a holistic reading of an essay written as part of the placement test, and the student's background and behaviour during the test and the interview. While many other programs simply may not have the time to invest in a thorough assessment of their students during placement testing, and while Malu herself admits that "the major difficulty apparent to all who participate in this programme is the amount of time it takes to implement [the] procedures" (p. 211), Malu stresses the importance of accurate placement through a thorough testing procedure.

LeBlanc and Painchaud (1985), in their discussion of self-assessment as a placement instrument, also approach the importance of accurate placement from a different perspective. Rather than focusing on the refinement of a traditional testing method, they look at a technique typical of alternative assessment (as defined by O'Malley and Pierce, 1996): planned self-assessment. They maintain that accurate placements may be achieved through the use of questionnaires and that self-assessment testing can be a viable alternative to standardized testing. The authors describe a research project which aimed to answer the questions of whether students registering for second language courses at the university in which the authors worked could assess their own language proficiency, whether the type of self-assessment instrument used could influence the quality of this assessment, and whether self-assessment could be used as a

placement instrument. To answer the first question, they randomly selected 200 students, who completed self-assessment questionnaires prior to taking the university's proficiency test. Afterwards, correlations were drawn between self-assessment and proficiency test scores, and it was found that students could indeed assess their own knowledge to some degree. To answer the second, they gave two forms of the self-assessment questionnaire to students taking part in the fall registration at the university; one included metalinguistic vocabulary, the other did not. Correlations between the two forms were drawn, and it was found that the format seemed not to have any bearing on the quality of the answers. To answer the third, they revised level descriptions of second language courses, enlisted teachers to contribute "representative descriptors for each of the six levels in both listening and reading" (p. 683), and prepared a questionnaire based on the result. Afterwards, they tracked the percentage of level changes as the result of misplacement using the proficiency test in one academic year and the self-assessment test the next, and they found that "the self-assessment results placed the students at least as well as the standardized tests previously used" (p. 684). In fact, in all sessions, self-assessment seemed to have placed the students better than the standardized tests; level changes dropped between 1.5 to 3.7 percent from one year using the standardized tests to the next using the self-assessment placement test.

Although LeBlanc and Painchaud maintain that their students have the ability to assess themselves, and while all of their correlations are statistically significant at the .05 level, the correlations are, the authors admit, "not of the highest level" (p. 679), the one for the self-assessment speaking test, for example, being as low as .39. Also, the authors maintain that the format of the questionnaire did not matter as long as "students can understand the language used in the questions" (p. 682) and as long as the questionnaire was well-constructed. This particular

group of students seemed to have little difficulty with the language of the questions in this case. However, new groups of students and their abilities may be different from those who come before them. What might happen if another group does have difficulty? Furthermore, while the authors state that the questionnaire was well-constructed, they give only a brief account of its construction, never really explaining what they mean by "well-constructed." Finally, although level changes decreased by a few percent, these changes may have been made for reasons other than those related to misplacement—a fact the authors do admit—and the decrease itself was not examined to find out whether it is statistically significant. In spite of these shortcomings, LeBlanc and Painchaud suggest that self-assessment may be an accurate and valuable alternative to traditional placement instruments.

The Ethics and Effects of Testing

So far, testing has been observed from the point of view of the placement test itself: its definition, its component parts, its creation, and its accuracy in a variety of placement situations. No literature review would be complete, however, without a discussion of the ethics and effects of testing.

Shohamy (1993) argues that "few devices are as powerful, or are capable of dictating as many decisions, as tests" (p. 1). She maintains that

> results obtained from tests have serious consequences for individuals as well as for programs, since many crucial decisions are made on the basis of test results. Among these are the placement of students in class levels, the granting of certificates or diplomas, determinations as to whether students are capable of continuing in future

studies, the selection of students most suitable for higher-education institutions, and the acceptance of job applicants and program candidates." (p. 2)

Shohamy holds that rather than providing information, tests have become tools for power and control, and she provides three examples in support of her argument: the first, the impact of the introduction of a test of Arabic as a second language; the second, the impact of a new EFL oral test; and the third, the impact of a reading comprehension test. Shohamy lists a number of findings, the most significant being those which follow. She finds that all three tests had some type of impact (as defined by Wall, 1997), and that the impact was complex and dependent upon the nature and purpose of the test. She notes, too, that the implementation of the tests caused instruction to become testlike, in other words there was "backwash," which is also known as "washback," (both terms as defined by Wall). She observes that the strength of the impact of these tests varied, depending on the type of test, subject relevance, the failure rate, and so on. Shohamy's paper is an important reminder that there is more to the creation of tests than just the simple assembly of test items: "tests are powerful devices and should be treated as such" (p. 18). Responsibility in test creation must begin somewhere, so "testers need to examine the uses that are made of the instruments they so innocently construct" (p. 19).

Responsibility in test creation could also be seen as a form of accountability. In reviewing and detailing the problems inherent in and development of accountability in language assessment, Norton (1997) points out that "language assessment practices should be accountable" (p. 313) not only to test-takers, who have been considered to be powerless stakeholders in the field of language assessment, but also to systems, because "schools, colleges and universities are under pressure to inform the public about what they are teaching and how

effective they are" (p. 317). Accountability is an important consideration, and test-creators must think carefully about a number of matters: the uses to which their tests will be put; appropriate training in testing and test use; and "recognition that test takers come from heterogeneous, culturally diverse backgrounds that must be taken seriously in the assessment process" (p. 314).

A good example of accountability is Peirce and Stein (1995), in which the authors describe their piloting a reading passage that was intended to be used as part of a South African college entrance examination for Black students. Because of concerns regarding the violence present in the then-current political climate, it was decided to pilot a reading passage to be given to Black students in a Johannesburg secondary school, for the reason that "if test takers became unduly disturbed by the content of the test, their performance might be compromised" (p. 54). It was found that the passage was interpreted as racist and was therefore rejected. Peirce and Stein's paper is a good example of accountability in that the authors were responsible to the test-takers and considering the possible results the test might have, piloted the passage before it was put into use.

CHAPTER III. METHODOLOGY

This chapter first restates the three research questions. In order to provide necessary background for the rest of the discussion of methodology, it then describes the site and the VELI testing narrative, that is, a description of the VELI placement test and testing process. After that, the chapter outlines the design and methodology used for the purposes of this thesis, beginning with an explanation of the qualitative design known as evaluative case study, and continuing with a chronology of the various methods—or multimethods (Schumacher & McMillan, 1993)—used in the collection of data. The chapter continues with a description of the participants in the study, and ends with a summary in point form which outlines the research questions and the methodology used to examine them.

RESEARCH QUESTIONS

1. What are the historical and institutional influences and limitations that have shaped the VELI placement testing process?
2. On what basis do those concerned think that the test is flawed, and how widespread is the problem?
3. Based on the findings of the previous research question, where might problems with the institute's placement testing and the resulting placements lie?

THE SITE

In order to understand the methodology used throughout this thesis, it is necessary to describe first the setting in which the research took place, and then the VELI placement test and

placement testing process. The setting is the Vancouver English Language Institute (VELI), a school which occupies the entire third floor of a downtown Vancouver office building and which caters to adult foreign students who come to the institute from around the world. An accurate, if brief, description of the institutional nature of the school can be found in the VELI promotional brochure, which states that

> [The] premises are new and modern. The [16] classrooms are large and comfortable, equipped with whiteboards and air conditioning. Each class has its own textbooks, dictionaries, reference books, tape recorders, listening centres and VCR's for students to use whenever necessary.

Although it welcomes new students and tries not to turn any away, VELI limits its enrollment for reasons related to the philosophy of the school as stated below. As a result, not more than approximately 35 percent of the total number of the student body, which fluctuates between 300 to 450 students, is composed of any one language group. Consequently, classes contain a variety of nationalities and cultures, with students from Taiwan, Japan, Korea, Spain, Switzerland, France, Brazil, and Mexico, to name a few. The diversity of the VELI student body is unlike that of many of the private language institutes in Vancouver, which appear to concentrate on one or two language groups for their student body, seemingly because of reasons related to world economies and school finances. Thus, at these schools, it is not unusual to find classes made up primarily of students from one or two financially-powerful countries.

According to the promotional literature of VELI, students, who must be over 18 years old and have a minimum of a high school education, study at the institute to improve their English

for business or career reasons or for admittance into colleges and universities in North America. They are expected to be mature and responsible, to make progress during their program of study, and to attend classes regularly. Those who cannot fulfill these expectations may be asked to leave the institute. According to the VELI Instructors' Manual, the institute, which was founded in 1988 and now consists of a faculty of up to 30 teachers and a student body of approximately 350 students, places particular emphasis on "high standards of teaching and learning, innovative materials and methodology, and excellence in student service." The stated philosophy of the institute is to provide a quality EFL education to motivated adult international students for both short- and long-term periods, to "encourage [the] meeting of new friends and [the] mixing of different cultures," and to "foster an oral/aural approach to English language learning" through the policies and curriculum of the school. A particularly important policy is a strict English-only rule: all instruction regardless of level is carried out *exclusively* in English with no translation allowed (either teacher-to-student or student-to-student), and students *must* speak English everywhere within and around the school. Those who disregard this rule are subject to severe punishment: teachers may be disciplined (admittedly, a rare occurrence) and students may be dismissed, either from classes for the day if it is their first infraction or from the school if it is their third. With regard to the curriculum, it places a great deal of importance on a communicative approach to language learning, and listening comprehension and oral production in the form of audio tapes, video tapes (of movies, television programs, etc.), role play, discussions, debates, games, and so on play major parts. Reading and writing, although considered to be important, occupy lesser, complementary roles in the curriculum—to the point that some teachers feel uncomfortable if their classes are too quiet for a time because their students are occupied with a fair amount of reading or writing. In contrast, grammar also plays a

major part: there are classes that focus on grammar (e.g., the Grammar and Accuracy class, the Grammar and Writing class), and the overt teaching of grammar is carried on to some degree in all general classes. The reason for this dichotomy—communicative approach coupled with the teaching of grammar—is one of market forces; the early VELI curriculum did not allow for the teaching of grammar, so teachers avoided it, and the students and the agents who send students to the school complained—after all, a "serious" school such as VELI should have their students do more than just "sit around and talk." Since that time, VELI has incorporated grammar into its curriculum.

THE VELI PLACEMENT TESTING NARRATIVE

In order to understand further the methodology used throughout this thesis, it is also necessary to describe the VELI placement test and the placement testing process. This description will be divided into two sections. Detailing the VELI placement test, the first section examines the three parts of the test. Investigating the process through which the VELI placement test is administered, the second section explores the process of placement testing at VELI through the recounting of two perspectives: that of the test-takers and that of the teachers, who participate in the placement testing process from beginning to end.

Description of the Test

Placement testing of students new to VELI occurs on the first Monday of each month barring statutory holidays, since the institute has monthly intake—that is, students can begin their programs at the beginning of any month. The placement test is administered to as many as 150 students at one time, and for the record, one month (known as a "school month") at VELI

refers to a four-week period; in the course of a year, there are 12 four-week periods for a total of 48 weeks.

As can be seen in Figure 1, the test comprises three parts: the first, focusing on grammar, vocabulary, and reading, is divided into seven sections timed at 30 minutes and is entitled the "Entrance Evaluation" (EE); the second, focusing on writing, is divided into two sections timed at 8 minutes; and the third, an oral interview, has no time limit.

Figure 1

The Three Parts of the Placement Test

Part One	Part Two
Grammar Conversation Vocabulary More Grammar More Grammar Cont'd [sic] More Vocabulary Reading Comprehension	Writing, part 1 [guided sentence writing] Writing, part 2 [guided paragraph writing] **Part Three** Oral Interview

Part One

The first part of the placement test, the EE, appears in a booklet comprising five double-sided sheets of 8½-by-11-inch paper that are stapled together, making it a ten-page test. The answer sheet, which is separate from the test booklet and on which students circle the letter of the answer, is also printed on an 8½-by-11-inch paper. To return to the booklet, the first section of the first part of test, "Grammar," consists of five multiple-choice, fill-in-the-blank questions which focus on verb tense, verb aspect, word order in indirect statements, determiners, and

object pronouns followed by infinitives. The second section, "Conversation," consists of five sentences which function as the first part of a conversation, each followed by a choice of four answers which function as the next line of the conversation. The third section, "Vocabulary," consists of ten sentences in which one word or phrase is underlined, each sentence being followed by a choice of four words or phrases of which one means the same thing or something similar to the word or phase underlined. The fourth section, "More Grammar," consists of five error-recognition sentences similar to those found in the Written Expression section of the Test of English as a Foreign Language (TOEFL). The fifth section, "More Grammar Cont'd [sic]," consists of five multiple-choice, fill-in-the-blank questions of the same type found in the first section of this part. The sixth section, "More Vocabulary," consists of five sentences of the same type found in the third section of this part. The seventh section, "Reading Comprehension," consists of four reading passages, three of which are long (approximately 120 to 140 words), and one short (32 words), followed by multiple-choice comprehension questions. These texts are not authentic; three of them, however, are written in a style reminiscent of that of a popular newsmagazine, and one (the shortest) is written in that of a newspaper.

Part Two

The second part of the placement test appears on the side of the paper opposite the answer sheet. The first section of this part, appearing under the heading "Entrance Evaluation 1: Writing" and occupying the top half of the page, consists of six base-form verbs, which the test-taker must convert to the past tense and then write a sentence containing that verb. The second section, also appearing under the same heading as the first section of the second part and occupying the bottom half of the page, consists of a strip of five cartoon-like frames along the

left side of the page. This strip, however, does not tell a story; the frames contain drawings pertaining to weather phenomena. To be specific, the top frame contains a drawing of a woman wearing a hat, caught in a heavy downpour. The second frame contains a dark drawing of an angry-looking thundercloud with two bolts of lightning striking out beneath it. The third and fourth frames, side by side and occupying the same amount of space as one of the previous frames, contain respectively a drawing of a man wearing a hat, clutching his hands together in front of him, looking up at the snow that is falling on him; and a drawing of a man caught in a gust of wind, his umbrella almost torn from his grasp and being turned inside out, and his hat being blown off his head. The final frame contains a drawing of a smiling man wearing only a bathing suit and a hat, his arms outstretched while he basks in the sun. These frames are accompanied by directions to the test-taker to "write a paragraph about the weather" followed by a question related to the weather and some guidance in how to go about doing so: "How has the weather been since your arrival in Vancouver? Compare it to the weather in your country this month."

Part Three

The final part of the placement test is in the form of an oral interview, which is administered by a teacher after only the first part of all the tests that day have been marked (the second part containing the writing is not marked) and sorted. The oral interview consists of a meeting in an empty classroom between the test-taker and a teacher (who functions as an interviewer and, as stated earlier, to be known as the "TI" for the purposes of this thesis), who spends some time interacting with the test-taker—usually by asking questions for the test-taker to answer. These questions are not standardized. There are, however, three optional interview

aids available to the TI: a black-and-white drawing of a busy street corner city scene, a cartoon depicting a busy beach scene, and a list of questions based on grammar and divided into ability levels. In addition, a chart called the "Entrance Evaluation and Placement Recommendations" is available to the TI. It is given to every TI on registration day and contains criteria which the TI may use in order to place test-takers. Use of the chart is not mandatory: TIs may choose to follow its guidelines—or not. Some do. Some do not. Concerning the chart itself, it consists of rows of ranges of scores and columns outlining the type of writing to be expected from a test-taker falling within that range, a recommendation of level, and the type of "conversation" during the interview that can be expected from a test-taker falling within that range.

In any case, once the TI has heard enough of a sample of the test-taker's oral ability, the TI decides on the appropriate level for the test-taker and ends the interview by telling the test-taker into which level the TI will place him or her. Because training in interviewing technique is minimal (one teacher mentioned that she "had no training at all"), TIs assess test-takers based almost entirely on the TIs' experience and expertise teaching at certain levels of instruction. As a result, prior to the interview, test-takers are assigned to certain TIs based on the test-taker's performance on the first part of the VELI placement test. That is, test-takers who receive low scores on this part of the test are interviewed by TIs who usually teach beginner classes; those who receive middle scores are interviewed by TIs who teach intermediate classes; and those who receive high scores are interviewed by TIs who teach advanced classes.

Description of the Process of Placement Testing

As detailed previously, the placement test, given on registration day to as many as 150 students at a time, comprises three parts. Part one focuses on grammar, error recognition,

vocabulary, and reading comprehension. Part two focuses on writing. Part three focuses on speaking. The process through which the VELI placement test is administered can be described from two perspectives: that of the test-takers and that of the teachers.

The Test-Takers' Perspective

From the test-takers' perspective, upon entering the institute for the first time on what is known as "registration day," the test-takers are ushered into a common area of the school, "the student lounge" (in reality, a large, brightly-lit, L-shaped cafeteria-like room that can seat around 150 people), where they undergo what Brown (1989) has described as the "cold and detached experience" (p. 65) of taking the placement test. Once in the lounge, the test-takers begin to sit down in more or less random places, six to eight per table. At this point, there is much noise: some of the latecomers are finding their way to their seats, moving chairs and tables; some of the more adventurous types are speaking to one another, introducing themselves—in English; and still others are rummaging through their bags and briefcases, looking for pencils, pens, and paper. While the testing situation at this school is analogous in many ways to that of Malu (1988) in her description of the process of entrance testing and course placement for the ESL program at the United Nations International School, the commotion at the institute in Vancouver is certainly the opposite of what Malu has described as a "small relaxed setting" (p. 207).

In any case, after a while, the amplified voice of one of the directors or administrative assistants cuts through the din, and truly, the students begin their testing experience by listening to one of the directors or administrative assistants review important information regarding the school such as school rules, what to expect during the testing process, what to expect in class, and so on. Soon after this introductory speech, the test-takers receive a test booklet and an

answer sheet (as described above) and receive instructions on how to complete the test. Unlike Malu, whose "students work through the test at their own pace and extra time is given, if need be, to ensure all students understand the test directions" (p. 207), the test-takers at VELI, regardless of whether they have understood the directions, get down to the business of working on the first part of the test by marking on their sheets answers to multiple-choice questions within the time allotted. For thirty minutes, a somber silence falls over the student lounge, interrupted periodically only by the soft sound of erasers rubbing or by pages turning. Again, unlike Malu, who states that if "a student is overwhelmed by the test, testing is stopped" (p. 207), because of the sheer number of test-takers involved in the testing (anywhere from approximately 70 to 150 at one time), test-takers who cannot answer the test simply sit and wait for the half-hour to end. At the end of the thirty-minute time limit, the test-takers are directed to turn over their answer sheet, in order to continue working on the second part of the test; then, for another eight minutes, the test-takers quietly work on the test, this time writing sentences and a short paragraph. At the end of this time limit, the test-takers (often reluctantly) hand their answer sheets to a director or an administrator for marking, and then participate as best they can in an oral information-gap exercise with the other test-takers (essentially, a filler exercise to occupy the test-takers) while waiting for the last step in the placement process: the interview.

During the information-gap exercise, after some time has passed, one of the directors or administrative assistants starts to call out numbers. These have previously been assigned to the test-takers so that they can be summoned for their interview. Once the number of a specific test-taker is called, that person is sent to a classroom, outside the door of which, in the hallway, is a chair. The test-taker sits passively in that chair in the hall along with all of the other test-takers whose numbers have been called and who are sitting in *their* chairs outside the doors of the

classrooms where *they* will have *their* interviews. Once the TI is ready to receive his or her first interviewee or has finished a previous interview, the test-taker is conducted into the classroom and is invited to sit near or across from the TI. The test-takers are often nervous at this point and many are jet-lagged, having arrived in Vancouver only a day or two before, many from overseas; thus, even those who have a good command of English often stumble through their interview. Depending on the ability of the test-takers, after answering a variety of Yes/No questions and information questions, they are informed in which level they have been placed and are told some basic information about reporting to the school the next day. At that point, the interview is over and rather than immediately attending classes, the test-takers are told that they have finished for the day and to return the next day for classes.

The Teachers' Perspective

From the teachers' perspective, "registration day" begins at approximately 9:00 a.m., when the teachers start to arrive at the institute. Not all teachers participate in every registration day; teachers are scheduled to participate in alternate months, with approximately one-half of the faculty attending registration day while the other half has the day off. Because the testing process for the teachers does not really begin until 9:30 a.m., they congregate in the "teachers' room," a converted classroom containing a large, industrial-strength photocopier immediately apparent upon entering the room, a blue recycling bin for discarded paper, shelves of texts along one of the walls of the room, a number of file cabinets, shelves along the other side of the room containing videos and tapes, and a refrigerator and two computers at the far end of the room. In the middle of the room sit two long tables placed side by side, around which are found ten or twelve chairs.

From approximately 9:00 a.m. to 9:30 a.m., the teachers prepare for their classes in the upcoming weeks. A serious air falls over the room, the photocopier in constant use. Papers are shuffled and then shuffled again. Some teachers confer with others about their classes: the grammar to be reviewed, the themes to be discussed. Other teachers have their heads down in their books, their fingers furiously flipping through pages. Still others are elbow-deep in the filing cabinets, fishing for file folders containing the contents of a lesson plan. At around 9:30 a.m., the time for the teachers to begin the testing and interview process, the tone of the room changes: the photocopier is still, the books are closed, the file cabinets are shut. Some light conversation begins as the teachers leave the teachers' room and move to Room 3.

Room 3 is simply a classroom in which the teachers congregate to perform the functions necessary to start off the testing and interview process. After sitting down around four tables placed end-to-end in a square, the teachers settle down to listen to one of the directors, who begins by reviewing placement information important to the teachers. This information consists of classes that are full, time-slots that are available or unavailable, courses that are offered or not offered for the coming month, and so on. After a while, the answer sheets from the EE are brought into Room 3, and the teachers get down to marking the multiple-choice section of the test. The teachers select, at this point, which "job" they would like to do: "marker" or "counter." Markers assess the tests using a scoring guide, which is overlaid onto the answer sheet. After marking the wrong answers with a felt pen, the markers then pass the tests to the counters, whose job it is to count the number of mistakes and then write the number of correct answers at the side of the answer sheet. Once all of the tests have been marked, matched to registration cards previously filled out by the test-takers, and sorted into approximate levels (i.e., beginner, intermediate, advanced) based on the number of correct answers, the teachers receive

six to eight tests-plus-registration cards and then head off to the classroom where they will conduct their interviews.

Upon arriving at the classroom, the TIs sit and prepare for the arrival of the first interviewees. When the test-takers arrive, the TIs welcome them to the classroom and then ask them to sit down. The first questions the TIs ask are usually basic in order to gauge the approximate level of the test-takers and to judge whether to continue with the interview in the same vein or whether to adjust to an easier or to a more difficult mode of communication. At this point, some TIs use the interview aids provided, for example, asking the test-takers to talk about what they see in a picture. After asking a variety of questions and, depending on the abilities of the test-takers, after some small talk or open-ended light conversation, the interviewers fill out a form indicating the score on the EE; the level (from one to eight) of the test-takers in terms of their writing skills (as demonstrated solely in the sample found in the second part of the placement test), their listening comprehension, their speaking skills (which is broken down into vocabulary, structure, and fluency); and their confidence level (low, medium, or high) during the interview. As explained earlier, TIs assess test-takers based almost entirely on the TIs' experience and expertise at teaching certain skill levels. Placement guidelines do exist and are found in the VELI Registration Day packet—as mentioned previously, the chart called the "Entrance Evaluation and Placement Recommendations" is available to the TI. However, informal interviews conducted for this thesis between the researcher and TIs revealed that the placement guidelines for the most part are not very useful and do not figure much in the TIs' ultimate placement of the test-taker. To return to the explanation of the process, the TIs also indicate the recommended level from one (lower beginner) to eight (very high advanced), along

with an assessment of where the test-taker "fits" in that level: low, medium, or high, and, if warranted, in which recommended specific skill classes the test-takers would like to study. The last step in the process for the TIs is to sign the bottom of this form, inform the test-takers at which level they have been placed (and often, why they have been placed there), and give the test-takers further information about classes the next day: what time to arrive at the school, how to find the classroom in which to study, and so on. At this point, the interview ends, and the test-takers leave the room. The TI immediately follows, going out into the hallway and welcoming the next test-taker into the classroom. For the TI, the interview process begins anew. Welcome, interview, good-bye: the TIs process the test-takers one after another, the entire procedure taking approximately two-and-a-half to three hours.

DESIGN

A qualitative design known as evaluative case study was used in this thesis to describe and analyze the placement testing process, as "case study designs are appropriate in evaluation studies when the program or innovation must be systematically studied" (Schumacher & McMillan, p. 377). While "a great deal of debate, misunderstanding, and confusion currently surrounds the use of qualitative research methods in the applied linguistics field" (Davis, 1995, p. 427), for the purposes of this thesis, qualitative design will refer to that type of method which "[investigates] behavior as it occurs naturally in noncontrived situations, [with] no manipulation of conditions or experiences" (Schumacher & McMillan, p. 37). The behavior that will be described in this thesis is natural and non-contrived in that it occurs during the placement testing process at VELI. In addition, because this thesis describes the placement testing process in its natural, noncontrived form, there will be no manipulation of conditions or experiences.

METHODS

Schumacher and McMillan (1993) note that multimethods are useful in evaluative case study. The term "multimethods" is defined by the authors as "multiple strategies to corroborate the data obtained from any single strategy and/or ways to confirm data within a single strategy of data collection" (p. 416). Multimethods can include the use of a wide variety of techniques; those used in this thesis, mostly in the order in which they were employed, were document analysis, to examine the historical and physical influences; questionnaire, to determine the basis on which the administration and faculty thought that the test was flawed and how widespread the problem was; participant observation, to examine historical and physical influences, to determine the basis on which the administration and faculty thought that the test was flawed and how widespread the problem was, and to determine where problems might lie with the institute's placement testing and the resulting placements[2]; interview, to determine the basis as noted above; and further document analysis, to determine where problems might lie.

Document Analysis

Johnson notes that "researchers may gather a variety of other types of data that shed light on the research questions [and] in case studies of programs, such as . . . ESL/EFL programs, researchers usually gather as many documents as possible about the program" (p. 89). For this thesis, documents were analyzed as a first step in order to shed light on the first research question: What are the historical and institutional influences and limitations that have shaped the VELI placement testing process? Documents collected and analyzed for this thesis include a

[2] See the summary at the conclusion of this chapter for a point-form layout of the research questions and the methods used to explore them.

now-unused VELI test booklet, which contains the original VELI placement test; a current VELI test booklet, which contains the VELI placement test used presently; a VELI Registration Day packet, which is issued to all teachers who interview students on registration day; a VELI Instructors' Manual, the document given to all instructors which outlines their daily duties, responsibilities, and so on, at VELI; the VELI Students' Manual, a booklet given to all students entering VELI which contains important information about the school; internal VELI documents pertaining to the scheduling of teachers for registration day; a VELI promotional brochure, a glossy, multi-paged booklet used for advertising purposes; e-mail correspondence between the researcher and the Director of the Toronto campus of VELI; personal communications between the researcher and both the Director of the Vancouver campus of VELI and the Assistant to the Director; and a VELI Level Curriculum, a detailed internal document for instructors which outlines the eight levels of ESL classes at VELI and what is to be taught at each level.

Questionnaire

A questionnaire was devised as the next step in order to answer the second research question: On what basis do the administration and faculty think that the test is flawed, and how widespread is the problem? The use of questionnaires is well-documented as a viable approach to data collection in qualitative research (Banerjee & Luoma, 1997; Brown, 1991; Johnson, 1992; Richards & Lockhart, 1995). The questionnaire used in this thesis (see Appendix A) was distributed at VELI to all 28 members of faculty and 2 members of the administration, to which 18 instructors and 2 administrators responded. It contained questions which sought to discover which part or parts of the placement test and placement process the administration and the

teachers, in their roles of both classroom teacher and registration day interviewer, believed was of great value and of little value in assisting them to assess students and to assign students to levels of instruction at the institute. Specifically, the questionnaire contained seven questions asking whether the respondent was an instructor or administrator; the levels of instruction (lower and upper beginner, etc.[3]) at which the respondent usually interviewed during registration day; how useful the respondent found the first, second, and third parts of the VELI placement test; how many students the respondent believed were misplaced per class taught; and how effective the respondent found the entire placement process. Their responses indicated that the placement test and placement testing process was suspect and that a problem existed in which students were being misplaced into classes after registration day. As a result, investigation was undertaken in the form of participant observation, interview, and further document analysis.

Participant Observation

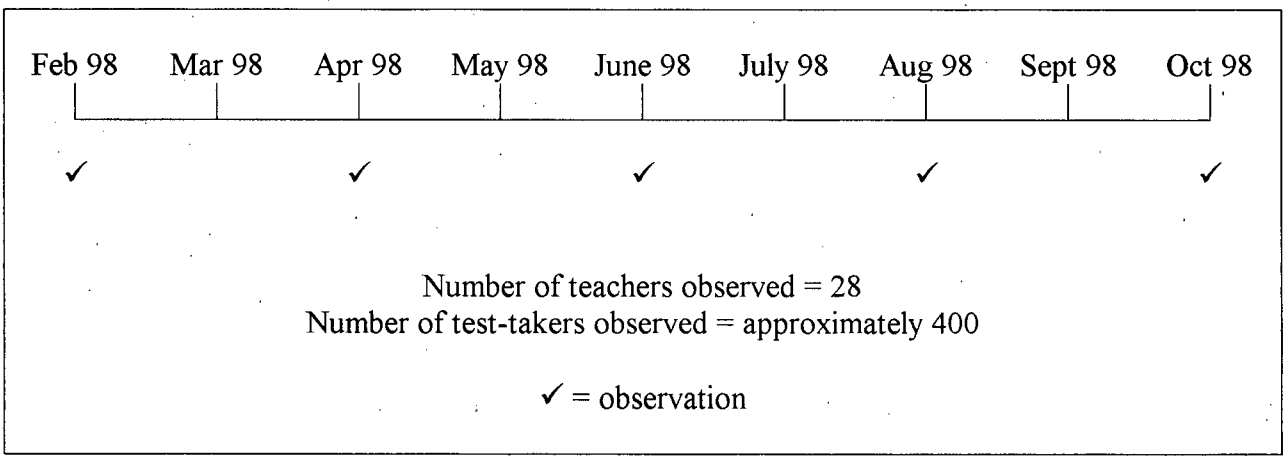
The technique known as participant observation was used as the ensuing step in order mainly to determine on what basis the administration and faculty thought that the test was flawed and how widespread the problem was, and to a smaller part to contribute to the examination of the historical and institutional influences and limitations that have shaped the VELI placement testing process. In addition, participant analysis was used in the analysis of the interview portion of the placement test in order to determine where the problems with the placement test might lie. As Johnson notes, "to try to see reality from the participants' point of view requires that the researcher spend a great deal of time on-site.... [a] long term stay is necessary" (1992, p. 143).

[3] See below for a more detailed explanation of the VELI placement testing process.

The researcher of this thesis has been a qualified ESL instructor since 1981, and has been employed in that position at VELI since its inception in 1988. In the ensuing decade, he has observed the daily goings-on at the school. Over a nine-month period, from February to October, 1998, the researcher observed and kept a log of what went on among approximately 400 test-takers and 28 teachers on five registration days, each at the beginning of February, April, June, August, and October. These observations are summarized in Figure 2.

Figure 2

Observations in 1998



The researcher was an active participant during this time period, participating in the placement testing process at VELI.

Interview

Because of the results yielded by the questionnaire, the technique known as interview was used as the next step in order to determine on what basis the administration and faculty thought that the test was flawed and where might problems with the institute's placement testing

and the resulting placements lie. Interviews may have many forms; the two used for the purposes of this thesis were, in the order in which they were administered, the informal conversation interview and the standardized open-ended interview.

Schumacher and McMillan state that in the informal conversation interview, "the questions emerge from the immediate context and are asked in the natural course of events; there is no predetermination of question topics or phrasing. Informal conversations are an integral part of participant observation" (p. 426). Of the 28 VELI faculty members who had been contacted initially through a letter in which the researcher outlined the research topic and generally requested participation in a questionnaire and interview, 11 instructors (8 female, 3 male) agreed to be interviewed informally at the school during the course of the regular VELI day, especially during registration day or immediately after it. Participants were asked a variety of questions concentrating on the VELI placement process, on the individual parts of the placement test, and on the training and competency of interviewers. These interviews ranged in time from a minimum of approximately five minutes to a maximum of approximately 15 minutes or more. For the most part, however, interviews tended to be short so as not to inconvenience the respondent; after all, the respondents were interviewed in passing, and as noted previously, during their work day—a time when most people are pre-occupied with various duties and therefore have their minds on the task at hand. In any case, answers to the questions were written down in point form by the researcher during the interview, and the responses were read back to the interviewees immediately afterwards (if possible) to check for accuracy. Each respondent and that respondent's pseudonym, gender, age range, and years of experience teaching English as a second language classes are summarized in Table 1, in which respondents are listed according to the order in which they were interviewed.

Table 1

Respondents to the Informal Conversation Interview

Number of Respondents (n=11: female = 8, male = 3)			
Respondent	Gender	Age Range	Years of Experience (ESL)
ICI-1	F	40 – 45	7
ICI-2	F	40 – 45	9
ICI-3	F	30 – 35	5
ICI-4	F	25 – 30	2
ICI-5	M	35 – 40	11
ICI-6	M	50 – 55	17
ICI-7	F	25 – 30	4
ICI-8	F	30 – 35	5
ICI-9	F	30 – 35	4
ICI-10	F	30 – 35	5
ICI-11	M	35 – 40	17

As was the case in regards to the questionnaire, the responses of the participants in the informal conversational interview indicated that the placement test and placement testing process were problematic and that a problem existed in which students were being misplaced into classes after registration day.

Schumacher and McMillan state that in the standardized open-ended interview, "participants are asked the same questions in the same order, thus reducing interviewer effects and bias" (p. 426). Reliable and valid interviewing therefore necessitates the use of a list of questions to aid the interviewer. This list is called an interview guide, in which open-ended questions are pre-specified and which the interviewer follows closely. A brief interview guide was used by the researcher, who as noted previously had contacted all 28 VELI faculty members initially through a letter in which the researcher outlined the research topic and generally

requested participation in a questionnaire and interview. All were later approached in person to be interviewed first informally, with (as noted above) 11 of the 28 faculty members agreeing to participate, and later to be interviewed formally, with 12 (7 female, 5 male) of the 28 faculty members agreeing. As was the case with the informal conversation interviews, these standardized, open-ended interviews ranged in time from a minimum of approximately five minutes to a maximum of approximately 15 minutes or more, with interviews tending to be short so as not to inconvenience the respondent. There was some overlap of participants between those taking part in the informal conversation interview and those in the standardized, open-ended interview; of the 12 participants in the standardized, open-ended interview, 7 had already participated in the informal conversational interview.

In any case, all interviews took place in a setting comfortable for and familiar to the respondents: either in the teacher's classroom or in the staff room. The respondents were asked two questions, the answers of which would contribute to the answer to the second research question of this thesis. The two questions asked were "How do you know when students have been misplaced into the classes you teach?" and "How long after classes have begun can you tell that students have been misplaced into the class you are teaching?" These questions are based on the presupposition, of course, that some students are in fact misplaced into classes. The reason for the presupposition is that, in the researcher's experience and in reference to data culled from the questionnaire (see Table 8), no teacher at VELI has indicated the absence of misplaced students. That is, all teachers have stated that they have had students misplaced into their classes.

To return to the standardized, open-ended interview, answers to the questions were written down in point form by the researcher during the interview; the responses were read back

to the interviewees immediately afterwards to check for accuracy. Each respondent and that respondent's pseudonym, gender, age range, and years of experience teaching English as a second language classes are summarized in Table 2, in which respondents are listed according to the order in which they were interviewed.

Table 2

Respondents to the Standardized, Open-Ended Interview

Number of Respondents (n=12: female = 7, male = 5)

Respondent	Gender	Age Range	Years of Experience (ESL)
SOEI-1	M	35 – 40	8
SOEI-2	F	40 – 45	7
SOEI-3	F	30 – 35	5
SOEI-4	F	35 – 35	5
SOEI-5	F	25 – 30	2
SOEI-6	M	40 – 45	13
SOEI-7	M	40 – 45	8
SOEI-8	F	25 – 30	4
SOEI-9	M	50 – 55	17
SOEI-10	F	25 – 30	5
SOEI-11	F	35 – 40	10
SOEI-12	M	35 – 40	17

As was the case with the informal conversation interview, the responses of the participants in the standardized, open-ended interview indicated that the placement test and placement testing process were problematic and that a problem existed in which students were being misplaced into classes after registration day.

Further Document Analysis

Placement testing at VELI seemed to be problematic—at least, in the eyes of the respondents to the questionnaire and to the interviews. As a result, the final step was to analyze the VELI placement test in order to shed light on the third research question. The three parts of the test were examined, and particular attention was paid to the first part of the test, the Entrance Examination (EE), which consists of grammar, vocabulary, and reading comprehension sub-tests with multiple-choice questions. The analysis of the VELI placement test was performed in the following manner.

To begin with, answer sheets from all 571 test-takers who took the first part of the VELI placement test (the EE) between March, 1998 to August, 1998 were collected month by month. Once all the data had been gathered, information from each answer sheet was then input into the spreadsheet program Microsoft Excel '97. This information comprised the following data:

- a pseudonym for each test-taker, expressed as a seven-digit number (the first two digits signifying the year, the next two signifying the month, and the final three signifying the sequence number of the test-taker).
- the score (the number of correct answers out of a total of 48) representing the performance of the test-taker, i.e., the score that the test-taker received on the test.
- the level (lower or upper beginner, etc.)—matching the score above—as indicated by the "Entrance Evaluation and Placement Recommendations," a chart which, as mentioned earlier, is issued to all teachers who also function as interviewers during registration day and which consists of rows of ranges of scores and columns outlining the type of writing to be expected from a test-taker falling within that range, the type

of "conversation" during the interview that can be expected from a test-taker falling within that range, and a recommendation of level.

- the level as assigned by the teacher-interviewer after the interview with the test-taker.

Next, the level as indicated by the chart entitled "Entrance Evaluation and Placement Recommendations" (that is, the level which matched the score that the test-taker had received on the EE) was then correlated with the level as assigned by the teacher-interviewer. The correlation was performed using the PEARSON worksheet function of the Microsoft Excel '97 program, a function which yields the Pearson product moment correlation coefficient. Because the two parts of the test, the EE and the interview, have been considered to be separate and unrelated to each other and because the interview is believed to the arbiter of level assignment, it was expected that correlating the EE level recommendations with those of the teacher-interview would yield an indication of whether the EE was somehow flawed.

Subsequently, the second part of the test, known as "Entrance Evaluation 1: Writing" and consisting of sentence and paragraph writing, was examined briefly, as this part is not marked during the VELI placement testing process and seems not to be instrumental in assigning test-takers to levels of instruction. Nonetheless, 121 writing sub-tests were collected, this number representing the total number of test-takers who took the placement test in August, 1998, a month chosen for its high number of test-takers. The writing sub-test was broken down into its two sections, guided sentence writing and guided composition, and was analyzed in terms of the number of students who were able to complete each section and in terms of any trends that might emerge regarding anything the test-takers may have experienced in working on this part of the

test. Specifically, information from each answer sheet for the writing part was input into the spreadsheet program Microsoft Excel '97. This information comprised the following data:

- a pseudonym for each test-taker, expressed as a two-digit number (signifying the sequence number of the test-taker).
- the level (lower or upper beginner, etc.) as indicated by the "Entrance Evaluation and Placement Recommendations."
- the level as assigned by the teacher-interviewer after the interview with the test-taker.
- an indication of completion of each of the two sections of the writing test (on the spreadsheet, the number one was used to indicate that the test-taker was able to complete the section, the number zero to indicate the opposite).
- comments, if applicable.

It was expected that the ability to complete the section and the trends which might arise would serve to indicate whether this part of the placement test was problematic. For the sake of clarity, the ability to complete the section was taken to mean that the test-takers understood the directions and wrote something wherever required, thus finishing the section.

Finally, the last part of the test, the interview between the test-taker and a teacher—who functions as an interviewer and who is known as the "Teacher-Interviewer" (TI) for the purposes of this thesis, was studied in detail, as the interview is considered to play the most important role in determining levels of instruction. Because there is no observable product stemming from this interview (the interviews at VELI are neither audio- nor video-taped), it was examined from the point of view of the TIs through the use of the questionnaire (as explained earlier), further

interviews between the TIs (also as explained earlier) and the researcher of this thesis, and participant observation. As noted earlier, the questionnaire asked how useful the oral interview was as a means of placing new students in the appropriate level, the interview sought to discover what the TIs believed about the test, and the researcher participated in his normal registration day duties.

THE RESPONDENTS

The respondents in this thesis can be divided into two groups: VELI administration and VELI faculty. While the full complement of VELI administrative personnel comprises a Director, a Director of Studies, an Assistant to the Director of Studies, a Registrar, and three administrative assistants, only those directly involved in the placement testing process were approached to participate in the research for this thesis. Thus, the two who were contacted (and who subsequently agreed to participate in the research) were the Director of Studies, who reports to the Director and who is responsible for the academic direction of the institute, and the Assistant to the Director of Studies, who is responsible for the scheduling of students into classes based on the results of new students' placement tests or on returning students' requests. Both administrators are native speakers of English and are experienced in administration. The Director of Studies, a female, is a career ESL instructor and administrator, having been in the field for decades. The Assistant to the Director of Studies, also a female, is a career administrator, having worked at VELI since 1993.

The VELI faculty comprises teachers who are all native speakers of English and are experienced, qualified ESL instructors with at least an undergraduate degree in teaching English as a second language (or a degree in a closely related field). Their ages range from mid-

twenties to mid-fifties, and their ESL teaching experience ranges from a few years to a few decades. Many have worked at other institutes, both public and private, and many have taught overseas. All faculty have in-depth experience in the VELI placement testing process, as participation in registration day is mandatory—as explained previously. The VELI administration and faculty who participated in this study are summarized in Table 3, in which respondents are sorted according to their age range and gender.

Table 3

All Participants

Total Number of Participants (n=30: female = 22, male = 8)				
Participant	Gender	Age Range	Job Classification	Years Experience (ESL)
1	F	25 – 30	Instructor	2
2	F	25 – 30	Instructor	4
3	F	25 – 30	Administrator	5
4	F	25 – 30	Instructor	8
5	F	25 – 30	Instructor	5
6	F	25 – 30	Instructor	5
7	F	25 – 30	Instructor	2
8	M	25 – 30	Instructor	3
9	F	30 – 35	Instructor	5
10	F	30 – 35	Instructor	5
11	F	30 – 35	Instructor	5
12	F	30 – 35	Instructor	8
13	F	30 – 35	Instructor	9
14	F	30 – 35	Instructor	4
15	M	30 – 35	Instructor	8
16	M	30 – 35	Instructor	5
17	F	35 – 40	Instructor	10
18	F	35 – 40	Instructor	7
19	F	35 – 40	Instructor	7
20	F	35 – 40	Instructor	7
21	M	35 – 40	Instructor	11
22	M	35 – 40	Instructor	17
23	F	40 – 45	Instructor	7
24	F	40 – 45	Instructor	9
25	F	40 – 45	Instructor	9
26	F	40 – 45	Instructor	1
27	M	40 – 45	Instructor	13
28	M	40 – 45	Instructor	8
29	F	50 – 55	Administrator	17
30	M	50 – 55	Instructor	20

SUMMARY

This thesis addresses three research questions. Those questions are listed below, along with the methodology, in point form, used to examine each.

Research Questions

1. What are the historical and institutional influences and limitations that have shaped the VELI placement testing process?
    - Document analysis
        - a now-unused VELI test booklet, a current VELI test booklet, a VELI Registration Day packet, a VELI Instructors' Manual, the VELI Students' Manual, internal VELI documents pertaining to the scheduling of teachers for registration day, a VELI promotional brochure, e-mail correspondence between the researcher and the Director of the Toronto campus of VELI, personal communications between the researcher and both the Director of the Vancouver campus of VELI and the Assistant to the Director, a VELI Level Curriculum.
    - Participant observation
        - observation and log over a nine-month period, from February to October, 1998, concerning what went on among approximately 400 test-takers and 28 teachers on five registration days, each at the beginning of February, April, June, August, and October.

2. On what basis do those concerned think that the test is flawed, and how widespread is the problem?
    - Data analysis
        - questionnaires, distributed at VELI to all 28 members of faculty and 2 members of the administration, to which 18 instructors and 2 administrators responded.
        - informal conversation interviews, approximately 5 to 15 minutes each, to which 11 (8 female, 3 male) of the 28 VELI faculty members responded.
        - standardized, open-ended interviews, approximately 5 to 15 minutes each, to which 12 (7 female, 5 male) of the 28 VELI faculty members responded.

- Participant observation
    - observation and log over a nine-month period, from February to October, 1998, concerning what went on among approximately 400 test-takers and 28 teachers on five registration days, each at the beginning of February, April, June, August, and October.

3. Based on the findings of the previous research question as well as test analysis and participant observation, where might problems with the institute's placement testing and the resulting placements lie?
    - Document analysis (a three-part analysis of the VELI placement test).
        - Part One of the VELI placement test (grammar, vocabulary, and reading comprehension sub-tests with multiple-choice questions)
            - answer sheets collected from all 571 test-takers who took the first part of the VELI placement test between March, 1998 to August, 1998.
            - data input into Microsoft Excel '97 (pseudonym, score out of 48, level as indicated by the "Entrance Evaluation and Placement Recommendations" chart, level as assigned by teacher-interviewer).
            - level as indicated by the "Entrance Evaluation and Placement Recommendations" chart correlated with level as assigned by teacher-interviewer using Pearson product moment correlation coefficient.
        - Part Two of the VELI placement test (sentence and paragraph writing)
            - answer sheets collected from all 121 test-takers who took the second part of the VELI placement test in August, 1998.
            - test broken down into its two sections, guided sentence writing and guided composition.
            - data input into Microsoft Excel '97 (pseudonym, level as indicated by the "Entrance Evaluation and Placement Recommendations," level as assigned by teacher-interviewer, whether section was completed or not, comments).
            - data analyzed in terms of the number of students who were able to complete each section and in terms of any trends that might

emerge regarding anything the test-takers may have experienced in working on this part of the test.

- Part Three of the VELI placement test (interview)
  - Data analysis
    - questionnaires, distributed at VELI to all 28 members of faculty and 2 members of the administration, to which 18 instructors and 2 administrators responded.
    - informal conversation interviews, approximately 5 to 15 minutes each, to which 11 (8 female, 3 male) of the 28 VELI faculty members responded.
  - Participant Observation
    - observation and log over a nine-month period, from February to October, 1998, concerning the actions of approximately 400 test-takers and 28 teachers on five registration days, each at the beginning of February, April, June, August, and October.

CHAPTER IV. FINDINGS

This chapter will review the findings in terms of answering the three research questions, which were the following:

1. What are the historical and institutional influences and limitations that have shaped the VELI placement testing process?
2. On what basis do those concerned think that the test is flawed, and how widespread is the problem?
3. Based on the findings of the previous research question as well as test analysis and participant observation, where might problems with the institute's placement testing and the resulting placements lie?

In reviewing the findings, this chapter has been divided into three sections. The first section, "Research Question #1," provides an answer to the first research question and is an exploration both of the history of VELI from the time of its founding up to the present and of the institutional influences and limitations that have affected placement testing at the institute. The second section, "Research Question #2," provides an answer to the second research question and is an examination of why those concerned think the test is flawed and an exploration of how widespread the problem is. The third section, "Research Question #3," provides an answer to the third research question and is a detailed dissection of the three parts of the current VELI placement test.

RESEARCH QUESTION #1

In order to answer the first research question (what are the historical and institutional influences and limitations that have shaped the VELI placement testing process?), data was collected from a variety of sources, including the recollections of the researcher; personal communications between the researcher and the Director of Vancouver campus of VELI, the Assistant to the Director, and the Director of the Toronto campus; and internal VELI documents.

Data aside, most discussions of placement testing focus on the test itself, as if it existed as an entity separate from its surroundings. This may certainly be the case when placement tests come into use from external sources, such as a commercial test or one developed in a department external to that providing ESL classes. However, a placement test that is developed in-house, such as that used at VELI, must have come from somewhere—it did not spontaneously come into existence—and thus must have historical and institutional influences and limitations specific to the institute that have shaped the test into what it is today.

An analysis of the data revealed that historical and institutional influences and limitations have shaped the VELI placement testing process. The history of placement testing dates back to the early years of the institute. In 1988, the year in which the school was founded, VELI was a small school catering primarily to young, privileged adult foreign students who had come to Canada temporarily to learn English or to improve their already-acquired level of English. Students studied at the institute either because they hoped to better their job prospects once they returned to their own countries or because English was their hobby and they wished to study further in an environment full of native English speakers. Doing so would thus allow the students at this institute to practise and use what they had already learned and were currently learning.

VELI in 1988 had seven classrooms, only five of which were initially in use. The maximum number of students per class was set at 14; in extremely rare circumstances, though, 15 were placed into a class if absolutely necessary. While no records now exist listing exact numbers, assuming a maximum number of 14 students per class, the total student population at that time was perhaps 70. Realistically, however, the number was smaller, perhaps 50 to 60 students at first. As one of the owners of the institute proudly pointed out on numerous occasions in the initial years of VELI, these students came to the institute almost exclusively by word of mouth (R. Mainn[4], personal communication, 1988) because at that time, the school neither advertised nor used the services of agents.

The monthly intake system, as mentioned above, was in place at that time; however, because of the small population of students and because of a relatively minor turn-over, registration day would see a very light intake: perhaps as few as five to seven students. An entrance examination was thus unnecessary, as placements were accomplished entirely through oral interviews perhaps lasting a maximum of five or ten minutes each. Originally, one of the two owners of VELI, who also functioned as the director, conducted the interviews, and since there were only five levels of instruction (two beginner classes, two intermediate, and one advanced) and a faculty of five or six teachers, the owner-director could easily place students into the class at the appropriate ability level. The ease with which the owner-director could make these placements was due to the fact that there were only five divisions of student ability level with which to deal and to the fact that an intimate working relationship was maintained between the owner-director and the small faculty. Thus, the owner-director had a very good feel

[4] A pseudonym.

for what was being taught in each of the five levels and for how the students were performing in each of those levels.

The institute experienced great changes in the early 1990s. One change was the expansion of the institute: an additional, much larger campus was added two blocks away from the original. The new location became the main campus. As a result, the owner-director and support staff moved there, while the original location became a satellite campus containing classrooms only. Another change was the division of the institute in terms of the type of classes offered. Until this time, there were few "content" classes—as those that focus on specific skills are called at VELI. Almost all classes taught at the institute were general in nature and were even called the "general" classes; that is, all skills (listening, speaking, reading, writing) were covered in each class at each level. After the expansion, however, a number of "content" classes were added, and among those offered were a TOEFL preparation class, an oral production class, an academic preparation class, and a business class. These classes were specific in nature, were different from the general classes, and focused mainly on only one or two skills or areas; thus, they were dubbed the "special skills" classes. The division of the institute therefore involved the location in which the "general" classes and the "special skills" classes were conducted: the "general" classes, considered to be the backbone of the institute, were held mostly in the new campus, while the "special skills" classes, considered to be complementary to the "general" classes, were held mostly in the old campus.

Because of the expansion of the school and the addition of the content classes, VELI and its student population became much larger. The institute now had 25 classrooms, most of which were in use, while the maximum number of students per class remained set at 14. Again, although no records now exist listing exact numbers, assuming a maximum number of 14

students per class, the maximum student population at that time was approximately 300. While some of these students came to the institute by word of mouth, VELI started to advertise and to use the services of agents. Of course, the maximum population of 300 students was not achieved initially, but the number of students entering the school on registration day rose dramatically in a very short period of time. Indeed, registration day now saw an intake of students numbering in the 40s or even higher. It was now difficult for the owner-director of the institute to place students through oral interviews alone: with interviews lasting a minimum of just ten minutes each, it would take at least approximately six hours to complete just the interview process. This time, of course, does not even take into account the additional time it would take to schedule these students into the now increased number of classes or to resolve scheduling conflicts which might result from the placements.

Furthermore, the number of the levels of instruction increased to seven (two beginner classes, lower and upper; three intermediate classes, lower, mid, and upper; and two advanced, lower and upper), the number of "special skills" classes increased, and the faculty grew to over 25 teachers. Along with the much greater number of students to interview, an intimate working relationship could no longer be maintained between the owner-director and the now larger faculty. It was no longer possible for the owner-director either to have an intimate knowledge of what was being taught in all of the levels and in all of the myriad special skills classes or to know how all the students were performing in all of those classes.

Consequently, it was decided that an entrance examination along with a brief interview was needed to process the large numbers of students quickly and efficiently during registration day. It was felt that the most qualified of the faculty to create an entrance test would be the researcher of this thesis, who had had some prior experience in ESL entrance testing at a large

local community college in addition to a basic knowledge of testing from teaching a TOEFL preparation class at VELI and at another institute for a number of years. The researcher was therefore enlisted to create the entrance test. The result was known as the "Entrance Evaluation Test (EET)," the precursor to the EE used today. Consisting of three sections (a short grammar section with fill-in-the-blank and error correction questions, a vocabulary section, and a reading comprehension section—all sections with multiple-choice questions), plus an oral interview and loosely based on the TOEFL, the EET was normed on the entire population of the school at that time. The resulting norms alone were used to develop a system to assign students to levels based on their performance on the EET and on their oral interview. In spite of the care taken in the rather brief and unsophisticated statistical review of the EET, the EET came to be perceived to be problematic, both by instructors and administration alike: it was somehow flawed and misplacements were regularly occurring as a result. In fact, according to the director of the VELI Toronto campus, who oversaw the subsequent analysis of the EET, "[the EET] had proven unsatisfactory . . . . [O]ur analysis of 100 students' errors [on the EET] showed that, for some questions, just as many advanced students as beginners answered incorrectly. In other words, the question didn't help us distinguish between levels. However, quite a few of the test questions—namely, those which showed some efficacy—were retained" (J. White[5], personal communication, August 19, 1998).

The institute experienced more changes in the mid 1990s. One change was that in 1994, the satellite campus in Vancouver was closed, and all its students and courses were transferred to the main institute. Another change was that, with the loss of the eight classrooms resulting from

[5] A pseudonym.

the closure, course times at the main campus were adjusted in such a way as to maintain and even to increase the student capacity of the school. Rather than having one set of full-time students studying from Monday to Thursday in two three-hour classes per day, the institute now had two separate sets of full-time students studying from Monday to Friday in two two-hour classes per day. For the sake of convenience, each two-hour class time was called as a "block," with the blocks lettered consecutively from A to D. To clarify, one set of full-time students studied in Block A and Block B from 8:30 a.m. to 1:00 p.m. and then went home, while another set of full-time students studied in Block C and Block D from 1:30 p.m. to 6:00 p.m. With two sets of full-time students in a total of 16 available classrooms at a maximum of 14 students per class, the total student capacity of the school rose to approximately 450. While this number was not achieved initially, the student population increased and more teachers were hired accordingly. A further change occurred in April, 1995, when the institute expanded again; this time, another smaller campus containing eight classrooms was opened in Toronto, Canada.

One result of these changes was an exacerbation of the perception that the EET was flawed. In Vancouver, this increased awareness of problems with the EET arose because as the number of students to be placed and the number of classes increased, the EET seemed to be less and less accurate. As teacher-interviewers who interviewed and placed test-takers on registration day often pointed out, the test yielded results that at times seemed in no way to correspond with the final placements of the students. Even in Toronto, which registered and tested its students separately from the campus in Vancouver, there was a heightened awareness of a problem with the EET (J. White, personal communication, August 19, 1998), and thus, in 1996, it was reviewed and modified. As noted above, the entrance test expanded from four components (a short grammar section with fill-in-the-blank and error correction questions, a vocabulary section,

a reading comprehension section, and an oral interview) to ten (see Figure 1). Because of staff, schedule, and time constraints, the modification of the test took place at the Toronto campus (E. Forester[6], personal communication, August 10, 1998), where the new form of the test was normed on its population. The resulting norms were used to develop a system to assign students to levels based on their performance on the EE, on the writing test, and on their oral interview. Similar to what had happened with the EET, in spite of the care taken in its statistical review, the revised placement test, like its predecessor, came to be perceived to be problematic, both by instructors and administration alike: it, too, was somehow flawed, and misplacements were occurring as a result.

Indeed, misplacements at one point became such a problem that VELI administration began the practice of internally publishing the names of the students who were misplaced and the TIs who had interviewed them, so as to identify at post-registration day meetings those who had conducted those interviews. It would seem that the implication on behalf of the administration at that time was that the fault of the misplacements lay with the teacher-interviewers (TIs), and it was hoped that by identifying them, the TIs would somehow become cognizant of what they were doing wrong and adjust whatever they were doing so as to become more accurate. After a time, it was found, however, that there seemed to be no pattern to the misplacements that could be attributed to the TIs alone: most, if not all, of the TIs appeared on that list in no particular order or in no particular pattern. In the end, it appeared that the fault may lie elsewhere. The placement test was indeed problematic, and at least one or two highly disruptive misplacements *per class* were occurring *each month*.

---

[6] A pseudonym.

In retrospect, then, historical and institutional influences and limitations have indeed shaped the VELI placement testing process. It has undergone a metamorphosis of sorts, starting out as an informal, blunt instrument and changing, due to the historical and institutional changes in the institute, over time into a formalized, well-honed instrument—yet one which could possibly be further refined.

RESEARCH QUESTION #2

In order to answer the second research question (on what basis do those concerned think that the test is flawed, and how widespread is the problem?), a questionnaire was distributed to all faculty of the school and to the two administrators whose jobs are directly affected by the placement process: the Director of Studies, who administers the test and leads the teachers through the registration day pre-interview meeting, and the Assistant to the Director of Studies, who schedules the students into classes based on the final recommendation of level assigned by the interviewing teacher. In addition, faculty and administration were interviewed about their beliefs concerning the placement test and the placement testing process.

Why Do Those Concerned Think that the Test is Flawed?

An analysis of the questionnaire and interview data revealed that the faculty and administration felt that the test was flawed based on their belief that components of the placement test were of limited usefulness and that the entire placement process was only somewhat useful to them. In reference to the questionnaire[7], respondents were posed with three

[7] See Appendix A for the wording of the questions.

questions regarding how useful they felt as a means of placing new students in the appropriate level (1) the first part of the test (the Entrance Examination) was, (2) the second part (the writing component) was, and (3) the third part (the interview component) was. Additionally, respondents were posed another question regarding how effective they felt the entire placement process was. For the most part, respondents were split on the usefulness of the Entrance Examination; one felt that it was very useful and eleven that it was only somewhat useful, yet the remainder, eight respondents, felt that it was neither useful nor useless, somewhat useless, or useless. The results are summarized in Table 4.

Table 4

Usefulness of the Entrance Examination

	Very Useful	Somewhat Useful	Neither Useful Nor Useless	Somewhat Useless	Useless
Teachers and Administrators (n=20)	1	11	3	3	2

Respondents were also split on the usefulness of the writing component of the placement test; one felt that it was very useful and nine that it was only somewhat useful, while the remainder, ten respondents, felt that it was neither useful nor useless, somewhat useless, or useless. The results are summarized in Table 5.

Table 5

Usefulness of the Writing Test

	Very Useful	Somewhat Useful	Neither Useful Nor Useless	Somewhat Useless	Useless
Teachers and Administrators (n=20)	1	9	2	5	3

In contrast to the apparent lack of usefulness of the first two parts of the test, respondents seemed to feel that interview component of the placement test was a little more useful than the other two components. Three respondents felt that it was very useful, while seventeen found it only somewhat useful. No respondents felt that it was neither useful nor useless, somewhat useless, or useless. The results are summarized in Table 6.

Table 6

Usefulness of the Interview

	Very Useful	Somewhat Useful	Neither Useful Nor Useless	Somewhat Useless	Useless
Teachers and Administrators (n=20)	3	17	0	0	0

Respondents seemed to have a mixed reaction to the effectiveness of entire placement testing process. Three felt that it was very effective, fourteen that it was only somewhat effective, and three that it was ineffective. The results are summarized in Table 7.

Table 7

Effectiveness of the Entire Placement Process

	Very Effective	Somewhat Effective	Neither Effective Nor Ineffective	Somewhat Ineffective	Ineffective
Teachers and Administrators (n=20)	3	14	0	3	0

In reference to the interview, it also revealed a belief that components of the placement test were of limited usefulness and that the entire placement process was only somewhat useful. While one teacher felt that the first part of the test was useful ("the grammar test works well for me at the lower levels"), other teachers believed that that part of the test was problematic ("the test gives me an indication of the level of the student, but often the results of the interview are different from what the test is suggesting"). While some teachers felt the writing test was useful to a degree ("if there's nothing on the page, then that could mean the difference between a Lower Beginners student and an Upper Beginners"), others questioned its value ("the written test means not much [to me]" and "I don't even look at it"). In addition, some teachers believed that the interview part of the placement process was problematic ("the interview is difficult for students;" "it's very intimidating;" "not a relaxed environment;" and "the interviewee is nervous"). In

addition, teachers referred to the testing and placement process as "subjective" and "not accurate." One teacher mentioned that the process "is less accurate than the criteria used for end of month promotion," exit criteria which are used by the school to determine which students should be promoted to the next level of instruction.

How Widespread is the Problem?

The interview data (from the informal conversation interview, specifically) hinted that the problem was widespread, if only because respondents were not asked directly to comment on how many misplacements they experienced in their classes. In passing, however, the Assistant to the Director of Studies mentioned that there are "lots of misplacements at registration day," and one teacher stated that there are "always a couple of obvious ones [i.e., misplacements] on the first day [of classes], but also a couple of difficult ones [i.e., misplacements that are not immediately apparent]."

The questionnaire data, in contrast, revealed that the problem was widespread; a number of misplacements severe enough to warrant movement into a more appropriate class were occurring as a direct result of the placement testing process. Of the twenty faculty and administration who responded to the questionnaire, nine teachers and/or administrators stated that one student *per class* was misplaced immediately after registration day, six stated that two were misplaced, four stated that three were misplaced, and one stated that four were misplaced. No respondents stated that there were zero misplacements. The results are summarized in Table 8.

Table 8

Number of Misplaced Students, According to Teachers and Administrators

	Number of Students Misplaced *Per Class*				
	Zero Students	One Student	Two Students	Three Students	Four Students
Teachers and Administrators (n=20)	0	9	6	4	1

It is important to note that those students who are recognized as misplaced are moved into a more appropriate class for their skill level, and that once they are placed into another class, the students (and the teachers) appear to be happy with the adjustment. In fact, there has never been any case at VELI of a student being identified as misplaced and thus moved by one teacher, only to have the teacher of the other class identify that student as misplaced and recommend further movement (Assistant to the Director of Studies, personal communication, April 14, 1999).

An Additional Consideration: Recognition of Misplaced Students

It was found that the basis for the feeling that the test is flawed is the faculty and administration's belief that components of the placement test were of limited usefulness and that the entire placement process was only somewhat useful to them. In addition, it was found that the problem was widespread. These two points, however, beg the question: how, in fact, do teachers know that students have been placed incorrectly into their classes—thus necessitating the movement of the misplaced students out of the class and into classes more appropriate to

their abilities? In order to investigate this question, twelve of the thirty VELI faculty who taught beginner, intermediate, and advanced classes were asked two questions in a standardized, open-ended interview: "How do you know when students have been misplaced into the classes you teach?" and "How long after classes have begun can you tell that students have been misplaced into the class you are teaching?"

In response to the first question, interview data revealed that although, as one teacher put it, "it may be more difficult to tell [that they have been misplaced] because students may be able to use strategies to hide [their misplacement], especially coping strategies," students exhibited various signs of misplacement severe enough to warrant movement into a more appropriate class. One was the inability to produce language. On that topic, one teacher noted that students "can't do the work, or they can't do it easily." A second observed that they "don't participate [and] when asked direct questions, they are uncomfortable." A third noticed that they "don't produce the language expected of them at that level." Another sign was the inability to understand. One instructor said that she knows students have been misplaced when "they don't understand what I'm talking about—as in following my instructions [and] if they say, 'Huh?'" Another noted that they have "poor comprehension right off the bat when you tell them something." Similarly, another observed that they display an "inability to follow instructions right off the bat." Another sign was the inability to interact with other members of the class. One teacher mentioned that she knows that students have been misplaced when in addition to other signs, the misplaced students "are not making contact with other students." Another noted the lack of participation in the class "when students are in groups and one guy is sitting by himself."

In response to the second question, interview data revealed that regardless of the level the instructors taught, they all stated that within the first two or three days of classes, the instructors

could identify misplacements necessitating a movement into a more appropriate class. Examples of the responses are the following: "[I know] in the first day—sometimes a suspicion in the first day, but within two days or so usually"; "two days, I can tell"; "at the lower levels, we can tell by the time the first class is over"; and "the second or third day—I'm sure that they've been misplaced."

RESEARCH QUESTION #3

In order answer the third research question (where might problems with the institute's placement testing and the resulting placements lie?), a review of the individual parts of the placement test—Parts One, Two, and Three—was undertaken. Determining whether each part of the test is problematic should of course also serve to identify where the problems lie.

Analysis of Part One

As mentioned earlier, Part One, the Entrance Examination (EE), contains questions on grammar, conversation, vocabulary, and reading comprehension to be completed within a time limit of 30 minutes; this portion of the test was examined closely because it is this part and its score on which the initial assignment of level is based. That is, the score on this part of the test is first used immediately after the marking stage to indicate whether test-takers should be sent to a teacher-interviewer (TI) who specifically interviews at the lower or upper beginner level; the lower, mid, or upper intermediate level; or the lower or upper advanced level. It is also this part of the test that faculty and administration perceive to be most flawed, because the level as indicated on the test often does not match the performance of the test-taker during the interview.

To continue, 571 EEs dating from March, 1998 to August, 1998 were examined. This number represents the total number of students entering the school from March to August, so chosen because it is during these months, the spring and summer months, that the school receives the highest number of students. For this reason, it was felt that a greater number of tests would yield a more accurate appraisal of whether or not the EE was flawed. As explained earlier, the TIs fill out a form, first writing down the test-takers' score on Part One (the EE), and later indicating the level of the test-takers based mostly on the result of the interview. This form was used in the examination of this part of the test: the score of Part One and the resulting level as indicated by the Entrance Evaluation and Placement Recommendations—a chart given to every TI containing rows of ranges of scores followed by the type of writing to be expected from a test-taker falling within that range, a recommendation of level, and the type of "conversation" that can be expected from a test-taker falling within that range—was correlated, using the Pearson product moment correlation coefficient, with the final level as assigned by the interviewer.

When the level as recommended by the EE, the first part of the placement test, was checked against the level as assigned by the TI, it was discovered that the Pearson product moment correlation coefficient between the two was higher than expected ($r=0.84$). Such a good correlation seems to indicate, at first glance anyway, that there is a reasonable relationship between the assignment of level by Part One of the test and that by the TI and that the problem may not lie with Part One. However, this result contradicts the seeming strong belief of the faculty and administration that the test—particularly the EE—is somehow flawed. Is the EE, then, really problematic?

In spite of the good correlation, the EE could be viewed to be problematic for the following reason: the simple correlation of the two data sets does not take into account at least

two mitigating factors. First, while the correlation is high, it is of course not perfect, indicating that mismatches between the two data sets must still occur. If the test-taker's level as assigned by the TI is taken to be the more accurate of the two data sets (as is the case at VELI: students are assigned to classes based *solely* on the TIs' recommendation), the conclusion must be that the EE is problematic. Second, when the level as assigned by the EE and the level as assigned by the TI were compared, there were 308 mismatches between the two data sets, and of these, 248 were different by a factor of only one level (e.g., a situation in which the EE places the test-taker into a Lower Advanced class, the sixth level, while the TI assigns the test-taker to a Lower Intermediate class, the fifth level). While this situation may result in a good correlation, it cannot account for the fact that at VELI, the difference between each of the eight levels is not equal. For example, the difference between Mid-Intermediate (the fourth level) and Upper Intermediate (the fifth) is believed at VELI to be small; the abilities of the students studying in either level are similar, the difference being mostly one of fluency level. However, the difference between Upper Intermediate and Lower Advanced (the sixth level) is believed to be much greater; the abilities of the students studying in either level can be markedly different, with those in the higher of the two knowing more grammar and vocabulary and being much more accurate and much more fluent. Thus, superficially, the correlations may look good, but they cannot take into account the fact that there may be greater differences between the levels that are simply not revealed by correlation alone. Once again, if the test-taker's level as assigned by the TI is taken to be the more accurate of the two data sets, the conclusion must be that the EE is problematic.

For these two reasons, it was found that Part One of the VELI placement test, the EE, is problematic, and it is in this section where at least one of the problems with placement testing at VELI lie.

Analysis of Part Two

As mentioned earlier, Part Two, the assessment of the test-taker's writing, is divided into two sections to be completed within a time limit of eight minutes: the first contains verbs which the test-taker must convert to the past tense and then write a sentence containing that verb, while the second contains a strip of five frames along the left side of the page, with directions to the test-taker to "write a paragraph about the weather" followed by two questions related to the weather. This part was not examined in as much detail as the previous because it is not marked and seems not to play very much of a role if any in the interviewers' final placement of the test-taker into a level. In fact, of the 20 respondents to the questionnaire, only one felt that the writing assessment was very useful, a number (9 respondents) felt that it was only somewhat useful, while a small majority (10 respondents) indicated that it was useless, somewhat useless, or neither useless nor useful as a means of placing test-takers into the appropriate level. These results are summarized in Table 9.

Table 9

Usefulness of Part Two

	TIs' Opinions of Part Two				
	Very Useful	Somewhat Useful	Neither Useful nor Useless	Somewhat Useless	Useless
n=20	1	9	2	5	3

Therefore, of the 571 tests gathered for analysis of Part One, for Part Two only 121 of the 571 were examined, this number representing the total number of students entering the school in August, 1998, a month chosen for its high number of test-takers and its proximity in time to the writing of this thesis.

In any case, the writing test was divided into its two sections (as described above) and analyzed both to ascertain how many test-takers completed each section and to determine from the product of their writing if any trends might emerge regarding problems the test-takers may have experienced in working on this part of the test. The ability to complete the section and the trends which may emerge should serve to indicate whether this part of the placement test is problematic. For the sake of clarity, "completion" in this sense is taken to mean that the test-takers understood the directions and wrote something where required, thus completing the section. The numbers of test-takers who completed each section are summarized in Table 10.

Table 10

Number of Test-takers Completing Each Section of the Writing Test

Level	*n*	Section One		Section Two	
		Complete	Incomplete	Complete	Incomplete
Lower Beginner	22	17	5	6	16
Upper Beginner	14	11	3	5	9
Lower Intermediate	18	14	4	9	9
Mid Intermediate	22	18	4	16	6
Upper Intermediate	21	20	1	18	3
Lower Advanced	19	18	1	14	5
Upper Advanced	5	5	0	4	1
Total	121	103	18	72	49

Concerning the completion of both sections of the writing test, it was discovered that most test-takers had little difficulty with the first section of the test, the majority of them able to transform the verb to the past tense and to complete, with a high degree of accuracy, the sentence based on that verb. It was discovered, however, that the level of English used by the test-takers was not all that different, regardless of native language, ability, or final placement, and thus could certainly not be useful in distinguishing level. For example, there is little difference among the following examples in table 11.

Table 11

Examples of Writing on Section One of the Writing Component

TI Assigned Level	Native Language	Examples
Lower Beginner	Chinese	"I had a class on yesterday morning." "I did homework."
Upper Beginner	Korean	"I had a watch last year." "I did watch a TV last night."
Lower Intermediate	Japanese	"I had a car in Japan." "I did homework last night."
Mid Intermediate	German	"He had a nice day." "I did a test."
Upper Intermediate	Korean	"I had my breakfast in the morning." "I did my best."
Lower Advanced	Chinese	"I had a job in Taiwan." "I did my homework."
Upper Advanced	German	"I had an accident last week." "Did you take the umbrella?"

By way of contrast, it was found that many test-takers had problems with the second section. In Table 10, of the 121 test-takers, 49 could not complete the test regardless of ability level. Those at the beginners levels, lower and upper, experienced a great amount of difficulty writing a paragraph of comparison about the weather. Many wrote nothing, while some wrote only a sentence or two simply describing one or two of the drawings in the frame on the left side of the page. Those at the intermediate levels experienced some difficulty, and although many did write something, they often were not able to write very much, their sentences often short—of

the three- or four-word subject-verb-object variety—and their paragraphs mostly unfinished. Those at the advanced levels had a little more success with this section, writing more than those at lower levels, but even so, one-quarter of the test-takers still did not manage to complete the section as directed. Interestingly, a trend that emerged from the analysis of this section was that a number of the test-takers, regardless of level, were confused by the strip of five cartoon-like frames along the left side of the page. Some ignored the directions and wrote only about what they could see in the frames, while others tried to relate the directions to the drawings but were unsuccessful; as explained earlier, the drawings simply do not correspond to the directions.

Because of the inability of the writing test in its first section to distinguish among levels, and because of both the inability of the test-takers to complete the second section and the confusion on behalf of the test-takers caused by the disparity between the directions and the drawings in the strip of frames at the side of the page, it was discovered that the writing part of the placement process at VELI is indeed problematic. It is in this part of the test where yet another one of the problems with placement testing at VELI lie.

Analysis of Part Three

As mentioned earlier, Part Three, the oral interview, is administered by a teacher and consists of a meeting between the test-taker and a teacher-interviewer (as explained earlier, known in this thesis as the "TI") who spends some time interacting with the test-taker, mostly by asking questions for the test-taker to answer. These questions are not standardized. At the TI's disposal, however, are three interview aids: a black-and-white drawing of a busy street corner city scene, a cartoon depicting a busy beach scene, and a list of questions based on grammar and divided into ability levels. In addition, as mentioned previously, a chart called the "Entrance

Evaluation and Placement Recommendations" is available to the TI. By its very nature, the oral interview is difficult to examine in depth because there is no observable product, aside from the final judgement of level as indicated by the interviewer. Therefore, the interview was examined from the point of view of the TIs since they are the ones who determine the level at which the test-taker will be placed. Interview data revealed that of all the parts of the entrance test, the interview was the most useful as a means of placing new students in the appropriate level. Of the 20 respondents to the questionnaire, 8 felt that the oral interview was very useful, while the rest, 12 respondents felt that it was somewhat useful. No respondents believed that the interview was neither useful nor useless, somewhat useless, or useless. In addition, some teachers mentioned that by interview alone, they could place students accurately and quickly. In fact, one stated how quickly he knew the test-taker's level by interview alone: "within 30 seconds, I know [in which level the student belongs]." Another commented on his using the interview alone: "I don't even look at the EE or the writing. I base my placements entirely on the interview."

Although this part of the placement test seems to be the strongest, the interviews are not free from criticism. First, while it is true that teachers new to VELI are not scheduled to interview during their first registration day, and that they must observe experienced TIs go about their business, TIs really receive no formal training in the art of interviewing. This point alone may introduce inconsistencies that ultimately lead to misplacements. Indeed, one teacher feels that "two different teachers could put the same students in two different levels." In addition to the lack of training is a perception that the competency of the TI during the interview may also play a role. One experienced VELI teacher believes that the competency of the TI "is greatly a factor in misplacement of students." Another mentioned the shortness of his interviews ("I could do that interview in just five minutes") in comparison with those of other teachers. This

instructor was commenting on the fact that for him, the short duration of his interviews was a result of his skill at interviewing, while for some teachers, the long duration of their interviews was a result of their lack of competency in interviewing.

In answer to the third research question, it was found that although the interview is felt to be the most useful part of the entrance test, because of the lack of formal training in interviewing and because of questions regarding the competency levels of some of the interviewers, the interview part of the placement process at the institute in Vancouver is also problematic—perhaps to a lesser degree than the other parts, but problematic nonetheless. Thus, it is also in this part of the test where yet another one of the problems with placement testing at VELI lie.

CHAPTER V. DISCUSSION

This thesis set out to describe and analyze the placement testing process at VELI in order to answer three research questions: First, what are the historical and institutional influences and limitations that have shaped the VELI placement testing process? Second, on what basis do those concerned think that the test is flawed, and how widespread is the problem? Third, based on the findings of the previous research question, where might problems with the institute's placement testing and the resulting placements lie? This chapter will discuss the findings in relation to each of the three research questions. Specifically, this chapter will examine placement testing at VELI, perceptions of the placement test, and recognition of misplaced students. In addition, it will discuss each of the three parts of the current VELI placement test.

PLACEMENT TESTING AT VELI

Historical and institutional influences and limitations have undoubtedly shaped the VELI placement testing process. Indeed, had VELI remained a small school with five levels of instruction and a monthly intake of under ten students, the placement testing process and the problems concerning misplacement would probably not be much of an issue. After all, there originally was no formal placement test, with placements being made solely by the director, a person who had an intimate knowledge of the teachers of the institute and the classes they taught. Additionally, there existed no demand for a placement test, and the interview was accepted by students entering VELI at that time as a means of placement; for a school of that size with such a small number of classes to begin with, a long, complicated placement process would have been felt to be inappropriate. There was little, if any, complaint about the placement process in the

beginning, and thus, there was no real need to develop a placement test or to institute a complicated placement testing process.

From a historical perspective, then, it could be said that the placement system in its current form at VELI is a result of the ever increasing expansion of VELI and the consequent growth in number of students taught, courses offered, and instructors hired to teach those courses. In retrospect, for each major increase in the capacity of the school, there has been a corresponding change in the placement process. Initially, as the school grew and became more popular, the number of students to be placed increased, and there arose a need to place them efficiently and effectively. The interview process alone was seen to be time-consuming and inadequate, and thus, the first placement test to be used at VELI was developed. It was based loosely on the TOEFL—in hindsight, a constraint—mostly because of the background of the test creator and partly because the multiple-choice format of the TOEFL was perceived at that time to carry much face validity. Subsequently, as the school expanded to Toronto and as class times expanded, there arose a need to modify the original placement test because it was perceived to be inaccurate. The corresponding change in the placement process occurred when the initial placement test was revised into what it is today: a combination of the EE, the writing test, and the oral interview. Now that VELI has continued to grow and has expanded once more—this time, a campus in San Diego, California—the placement process is felt to be again inaccurate, and presumably there will be a corresponding change in the placement testing process, as the current process is once again perceived to be inaccurate.

From an institutional perspective, it follows that the placement system in its current form at VELI is also a result of the ever increasing capacity of the school. As the capacity grew, so did the number of classrooms and courses offered, the result of which being the necessity for a

testing instrument to assign the expanding numbers of students into their classes in a time- and energy-efficient manner. Additionally, capacity has also played a major role in the shaping of the VELI placement process: a large cafeteria allows placement testing to occur on a large scale, with as many as 150 students taking the placement test at the same time. Were there no cafeteria, it is doubtful that the system would resemble what is in place today: within the confines of VELI, there simply is no other serviceable method to divide a large number of test-takers into smaller groups to be taken to individual classrooms to be briefed on the procedures of the school, to be tested, to be interviewed, and so on.

PERCEPTIONS OF THE PLACEMENT TEST

As stated in the findings, an analysis of the questionnaire and interview data revealed that the faculty and administration felt that the test was flawed based on their belief that components of the placement test were of limited usefulness and that the entire placement process was only somewhat useful to them. These perceptions of the placement process at VELI have important implications in light of what Bradshaw (1990) calls "consumer validity," a term "used to cover attitudes and feelings of both test-takers and test-users, the after-effects of testing procedures, and the possible debilitative effects on test scores of aspects of test design and test administration" (p. 26). Bradshaw believes that consumer validity is worthy of concern, as adverse effects may not be apparent through statistical analysis. At VELI, the after-effects of testing procedures to which Bradshaw refers may be that, because the placement test may appear to TIs as little more than an elaborate exercise, they may not devote the time and attention necessary to place test-takers accurately during registration day. After all, within the following few days after registration day, classroom teachers are able, post facto, to recognize

misplacements and to initiate a process that ends with the movement of the misplaced student to another more appropriate class—regardless of the difficulties that inaccurate placement at VELI poses for all concerned. The result of this problem may be a repeating cycle of misplacement: on any one registration day, the testing process is felt to be inaccurate and thus, TIs lose faith in it. Being classroom teachers themselves, the TIs may care less about the placements themselves, knowing that misplaced students will just end up being moved to another class; consequently, the TIs may not make the effort to place students accurately during the interview. Because this situation exists during this particular registration day and nothing is done afterward to rectify it, the loss of faith carries on to the next registration day, when the cycle repeats itself.

RECOGNITION OF MISPLACED STUDENTS

It is interesting to note that teachers easily and rapidly recognize misplaced students, and can do so with accuracy. There are at least two reasons. One is that VELI instructors know very well the classes they teach. They are familiar with the material they teach and with the capabilities of their students. Thus, a situation in which the in-class performance of a student does not match the teacher's expectations for students at that level is easily and quickly noticed. The other reason is that the signs of misplacement mentioned earlier are easily recognizable to those in a profession that caters to "paying customers" on whom the teachers depend for their salaries. Teachers in private institutions such as VELI are expected to maintain a constant surveillance of students in order to gauge their comfort levels in class, their performance levels, and so on, for unhappy students may result in poor publicity. Therefore, signs such as the withdrawal of students from participating with their classmates and the lack of appropriate performance are not easily missed.

THE PLACEMENT TEST

As stated in the findings, problems with the institute's placement testing exist in each part of the placement test. A discussion of each of these parts in relation to the findings follows.

Part One

The discrepancy between the level indicated by the first part of the test—the EE—and that assigned by the TI is easily observed regardless of the correlation between them, is often noted by interviewers during registration day, and results in the feeling that Part One is of questionable usefulness. The reasons for the discrepancy are many. One reason is that, because of a lack of resources, the statistical analyses performed during the creation of the EE and afterwards were understandably unsophisticated. In fact, the only statistic calculated was item difficulty. To do so, the administrator who developed the EE studied results on the original placement test of 100 then-current students spread throughout all of the levels of the school. The percentage of incorrect answers for each question on the test was gathered; from that, questions having a high percentage of incorrect answers for a certain type of student were kept in order to be used on the EE (J. White, personal communication, August 19, 1998). Once these questions were gathered and other new questions added, there were no further attempts to analyze the test or to monitor it. This is unfortunate, as monitoring a new test, according to Weir (1990), is especially important in order to "guarantee that tests are made as valid, reliable and efficient as possible" (p. 40).

Another reason for the discrepancy is that while the population of 100 students mentioned above may seem an adequate number when sampling populations, it merely represented the total number of students participating in the analysis. The population per level

was actually much smaller, the breakdown revealing that in some cases, the populations were extremely small—as can be observed in Table 12.

Table 12

Numbers of Students per Level Used in Gathering Questions

Level	*n*
Lower Beginner	16
Upper Beginner	13
Lower Intermediate	25
Mid Intermediate	22
Upper Intermediate	13
Lower Advanced	8
Upper Advanced	3
Total	100

A further reason for the discrepancy is that a multiple-choice test of what is essentially grammar, vocabulary, and reading comprehension may be an inappropriate placement tool in consideration of the fact that the philosophy of the school focuses on improving the aural/oral competencies of the student. Furthermore, the test-takers at VELI come from a variety of cultures and educational backgrounds. Some may be familiar with the multiple-choice format; others may have had no experience with it at all. Some may have studied in educational systems which emphasize rote learning of rules and facts; other may have studied in systems which emphasize individual interpretations of the world around them. On this point, teachers at VELI have commented that Asian ESL students seem to do well on paper-based tests of grammar but

are mostly unable to speak fluently, whereas Latin American students speak fluently but seem not do well on tests of grammar.

A final reason is the well-documented (in, for example, Hughes, 1989; Weir, 1990; Weir, 1997) problems with multiple-choice testing: the multiple-choice format used in this test and others has its disadvantages, among them the fact that test-takers may have guessed the answers; that they may have deduced the answer by elimination of wrong answers (a different skill from being able to choose the correct answer); that they may have determined the answers to the reading comprehension passages without ever having read the passage; that the format makes it easy for test-takers to cheat; and most importantly, that multiple-choice tests may not be valid as measures of language ability. What Weir (1990) writes concerning this last point is particularly illuminating:

> There is considerable doubt about their [i.e., multiple-choice questions'] validity as measures of language ability. Answering multiple-choice items is an unreal task, as in real life one is rarely presented with four alternatives from which to make a choice to signal understanding. Normally, when required, an understanding of what has been read or heard can be communicated through speech or writing. In a multiple-choice test the distractors present choices that otherwise might not have been thought of. If a divergent view of the world is taken it might be argued that there is sometimes more than one right answer to some questions, particularly at the inferential level. What the test constructor has inferred as the correct answer might not be what other readers infer, or necessarily be explicit in the text. (p. 45)

In spite of the problematic nature of the this part of the test, the EE serves two extremely useful purposes. First, it functions as a filtering or sorting device during the initial stages of registration day. After the EEs are marked, the tests are sorted by level and matched to the registration cards of the test-takers, with up to eight tests-plus-registration cards being assigned to each TI. This filtering allows administration to assign interviews to TIs on the basis of their level of expertise: instructors usually teach only certain levels and as a result, they are felt to have the most expertise interviewing at the levels they teach. For example, instructors who teach at the Upper Intermediate level will interview test-takers whose EE indicates their level as Upper Intermediate.

If there were no filter, presumably all the teachers at the institute would have to have great knowledge of and to be proficient in all eight levels. The time necessary for that to be accomplished is simply impossible; teachers at VELI estimate that it takes at least *four to six months* to become comfortable teaching at just one level. Consequently, were there no sorting device, the only teachers who could interview would be greatly experienced teachers who had taught at each level of instruction—a decidedly small number, given the fact that faculty or staff at any place of employment is in a constant state of turn-over due to holiday time, release time, pregnancy, sickness, attrition, and so on.

The second extremely useful purpose that the EE serves is the fact that it *looks* like a test. All interested parties—agents, test-takers, teachers, staff, administrators—recognize it as a test. In other words, the EE has a great deal of what Weir (1990) refers to as face validity, or what Bachman (1991) dismisses as face appeal. While the value of face validity is debatable—after all, it is highly subjective, according to Bachman—nonetheless, it is important, because, as Weir points out, "if a test does not have face validity though, it may not be acceptable to the students

taking it .... If the students do not accept it as valid, their adverse reaction to it may mean that they do not perform in a way which truly reflects their ability" (p. 26).

Part Two

The findings reported that most test-takers were able to complete the first section of the writing test and that there were few differences in the level of English used regardless of native language, ability, or final placement, thus rendering this section of Part Two of doubtful utility in helping determine level. The reason that this section did not present much of a challenge to the test-takers may be that the verbs tested were basic ("have," "call," "go," "see," "get," and "do"), most of which lending themselves to the simple subject-verb-object sentences that the test-takers wrote. Perhaps the test-takers chose to write such short sentences because of the eight-minute time limit for the two sections; faced with two sections to complete—the first a guided sentence writing exercise, the second an open-ended essay writing exercise—the test-takers possibly attempted to complete the first section in as short a time as possible in order to have more time to concentrate on the second section.

The findings also reported that test-takers had difficulties completing the second section. There are many reasons for the test-takers' lack of ability to finish this section. One is the position of this section of the writing test in relation to the other section and Part One. That is, after listening to the introductory speech of important information regarding the school and after spending 30 minutes on Part One and then a portion of eight minutes on the first section of Part Two, the test-takers may have experienced some sort of mental exhaustion that precluded their finishing this section. Another reason is that the eight-minute time limit may simply be too short to expect the test-takers to produce any writing of value. The fact that most students were able to

complete the first section but not the section is telling in this regard; it suggests that most test-takers, regardless of whether they finished the first section as quickly as possibly, more than likely spent much of the time limit on completing the first section. In regards to time limits themselves, Weir (1990) points out that "time pressure is often an unrealistic constraint for extended writing and writing timed essays is not normally done outside of academic life. For most people the writing process is lengthier and may involve several drafts before a finished version is produced" (p. 61).

A further reason is, as stated in the findings, that the directions and the strip of cartoon-like drawings likely confused the test-takers. This confusion is an example of what Kirschner, Spector-Cohen, and Wexler (1996) describe as a breakdown in communication, which "means that the students' comprehension of the question is undermined and that they therefore cannot perform the required task in the most efficient manner. This is because they must first work through a series of cognitive hurdles before completing the task" (p. 91). In other words, this confusion may have led to the test-takers' spending time trying to interpret what to do in order to complete the exercise: the exercise is to write a paragraph of comparison about the weather, according to the directions, but how can that comparison fit in with the drawings? A final reason is that the test-takers have differing backgrounds, cultural knowledge, and level of interest in the topic, all of which may affect how much they write or whether they choose to write at all. Thus, while weather may be a topic of considerable interest to those who live on the west coast of Canada, it may not be a subject of much interest for those who come from other parts of the world.

As with the EE, in spite of the problematic nature of Part Two, it does serve a useful purpose. The writing test also has a great deal of face validity. As Weir says, "the essay has

traditionally been accorded high prestige as a testing technique [and] it is a familiar testing technique to both the candidates and the users of test results. It thus has a superficial face validity in particular for the lay person" (p. 60).

Part Three

The findings examined Part Three from the point of view of the TI and reported that all of the respondents to the questionnaire given to the VELI faculty and administration felt that it was very useful or somewhat useful. This result is not surprising in that teachers interview test-takers who have been sorted into groups at the levels the teachers instruct. It is to be expected, then, that as experts at their own levels, the teachers should be able to assess test-takers somewhat accurately, especially those test-takers who have been pre-arranged into groups which the teachers might expect to be studying in their levels. It is also to be expected that since the philosophy of the school in Vancouver is one of an aural/oral focus, the interview would be the best method of assessing the listening and speaking abilities of the test-takers.

The findings also reported, however, that the teachers themselves brought up the point that differing competencies on their own behalf may be responsible for misplacements. Interestingly, this idea concerning competency varying from interviewer to interviewer may really be one of reliability, as noted by Weir (1990) and mirrored by Underhill (1991) in the section of his book in which he examines teacher assessment. Weir comments that "the problems of assessing speech reliably are even greater than those for assessing writing" (p. 74), and Underhill, in noting that the problems with teacher assessment revolve around reliability, makes two important points applicable to the situation at VELI. First, he states that "the more people involved in an assessment programme, the more difficult it is to be confident that the

results are comparable" (p. 28). At VELI, depending on enrollment and staffing levels, there are between 25 to 35 teachers in any one month, all of whom must participate in interviewing, with about half interviewing one month, the other half interviewing the next, and so on. It is indeed hard to be certain that the results from all these teachers are equivalent. Second, Underhill notes that "it may be difficult to arrange thorough training for all the teachers involved. Good teachers do not necessarily make good assessors" (p. 28). There is, for practical reasons related to constraints of time and money, little training at all in interviewing at VELI. Although teachers new to the institute are not expected to participate in interviewing until their fourth month at the school, presumably to give them time to acclimatize to the level at which they teach, the only real interview training they receive is to observe an experienced TI conducting interviews during one registration day.

SUMMARY

Placement testing and the placement process at VELI have been shown to be problematic. To review, the first part of the test, the EE, was discovered to be inaccurate and of doubtful usefulness for many reasons, among them, an unsophisticated statistical analysis and problems inherent in multiple-choice testing. The second part of the test, the two sections of writing, was also found to be of questionable utility for a variety of reasons: e.g., the first section did not discriminate among students at different levels, while the second caused confusion among the test-takers, leading to their having difficulty in its completion. The third part of the test, the interview, was found to be the strongest of the three, with teachers reporting that it was useful, yet it was still a cause for concern owing to the perceived varying competencies of the interviewers.

CHAPTER VI. RECOMMENDATIONS AND CONCLUSION

This thesis drew two sets of important implications from the findings, one for VELI in specific and the other for those involved in placement testing in general. The first implication, while specific to VELI and not to be generalized, can nonetheless serve as a stimulus to others working on placement testing. It is an examination of the insights to be gained, based on the findings of the three questions above and on the reflections of the researcher, as to what can be done in the future at VELI to refine the placement process or to contribute toward more accurate placements. In addition, it explores the constraints that might exist. The second implication, more generic in nature, can serve as a guideline for any institution working on placement testing. It is a review of steps and recommendations, based on the literature in the field, for other language institutes or programs to use in the creation of useful, accurate placement tests.

Thus, this chapter will discuss, in relation to the two sets of implications, both recommendations for a more accurate placement process at VELI and a set of steps and recommendations for accurate placement for other language institutes or programs doing in-house placement testing and placement test creation. In addition, this chapter will conclude with a summary of the thesis and an exploration of other implications.

RECOMMENDATIONS FOR PLACEMENT TESTING AT VELI

The VELI Placement Test should be Modified

To begin with, for the most part, the VELI placement test should be retained, albeit with some modification of its subtests. Although Weir (1990) rightly finds fault with the type of discrete point testing that is found in much of the current VELI placement test, and although he

advocates that direct extended writing tasks be adopted because of their "greater construct, content, face and washback validity" (p. 58), one reason for retaining it in more or less its current form is that, as a test of grammar, vocabulary, and reading, in this particular context, the current placement test carries a high degree of face validity, an extremely important consideration for privately-owned English language institutes, especially those which want to be seen by students and agents abroad as "serious" schools—in other words, those in which the students will have to work diligently to improve their abilities, as opposed to those schools offering only "conversation" courses. Another reason for the retention of the placement test is that at VELI, there still remains the need to sort the test-takers to allow administration to assign interviews to teachers on the basis of their level of expertise. Arguments in favor of keeping the placement test aside, if the test is to be retained, it must better reflect what is taught in each of the levels of the institute, i.e., the content of the test must be related to the curriculum "so that the reasons for separating students into levels in the program are related to the things that the students can learn while in those levels" (Brown, 1995, p. 122). Thus, if the placement test is to be retained, it must be revised. In addition, it must be renamed to distinguish it from its predecessor. For the purposes of this thesis, the revision of the VELI placement test will from this point forward be known as the Vancouver English Language Institute Placement Test (VELIPT).

Type of Test

If the placement test is to be rewritten, the first point for consideration is what kind of test should be chosen: a criterion-referenced test (CRT) or a norm-referenced test (NRT). Brown (1995) says that since the purpose of CRTs "is to assess the amount of knowledge or material known by each individual student, the focus is on individuals rather than on distributions of

scores" (p. 115). Brown also points out that, in contrast, the purpose of NRTs is to "generate scores that spread the students out along a continuum of general abilities or proficiencies in such a way that differences among the individuals are reflected in the scores" (p. 115). It is conceivable that on a CRT, all test-takers, if they know the material, could score 100 percent. While CRTs are mostly considered to be inappropriate as placement tests, which need to spread test-takers "out over a wide range of scores so that they can be sorted as efficiently as possible into class groups" (Harrison, 1983, p. 24), they have been used successfully in placement testing (Brown, 1989). However, in the case of VELI, the use of a CRT for the VELIPT would demand the specification of detailed performance criteria for each level at VELI and for each VELI special skills class—criteria which currently simply do not exist in any detail. While some specifications at VELI do exist, they are minimal, the only one currently existing being that related to what is known as the grammar curriculum. For this reason, while a CRT may be desirable in the long term, until detailed specifications are drawn up, a NRT is the better choice for the VELIPT.

Test Design

If it is to be a NRT, the next point for consideration is test design. Brown (1995) and Harrison (1983) agree that the test should be general and that the "main language skills" (Harrison, p. 26)—listening, reading, writing, and speaking—should be examined in subtests, which according to Brown tend to be relatively long at up to 50 test items each. This number of items per subtest, however, is inappropriate for the situation at VELI, for the reason that a placement test, according to Harrison, should provide results quickly; it is doubtful that four subtests with up to 50 questions each could yield results in a convenient period of time for the

administrators in Vancouver. The VELIPT, then, should be designed to have a number of relatively short subtests.

What to Test

If the VELIPT is to contain a number of subtests, the next point for consideration is what to test. The subtests should specifically focus on discrete grammar points, reading, listening, and writing. The interview, the third part of the testing process, will also function as a subtest, except as one of speaking, in the same way as it does in the current testing process at VELI. As a result, while the interview or speaking subtest is currently considered a separate entity, being conducted in a different room at a later time, it is nonetheless included as a subtest of the VELIPT for the sake of clarity and organization.

The Grammar Subtest

The first subtest of the VELIPT should be one of grammar, similar to what is found in the first part of the current test. Rea-Dickins (1997) notes that "the testing of grammar continues to feature as a component of many school examinations" (p. 87). By extension, the testing of grammar is expected by students, especially upon their entering an English language institute for the first time. Consequently, a discrete point grammar subtest has been retained because, even though the testing of grammar has fallen into desuetude, there remain some good arguments for its retention:

> One of the reasons why much grammar testing still reflects the best practice of the 1960s is that high reliabilities are appealing to educational decision makers. A second is that

> any move away from the objective decontextualised and decomposable approach to grammar testing raises certain difficulties 'Communicative' testing . . . places greater demands on teachers [and additionally, on test writers and administrators] and challenges their competence in English. More open ended writing tasks, through which grammar may be tested, require a new set of skills for test design, format, and item writing, with implications for more explicit marking schemes, e.g. the appropriate design and application of different rating scales. (Rea-Dickins, 1995, p 93).

As applied to the situation at VELI, the resources are simply not available to accommodate the type of communicative test mentioned by Rea-Dickins above. It would be time-consuming to design such a test, difficult to develop rating scales or rubrics, troublesome to mark effectively within a limited length of time, and burdensome to assess for reliability and validity. In any case, as mentioned, the discrete point grammar test should be retained, albeit with one exception: the testing of vocabulary, currently a separate section of Part One, should be part of the reading subtest.

The Reading Subtest

The second subtest of the VELIPT should be one of reading. Of primary concern is that unlike that of the current EE, the texts found in the reading subtest should be as authentic as possible. On the topic of constructing authentic reading tests, Weir (1997) comments that "although full genuineness of text or authenticity of task is likely to be unattainable in the second language reading tests we develop, we still need to select appropriate texts, to be read for realistic purposes, and we expect the reader to extract an agreed level of meaning under specified

performance conditions" (p. 39). In addition, because of the belief that a variety of test formats is best (Weir, 1990), rather than retaining the current four similar reading passages followed by comprehension questions, a number of techniques may better serve to assess the test-takers' reading ability. It is thus recommended that the reading subtest of the VELIPT comprise one authentic reading passage along with multiple-choice comprehension questions, a selective deletion gap filling passage, and a reading passage along with short answer questions. The reading passage plus multiple choice comprehension questions is a common testing technique (Courchêne, 1995) and thus has the value of recognition face validity. In addition, it is convenient for assessing vocabulary, which as mentioned above, has been moved to this section. The gap filling passage, while restricting test constructors to a limited range of enabling skills (as defined by Weir, 1990, p. 48), allows them to select items for deletion "based upon what is known about language, about difficulty in text and about the way language works in a particular text" (Weir, p. 48). The reading passage along with short answer questions has some disadvantages, such as difficulty in marking—especially in questions requiring inferencing skills; nevertheless, Weir (1990) recommends this technique, characterizing it as useful for testing reading comprehension. In addition, this technique also lends itself to the testing of vocabulary.

The Listening Subtest

The third subtest of the VELIPT should be one of listening. At VELI, listening is currently not tested directly (albeit perhaps indirectly through the interview) and is sorely lacking. Similar to the reading subtest, of primary concern is that the listening subtest be as authentic as possible. O'Malley and Pierce (1996) advise that "listening activities should

provide students with opportunities to hear and attempt to decipher language representing, as much as possible, that which occurs in the real world" (p. 60). Buck (1997) agrees, recommending that "texts should be as realistic as possible" (p. 70). Weir (1990) also concurs, stating "where possible listening tests should include an authentic performance task" (p. 58). The construction of a listening test is problematic; there is no one way that is best for all testing situations (Buck, 1997; Weir, 1990). Thus, according to Buck, compromises will have to be made, and according to Weir, there should be a balance of testing types (for example, discrete point, integrative, communicative). With these points in mind, it is recommended that the listening subtest of the VELIPT comprise one authentic taped listening passage together with multiple-choice comprehension questions, a taped listening passage along with short answer questions, and an information transfer passage. Information transfer, as defined by Weir, is the situation in which "the information transmitted orally is transferred to a non-verbal form, e.g., by labeling a diagram, completing a chart or numbering a sequence of events" (p. 50). The listening passage with multiple choice questions has the same advantages discussed in the reading comprehension subtest section of this paper, and the multiple choice questions have the same disadvantages as discussed in the grammar subtest section. The listening passage with short answer questions has the advantage of being realistic (Weir, 1990) and of the certainty that answers are not the result of chance, as they might with multiple choice or true/false questions. The information transfer passage has the advantage of its being "a realistic task for various situations and its interest and authenticity gives it a high face validity in these contexts" (Weir, p. 50).

The Writing Subtest

The fourth subtest of the VELIPT should be one of writing. In the current situation at VELI, the writing test is the second part of the placement test. However, the writing assessment should be moved physically to function as a subtest of the VELIPT rather than existing as a distinct part on a separate sheet of paper. While this change is for the most part superficial, the current separation may be interpreted by test-takers as the writing assessment's being an afterthought and therefore not especially important, especially since the writing test is on the reverse side of the answer sheet. As a result, making the writing assessment a subtest of the VELIPT should serve to emphasize the fact that three of the four skills (listening, reading, and writing) are being tested together because they are ostensibly equally important.

The construction of the writing test is, as was that of listening, problematic; as mentioned earlier in this thesis, due to time and labour constraints related to test development and marking, a detailed writing component would be difficult to mark effectively during registration day. Regardless, some recommendations can be made to enhance the writing assessment as it stands now and to increase its usefulness to VELI. The first recommendation deals with the first section of the writing test, that in which the test-taker must convert the given verb to the past tense and then write a sentence containing that verb. It was found that this part does not serve much of a purpose; it should as a result be deleted. The second recommendation focuses on the second section, that in which the test-taker must write a paragraph of comparison about the weather. It is confusing because of the strip of cartoon-like drawings along the left side of the page, as explained earlier. To eliminate the confusion, one of the two components of the second section must be removed: either the writing topic or the drawings. If the writing topic is to be eliminated, the directions must be made clear that the test-takers are to write a story about the

drawings, which themselves must clearly present some sort of story. Admittedly, the use of drawings in guided writing exercises such as this one presents difficulties; drawings can be interpreted differently depending on culture, and there is no guarantee that all test-takers will be able to interpret the drawings in order to come up with something to write. If the drawings are eliminated, more than one topic must be offered, to account for the fact that not all test-takers will have enough of an interest in just one topic to write something about it. The third recommendation is concerned with the time limit. The current limit of eight minutes is simply too short. While at least thirty minutes for a writing test is perhaps desirable, time is at a premium during registration day at VELI. Thus, it is suggested that the time limit be increased to a minimum of fifteen or perhaps twenty minutes. The last recommendation deals with marking. While detailed marking of the writing subtest of the VELIPT is impractical given the context in which this testing occurs, it might nonetheless serve the interests of the school better if the writing subtest were marked, if only at a cursory level. The way to go about marking would be to have most of the test marked in Room 3, with the exception of the writing subtest. Once the tests have been sorted, before the TIs take all the information out to the classroom for the interview, they could quickly mark the compositions based on a simple yet clear rubric. Because the tests will have been sorted into approximate levels at which the TIs teach, the essays may be easier and therefore less time-consuming to mark, given the fact that the TIs should be familiar with the writing skills of the VELI levels at which they teach.

The Speaking Subtest

The fifth and final subtest of the VELIPT should be one of speaking, to be tested as it is currently in the environment of an interview. Although the interview as it stands is considered to

be the strongest part of the VELI placement testing process, some recommendations can nonetheless be offered to increase its accuracy in helping TIs place students into the correct level of instruction. Presently, there is some doubt as to the competency of the TIs. In addition, there is some question about the usefulness of the three interview aids available to them; TIs may use one or two or three—or none at all—in assessing the speaking ability of the test-takers. In light of these two points, one recommendation is to standardize to as great a degree as possible what is done during the interviews. Weir (1990) refers to one such procedure as the "controlled interview" (p. 76), and outlines its advantages, among them the fact that because test-takers will be asked the same questions, it will be easy to make comparisons across performances, that the procedure has a high degree of content and face validity, and that intra-rater reliability can be high, given proper training of the raters. Although it is still questionable whether each TI would follow the procedure accurately, nonetheless, some standardization, at least, is better than none at all. It follows that with standardization, there must be a better assessment system. Thus, the next recommendation is to devise a comprehensive marking scheme. The final recommendation is to train the interviewers adequately. Weir summarizes these points by noting that "in oral testing, as in the assessment of written procedures, there is a need for explicit, comprehensive marking schemes, close moderation of test tasks and mark schemes, and rigorous training and standardisation of markers in order to boost test reliability" (p. 80).

A Detailed Statistical Review should be Performed

If the VELIPT is to contain the aforementioned subtests, the next point for consideration is a statistical review of the test. Where possible and convenient, a review should be performed, for "one fundamental concern in measuring anything is that the results should be the same every

time you measure it" (Brown, 1991, p. 98). To that end, both Harrison (1983) and Brown (1991) detail methods for statistical review. Harrison is a little more practical of the two, specifically describing both how to calculate distribution, mean, standard deviation, and reliability, and how to interpret the results. In addition, he offers methods for item analysis and item discrimination. Brown is rather more theoretical, describing in general terms reliability, standard error of measurement, test validity, and construct validity. In any case, to return to the VELIPT, not all of its subtests are particularly suited to a statistical review; for instance, given the constraints of time and labour, it would be difficult to review the writing and speaking subtest. However, the discrete point grammar subtest lends itself well to a statistical review, as does the reading and listening subtests—of course, depending on what type of test method is employed.

Training should be Offered

If the VELIPT is to be rewritten as a NRT containing a number of subtests which have been reviewed statistically, the last and perhaps most important consideration of all is proper and rigorous training. The reason, of course, is that with subjectively scored tests, such as the proposed writing and speaking subtests of the VELIPT, there is a distinct possibility of rater subjectivity and a lack of agreement with other raters, points which can lead to inconsistent and unfair assessments (O'Malley & Pierce, 1996). In order to combat this problem, in writing about developing and using authentic assessments, O'Malley and Pierce emphasize that "there is no substitute for effective professional development" (p. 21), development in this case being rater training and the establishment of clearly defined rubrics or scoring criteria. Weir (1990) agrees: "considerable attention should also be paid to the development of relevant and adequate scoring criteria and examiners must be trained and standardised in the use of these" (p. 86).

At VELI, there is little training, except that mentioned previously in reference to the interviews, with the result being an opportunity for an inaccurate testing process possibly leading to unreliable placements. With a greater emphasis on training, there should be a greater consistency of rating and therefore fewer opportunities for misplacements. In addition, any questions concerning varying competencies of raters, as was the case with the interviewing teachers, should be reduced if not eliminated: everyone should be starting off with the same degree of preparation.

Therefore, it is recommended that VELI devise comprehensive rubrics for whichever parts of the VELIPT that require subjective assessment. Along with the rubrics, there must be clear examples of performances to supplement the ratings. In addition, all teachers must receive training in marking every aspect of the test, whether objective or subjective, more than just initially at the time of hiring as is done currently. Training in objective marking, of course, may be done infrequently, as the technique is straightforward and there should be little to no cause for concern in inter-rater reliability. Training in subjective marking, however, should not just be a one-time proposition; it must be on-going. It is therefore recommended that prior to every registration day, the rubrics should be reviewed, and supplemental examples must be discussed.

STEPS AND RECOMMENDATIONS FOR PLACEMENT TEST CREATION

The second set of important implications drawn from the findings is a review of steps and recommendations, based on the literature in the field, for other language institutes or programs to use in the creation of useful, accurate placement tests. In order to explore this implication, detailed steps will be offered focusing on how to go about the creation of a placement test, followed by a set of recommendations focusing on individual parts of the placement test.

A key question to be asked at this point is why there is a need for a set of steps and recommendations. One reason is that, at institutes such as VELI, placement testing has evolved slowly and almost haphazardly in response to the external pressures of an increasing capacity for students and a widening selection of classes into which to place those students. The result of this is that, at such institutes, there is often no clear-cut plan for placement testing, and those tasked with producing such a test invariably have little training in placement test creation. Thus, the accuracy of a placement instrument created under these conditions is bound to be questionable. In these situations, then, a set of steps and recommendations would be invaluable.

Another reason is that while guides containing steps and recommendations for placement testing do exist, they tend to be somewhat outdated (Harrison, 1983) or mentioned briefly as part of an overall discussion of other matters (Brown, 1995; Hughes, 1989). Another reason is that the steps and specifications for the creation of any test are "a central and crucial part of the test construction and evaluation process" (Alderson, et al., 1995, p. 9), and that they are needed by a wide variety of people, such as the constructors of tests, the users of tests, the test-takers, teachers, administrators, students, and those responsible for establishing test validity (Alderson, et al., 1995). A final reason is that the steps and recommendations to be presented in this thesis should offer to those involved in devising placement tests a much needed systematic or rational basis for developing their tests—in other words, a guideline, or what Lynch and Davidson (1994) refer to as a "blueprint" that test writers and test administrators can use in the creation and administration of their placement tests. As Brown (1995) points out, "though all this may seem like a great deal of work, remember that in most language programs, any rational approach to testing will be a vast improvement over the existing conditions" (p. 119) and "the work is

worthwhile because of the information that can be gained and the satisfaction that can be derived from making responsible decisions about students' lives" (p. 123).

Steps to Take in the Creation of a Useful and Accurate Placement Test

Assemble an Assessment Team

Before the creation of a test can even begin, some preliminary steps are necessary. In discussing authentic assessments, O'Malley and Pierce (1996) suggest that the first step should be to assemble an assessment team. It matters not whether the assessment is to be authentic or otherwise: the idea of bringing together interested parties is an important one. This is the time to address individual stakeholders and their concerns on how to go about constructing the test. As Buck (1997) notes, "when designing tests, everything depends on the purpose of the test, and the decisions that need to be made regarding the test-takers' ability. There will be advantages and disadvantages with any design, and compromises will usually be necessary" (p. 71). Assembling an assessment team should help the creator of a placement test to define the purpose of the test and to arrive at any decisions regarding the test. Concerning the composition of the assessment team, it should consist of any administrators who are responsible for curriculum and for students, of coordinators who are responsible for the implementation of the curriculum and the like, of teachers who represent a cross-section of the classes offered at the institute, and even of students at different levels within the school system. Noting also that test design involves compromises, Bradshaw (1990) states that "there seems to be no reason why some degree of collection of test-takers' and test-users' reactions cannot be included as part of the design of any new test" (p. 27).

Define Characteristics of the Test-takers

After the objectives have been defined, the next preliminary step is to describe what type of test-taker will be taking the test. Carroll (1980) refers to this step as one of "participant identification" (p. 19) and includes it in the first of his recommended three phases of test construction. In identifying the test-taker, Carroll includes "relevant information about his identity and language background, such as his age, sex, nationality and place of residence as well as target language [and] mother tongue and any other languages learnt" (p. 19). Alderson, et al. (1995) include information on test-taker characteristics, such as age, gender, stage of learning, first language, cultural background, country of origin, type of education, reason for taking the test, personal and professional interests, and amount of background knowledge (p. 12). Having access to this information will help the test creator greatly both in choosing appropriate material and test techniques and in avoiding some of the flaws mentioned previously in this thesis, those for example experienced by Ilyin (1970).

Define Objectives for the Placement Test

After the test-takers have been defined or characterized, the next preliminary step is to define objectives for the test. Hughes (1989) notes that this step is essential in testing "to make oneself perfectly clear about what it is one wants to know and for what purpose" (p. 48). Harrison (1983) believes that objectives for placement tests are different from those for other tests "because placement tests cannot be geared to the learning which went before" (p. 26). He suggests that test creators should think in terms of "aims," which Harrison says are more general than objectives. Semantics aside, the important point here is for test creators to decide what to test and how to go about doing so. Although O'Malley and Pierce (1996) focus on authentic

assessment, their belief that this step should encompass the determination of the purposes of the assessment and the specification of objectives is applicable really to any type of testing. In specifying the objectives, O'Malley and Pierce suggest—like Harrison—that objectives should be obtained from, among other sources, curricula.

Decide on the Type of Test to be Used and Its Contents

Once the objectives for the test have been outlined, the next preliminary step is to decide what type of test is to be used and what to include in its contents. Concerning test type, three decisions must be made. First, should the test be direct or indirect, or a combination of the two. Second, should it be discrete point or integrative? Third, should it be norm- or criterion-referenced? Direct testing involves requiring the test-taker to perform the skill or skills to be measured; indirect testing involves measuring the abilities underlying the skill. Hughes (1989) believes that while "it is preferable to concentrate on direct testing" (p. 16), he does admit that for some types of testing, indirect testing can be useful. Hughes observes that "direct testing is easier to carry out when it is intended to measure the productive skills of speaking and writing" (p. 15), and that indirect testing offers "the possibility of testing a representative sample of a finite number of abilities which underlie a potentially indefinitely large number of manifestations of them" (p. 16). With Hughes' observations in mind, for placement testing, it is recommended that a combination of the two approaches be used, with direct testing for speaking and writing, and indirect testing for listening and reading.

Discrete point testing involves testing one thing at a time, item by item; integrative testing involves testing the combination of many elements in the completion of a task. Hughes notes that the distinction between the two "is not unrelated to that between indirect and direct

testing [and indeed] discrete point tests will almost always be indirect, while integrative tests will tend to be direct" (p. 17). Again, with Hughes in mind, it is recommended that for a grammar part (if included) and a listening part of a placement test, discrete point testing should be employed, while for a speaking part and a writing part, integrative testing should be employed. For reading, a combination of the two should be employed.

NRTs and CRTs have been explained previously in this thesis. Although Brown (1989) has successfully used a combination of NRT and CRT in what he has called a "new strategy for constructing language placements" (p. 73), NRTs are most often used for placement tests, for reasons discussed earlier, and as such, are recommended for most placement situations. Nonetheless, if detailed criteria for classes or levels are already in place, CRTs certainly offer a viable alternative to NRTs in placement testing, and should be considered.

Concerning test content, it is logical that the placement test should reflect the curriculum of the school. In discussing the general development of language tests, Brown (1995) suggests that "a program-specific placement test could be developed so that the reasons for separating students into levels in the program are related to the things that the students can learn while in those levels" (p. 122). In practice, however, test items do not always mirror what is actually taught in class, and an example of this point is found in Brown's preamble to one of his earlier journal articles: "We decided to develop a placement battery that would be related in content to the curriculum of our institute—a proposal that struck us as strangely novel" (1989, p. 66).

A number of authors have proposed recommendations to assist test creators in deciding what to include in their tests. For example, Alderson, et al. (1995) recommend that at this stage, test creators ask themselves a variety of questions, such as how many sections the test should have, how long the sections should be, and how they should be differentiated; what the target

situation is for the test and whether it should be simulated in some way; what text types should be chosen (written and/or spoken); what language skills should be tested; what language elements should be tested; what sort of tasks are required; how many items their should be in each section; and what test methods should be used. Chandavimol (1988) recommends that "the content of the placement test should directly reflect the parameters of the English programme concerned" (p. 3). Harrison (1983) advises both that the "contents of a placement test should be general" (p. 24) and that "the tests themselves should be fairly short, so that they do not take too long to answer or to mark" (p. 27). Most importantly, he recommends that "all four of the main language skills (listening, reading, writing, and speaking) should be tested" (p. 27).

In summary of these points, then, it is recommended that a NRT be used for a placement test and that depending on the context of school at which the test is to be employed, the test focus on the main language skills as they reflect the curriculum of the school.

Create the Test

After the preceding preliminary steps, the first main step is to create the test itself. This step is important, for as Kirschner, Spector-Cohen, and Wexler (1996) indicate, "test questions constitute a communicative interchange between the test writer and the test taker" (p. 89). As such, then, the test creator must devise the test in such a way as to be "as easy for test takers to process as possible" (Kirschner, et al., p. 89).

Once this point has been understood, the test creator must then continue with creating the test and deciding on the parts of the placement test. As mentioned previously, listening, reading, writing, and speaking should be tested. In addition, a grammar component should be considered if the curriculum of the school places emphasis on grammar. In any case, in order to assist those

who are tasked with the creation of a placement test, a brief set of general specifications—recommendations, really—for each part of a placement test follows these steps in a separate section. Of course, depending on the testing context and other considerations, not every placement test will include all the recommendations listed here; as Brown (1995) warns, "many language tests are, or should be, situation specific" (p. 119). Nonetheless, for the sake of completeness, recommendations for each of the individual parts of a placement test have been included in this thesis, and it is suggested that those who are involved in the creation of placement tests use only those recommendations that apply to each individual testing situation.

Develop Rubrics or Rating Scales for the Test

The second main step is to develop rubrics or scoring guides for the placement test. Doing so should contribute to the reliable scoring of samples of the test-taker's performance. Although Harrison (1983) refers to rubrics in terms of "information for the student on how to do the test, including instructions, examples, and the organisation of test procedures" (p. 142), rubrics are taken here to refer to scoring scales that assign a numerical value to a test-taker's performance depending on the extent to which it meets pre-designated criteria (O'Malley & Pierce, 1996). As such, they are applicable to subjective or open-ended parts of the placement test, such as those containing short essay answers or oral interaction, and can be either holistic or analytical.

Holistic scoring "involves the assignment of a single score to a piece of writing on the basis of an overall impression of it" (Hughes, 1989, p. 86) and has the advantage of speed: Hughes notes that experienced scorers can assess a one-page piece of writing in "just a couple of minutes or less" (p. 86). One caveat concerning holistic scoring, however, is that the scoring

scale must be very well conceived. Hughes points out that the rubric must "be appropriate to the level of the candidates and the purpose of the test" (p. 87). A second caveat is that there must be more than one scorer in order to ensure a high degree of scorer reliability. Analytical scoring requires "a separate score for each of a number of aspects of a task" (Hughes, p. 91) and has a variety of advantages. The most important of these are that scorers must consider certain aspects of the test-taker's performance that they might otherwise miss, that the results can be used for diagnostic purposes, and that "the very fact that the scorer has to give a number of scores will tend to make the scoring more reliable" (Hughes, p. 94). The main disadvantage with this type of rubric is that analytical scoring is time-consuming.

Which of the two types of rubrics should be developed by the creators of a placement test? Test creators must assess their testing situation and decide which to use. If time is at a premium, it is recommended that holistic scoring be used, for the reason that it is much more time and resource efficient, in that placement testing of objective items alone can be time-consuming, and testing is often done on-site with staff and/or faculty of the school in the role of test administrator and/or scorer. If there are enough time and resources, however, it is recommended that analytical scoring be used, for the reason that it can be the more reliable of the two and that the results can be used as a diagnostic tool by teachers of the classes into which the test-taker may be placed.

Test the Test

The third main step is to analyze the newly created placement test, a step to which Alderson, et al. (1995) refer as pretesting and analysis. They state that "it is essential . . . that all tests should be pretested" (p. 74), because regardless of the care with which the placement test

has been created, serious problems may exist with the test that cannot be identified during its conception. Harrison (1983) agrees, stating that "pretesting items is often regarded as essential . . . because trying them out with students shows how they work in practice, and it is only from this experimentation that bad items can be identified and amended or thrown out" (p. 127). Examples of problems with test items that may be identified at this stage are, according to Alderson, et al., (a) an abundance of items used in the test may be too difficult or too easy; (b) open-ended test items may confuse test-takers; (c) essay tasks may unintentionally result in less than adequate responses from the test-takers; and (d) multiple-choice items may be ambiguous and therefore open to disagreement (p. 74). Any one of these problems could serve to cause the placement test either to yield inaccurate results or not to work as intended—to spread students out on a continuum of language abilities.

Alderson, et al. thus suggest that the newly created test be trialed in exactly the same way as the final test will be on a number of students who are "representative of the final candidates, with a similar range of abilities and backgrounds" (p. 76). How many students are considered to be enough? As the authors point out, it is often difficult to find large numbers of students, so "the only guiding rule is 'the more the better,' since the more students there are, the less effect chance will have on the result" (p. 75). Once the test has been trialed, it should be analyzed. The authors suggest that objective test items, such as those of the multiple-choice variety, should be analyzed in terms of the facility value, which measures the level of difficulty of an item, and the discrimination index, which measures "the extent to which the results of an individual item correlate with results form the whole test" (Alderson, et al., 1995, p. 80)[8]. The authors also

[8] For an explanation of how to calculate the facility value and/or discrimination index of test items, see for example, Alderson, et al., pp. 80-86, Hughes (1989), pp. 161-162, or Harrison (1983), pp. 127-133.

suggest that subjective test items, such as those of the essay variety, should be analyzed in terms of "whether the items elicit the intended sample of language; whether the marking system . . . is usable; and whether the examiners are able to mark consistently" (p. 86).

Train the Scorers and Administrators of the Test

The last main step is to train the people who are going to be scoring and administering the placement test. As Underhill (1991) observes, "in testing, as in teaching, people are the biggest asset, and like any other resource, they can be used effectively or badly" (p. 15). Weir (1990) agrees that this step is important. He states that "considerable attention should . . . be paid to the development of relevant and adequate scoring criteria and examiners must be trained and standardised in the use of these" (p. 86). Alderson, et al. (1995) advise that

> the training of examiners is a crucial component of any testing programme, since if the marking of a test is not valid and reliable then all of the other work undertaken earlier to construct a 'quality' instrument will have been a waste of time. No matter how well a test's specifications reflect the goals of the institution or how much care has been taken in the design and pretesting of items, all the effort will have been in vain if the test users cannot have faith in the marks that the examiners give the candidates. (p. 105)

Alderson, et al. offer detailed advice concerning procedures for training the scorers of writing and speaking, and discuss the idea of having a Chief Examiner (p. 111) and standardization meetings (p. 112). While the creators of placement tests need not follow such a formalized

method, it is nevertheless recommended that they develop a system that provides scorers with on-going training in the assessment of subjective test items.

Concerning administrators of tests, that is, those people who deliver the test to the test-takers, Alderson, et al. note that "though the training of administrators need not be as complex as that provided for examiners, it is still important that the administrators understand the nature of the test they will be conducting, the importance of their own role and the possible consequences for candidates if the administration is not carried out correctly" (p. 115). It is thus recommended that creators of placement tests also develop a system that provides administrators with training so that the test can be delivered consistently and correctly.

Recommendations Focusing on Individual Parts of the Placement Test

Listening

Buck (1997) observes that

> the basic idea of most listening tests is to assess the ability to use knowledge of the language for the purpose of understanding spoken texts . . . . [T]est tasks must [therefore] require fast, automatic, on-line processing of texts which have the typical linguistic characteristics of spoken language—especially the phonological characteristics. [In order to do so] non-interactive listening tasks . . . are probably most useful and certainly easier to construct. (p. 71)

To that end, Weir (1990) believes that the tasks should be authentic and "in terms of the tasks, items and scoring, it might be desirable in certain components of the test to focus on discrete

items" (p. 52). The testing of listening should be accomplished, according to Harrison (1983), through the use of tape recordings, with the advantage being that the fact that the text is recorded makes it "more authentic, as if the students were actually listening to a radio talk or telephone message" (p. 29). Harrison conveniently omits the fact that not all listening tasks are conducted over the radio or telephone, but in spite of that, his point has merit for a very important reason: The test is more reliable because it is the same for each administration. As Harrison notes, "all students hear exactly the same text throughout all repeats and at all sittings of the test" (p. 29).

To summarize, the listening part of a placement test should seek to assess the ability of the test-taker to understand spoken language. The text should be authentic, and the tasks should be non-interactive. The use of tape recordings is advised, and discrete-point testing is recommended, both for the sake of convenience of administration and marking, and for the sake of reliability.

Reading

Weir (1997) observes that a reading test "should reflect as closely as possible the interaction that takes place between a reader and a text in the equivalent real life reading activity" (p. 39). Therefore, the approach to the reading part of a placement test should be direct. Hughes (1989) believes that there are at least four levels of reading that can be tested: low-level operations, grammatical and lexical abilities, macro-skills, and micro-skills. The term "low-level operations" refers, for example, to the ability to distinguish between letters of the alphabet, e.g., between "b" and "d." According to Hughes, there is no call for the formal testing of this ability in that information on this ability can be observed through informal observation. Grammatical and lexical abilities refer to the ability, for example, to use the present perfect aspect or to define

vocabulary. Information on these abilities can be collected, as Hughes notes, "through tests of grammar and vocabulary, not necessarily as an integral part of a reading test" (p. 117). Macro-skills refers to the ability to scan text to find specific information, to skim to obtain gist, to identify the support of an argument, and so on, while micro-skills refers to the ability to identify referents of pronouns, to use context to guess meanings, to understand transition words, and so on. While a test of macro-skills is possible, Hughes believes that "only at the level of . . . 'micro-skills' do we reach the point where we find serious candidates for inclusion in a reading test" (p. 117).

The text of the reading part of a placement test does not necessarily have to be authentic—a term that Hughes defines as "intended for native speakers" (p. 118). Instead, Hughes suggests that whether or not authentic texts are employed in any sort of reading test depends in part on what the test is intended to measure. Unfortunately, he does not offer any further information on what type of measurement demands what kind of text (authentic or not), but he does state that "even at lower levels of ability, with appropriate items, it is possible to use authentic texts" (p. 118). In light of that point, then, authentic texts are recommended, the types of which, according to Hughes, might include textbooks, novels, magazines, newspapers, journals, and timetables—to name a few. The type may be further specified, such as a two- or three-paragraph passage from a novel, an article from a magazine, or an advertisement in a newspaper.

A number of techniques can be employed in the testing of reading, but Hughes cautions that "we have to recognise that the act of reading does not demonstrate its successful performance. We need to set tasks which will involve candidates in providing evidence of successful reading" (p. 120). The difficulty is, however, employing techniques or tasks which do

so without interfering in the reading itself. Hughes offers a list of techniques, including multiple-choice, short answer, guided short answer, and information transfer. While the advantages and disadvantages of multiple-choice testing are outlined earlier in this thesis, it should be noted that multiple-choice testing is reliable and does lend itself well to rapid scoring. Short answer and guided short answer testing may provide a good indication of reading ability, but both techniques have the disadvantage of the potential for obscuring the test-taker's true ability because each demands the ability to write: to use Hughes example, "a student who has the answer in his or her head after reading the relevant part of the passage may not be able to express it well" (p. 122). Information transfer, on the other hand, has the advantage of minimizing the potential for obscuring the test-taker's ability in that this technique demands little or no writing ability.

To summarize, the reading part of a placement test should seek to assess the ability of the test-taker to understand written language. The text should be direct and authentic. A variety of techniques are recommended: multiple-choice for its ease of scoring, short answer or guided short answer for its indication of ability, and information transfer for its lack of dependence on the test-taker's ability to write. The choice of which technique to employ is a difficult one, and must be decided according to each individual placement testing situation. In light of that point, a combination of the above techniques is recommended.

Speaking

Underhill (1991) observes that "when we test a person's ability to perform in a foreign language, we want to know how well they can communicate with other people, not with an artificially-constructed object called a language test" (p. 5). In considering a test of speaking,

then, Weir (1990) believes that "the essential task for the test designer is to establish clearly what activities the candidate is expected to perform, how far the dynamic communicative characteristics associated with these activities can be incorporated into the test, and what the task dimensions will be in terms of the complexity, size, referential and functional range of the discourse to be processed or produced" (p. 74). Underhill (1991) takes a more humanistic approach to the testing of speaking by stating that "oral tests must be designed around the people who are going to be involved. This is a human approach; we want to encourage people to talk to each other as naturally as possible. The people, not the test instrument, are our first concern" (p. 4). To underscore that point, Underhill continues by suggesting that "the direct interview is the most common and most authentic type of oral test for normal purposes; there is no script and no preparation on the learner's part for any special activity" (p. 31).

Regardless of the type of speaking test, Weir cautions that "in oral testing . . . there is a need for explicit, comprehensive marking schemes, close moderation of test tasks and mark schemes, and rigorous training and standardisation of markers in order to boost test reliability" (p. 80). On this point, Underhill adds that accurately-worded rating scales will be of great benefit to those administering the speaking test (p. 13). In addition, Alderson, et al. (1995) suggest that the person administering the test of speaking is important "because it is always necessary for at least one person to elicit language from the candidate and to react in an encouraging way to keep the language flowing" (p. 116). Underhill echoes this point: "the interviewer should also know a lot about what happens in [the] classes. Ideally, she should be a regular class teacher herself so she knows the classes well and can ask herself questions like, 'How would I feel if this learner appeared in my class tomorrow?'" (p. 13). Also of importance,

according to Alderson, et al., is an environment that will not be intimidating to the test-takers, one "which will help candidates to feel at ease" (p. 117).

To summarize, the speaking part of a placement test should seek to assess the ability of the test-taker to communicate orally with other people, not with a testing instrument. A direct interview is recommended, as it is not scripted and requires no special preparation on the part of the test-taker. However, comprehensive marking schemes should be devised, and precisely-worded rating scales are recommended. The person administering the speaking test should be a regular class teacher, and the speaking test should be administered in a place that is comfortable to the test-taker.

Writing

Hughes (1989) assumes that "the best way to test people's writing ability is to get them to write" (p. 75); thus, the approach to the writing part of a placement test should be direct. The tasks contained in the writing test should, according to Hughes, be "representative of the population of tasks that we should expect the students to be able to perform [and] should elicit samples of writing which truly represent the students' ability" (p. 75), and do not represent other things such as the creativity, imagination, or intelligence of the test-taker. In addition, Hughes maintains that the samples of writing obtained from these tests should be scored reliably. While a large number of writing tasks is seen by Hughes as being ideal in terms of validity (p. 81), it is impractical in a placement test. As Hughes notes, "if it is a matter of placing students in classes from which they can easily be moved to another more appropriate one, then accuracy is not so important; we may be satisfied with a single sample of writing" (p. 82).

While a number of strategies exist for testing writing ability, Weir (1990) offers three viable suggestions: the summary task, the controlled writing task, and the essay test. Summarizing, however, is problematic in that it demands the production of a specific text, one which might be too narrow and thus beyond the knowledge or abilities of the test-taker. The controlled writing task, while necessary "where writing tasks are an important feature of the student's real life needs" (Weir, p. 61), is also problematic in that there may be situations "when the complexity of the stimulus obstructs the desired result, i.e., one needs to understand a very complex set of instructions and/or visual stimuli to produce a relatively straightforward description of a process or a classification of data" (p. 62). The essay test is problematic in that it is open-ended, timed, and time-consuming. In addition, among other problems associated with this form of testing, the ability to write freely on topics "may depend on the candidate's background or cultural knowledge, imagination, or creativity" (p. 60). Nevertheless, in spite of the problems associated with the essay test, it is a traditional method for testing writing ability, is familiar to a wide variety of test-takers, and thus holds much face validity.

To summarize, the writing part of a placement test should seek to assess the ability of the test-taker to write. A direct test is advised, as testing writing through the use of indirect, discrete-point items does not clearly give an indication of writing ability (Weir, 1990, p. 59). The essay test is recommended as the vehicle for testing writing ability, yet caution must be taken in the creation of topics for the essay: test-takers may be hampered in that the topic may be uninteresting or culturally biased. It is therefore recommended that a selection of topics be offered on a variety of subjects, with the test-taker writing on one topic.

Grammar

Rea-Dickens (1997) notes that the communicative approach to language teaching has lessened "the role of grammar as a respectable focus of teaching and learning" (p. 94), yet Hughes (1989) observes that

> there is often good cause to include a grammar component in the achievement, placement and diagnostic tests of teaching institutions. It seems unlikely that there are many institutions, however 'communicative' their approach, that do not teach some grammar in some guise or other. (p. 142)

While the testing of grammar has traditionally been accomplished through the use of multiple-choice items, other techniques are available and may even be preferable; rather than requiring the test-taker solely to recognize correct use, as is the case in most multiple-choice items, these techniques require that the test-taker use grammatical structures appropriately. Hughes lists three such techniques: paraphrase, completion, and modified cloze. Paraphrase requires the test-taker to write a sentence, the beginning of which is supplied, that is similar in meaning to one that is given. Completion requires the test-taker to complete sentences by supplying correct structures in context (for example, interrogative forms in the completion of questions, with the responses already supplied). Modified cloze requires the test-taker to complete sentences by supplying the deleted form (for example, prepositions or articles).

To summarize, because grammar is taught in some way or other, it should be considered as a component of a placement test. The multiple-choice and modified cloze techniques are recommended for their ease of scoring.

CONCLUSION

Summary

This thesis has shown that placement testing, while important and necessary, can be subject to problems, some of which include the removal of the test-takers from their own assessment; the tests themselves, which may be flawed in some way; and the resulting placements, which may be inaccurate because of a combination of the previous problems or because of other failings such as inadequate training of examiners, inferior test design, or poor test administration.

In consideration of these problems, this thesis has sought to describe and analyze the placement testing process at VELI, a large, privately-owned English language institute for adult foreign students in Vancouver, British Columbia, Canada, where placement testing has been perceived to yield inaccurate placements. The placement test at this institute has been examined closely, as has its history and the context in which it is administered, and steps and recommendations have been offered not only to VELI but also to any other institute in the creation of a useful and accurate placement testing instrument.

Other Implications

One implication arising from this thesis is that a measured, thoughtful approach to testing should yield accurate results, albeit through an initial heavy investment of time and perhaps, of funds. The main implication, however, arising from this thesis is that placement testing cannot be viewed as existing separate from the rest of the institute. There are two aspects to this point: the first is that a valid, reliable, and accurate test must be linked somehow to the curriculum, to

what is taught at the institute; after all, a placement test consisting of multiple-choice grammar questions is of little value, for example, to administrators faced with placing students into a program that has an aural/oral approach. The second is that institutional and historical influences can shape placement tests over time, with the result being a gradual movement away from accuracy. What seemed to be and accurate placement vehicle some time ago may not be precise today.

Suggestions for Further Research

This thesis has attempted to describe and analyze the placement testing process at one institute in order to offer to both this institute and others steps and recommendations toward more useful and accurate placements. Much has been covered in this thesis; however, there is still room for further research. One area that bears exploration is self-assessment as a placement tool. LeBlanc and Painchaud (1985) have shown that self-assessment can be a viable placement vehicle, as long as considerable work has been done to ensure that the test-takers (test-participants?) have enough information to make informed choices. LeBlanc and Painchaud's work merits repeating, with an addition: echoing LeBlanc and Painchaud, a comparison of a placement situation similar to that at VELI, in which a "traditional" placement test is used, and one in which informed self-assessment is used, could be conducted in order to determine whether there was a correlation between the results of the two. In addition, however, it would be useful to examine which of the two methods of placement the test-takers preferred. Although face validity can be important in placement testing, LeBlanc and Painchaud only address in passing the issue of whether informed self-assessment is wholly acceptable to the test-takers.

Another area that merits exploration is the beliefs of those instructors, administrative staff, and administration involved in the delivery and supervision of placement tests and the effect these people may have on the placement process. While it has been mentioned, albeit indirectly, in this thesis that TIs have specific beliefs with regard to the testing process, to the test itself, and to the abilities of other TIs, the possible negative effects arising from the fact that TIs may have lost faith in the placement process have only been touched upon. To attempt to ascertain whether those effects result in inaccurate placements would be beneficial.

A further area that is worthy of examination is a comparative look at placement testing at a number of institutes. The approach to placement may be quite different from one institute to the next, dependent upon philosophy, size, curriculum, and so on. These different approaches may give rise to some of the same questions posed in this thesis and, perhaps, to others: Have historical and institutional considerations and limitations affected the tests of other institutes in the same way that those at VELI affected its placement vehicle? How many institutes choose to create their own placement test? Why? Which skills do various institutes examine in their placement test? Why? How do they go about doing so? To examine how students are placed into classes in various institutes and how placement testing in those institutes has evolved may contribute to more accurate placement testing through determining which testing approaches and techniques seem to be the most useful and least inaccurate.

A final area that warrants investigation is the use of computer adaptive testing (CAT) in placement testing. CAT has held for some time the promise of "accurate, efficient, individualized assessment of knowledge and skills utilizing high-speed electronic machines" (Tung, 1986, p. 13). CAT, then, could be a highly desirable tool for useful and accurate placement testing. However, widespread use of CAT in general and in placement testing in

particular is still in a nascent stage, and it is as yet unknown whether its use may contribute to more accurate placements. It would be informative, then, to examine the results of a CAT placement test and those of more traditional forms of placement testing in order to determine whether CAT yields more accurate placements.

A Final Note

Placement testing is important, yet it can be problematic. With thought and careful planning, however, problems can be minimized, and placement testing can be more accurate.

REFERENCES

Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.

Bachman, L. F. (1991). What does language testing have to offer? *TESOL Quarterly, 25*(4), 671-704.

Banerjee, J. & Luoma, S. (1997). Qualitative approaches to test validation. In C. Clapham and D. Corson (eds.), *Encyclopedia of language and education, volume 7: Language testing and assessment* (pp. 275-287). Netherlands: Kluwer Academic Publishers.

Bradshaw, J. (1990). Test-takers' reactions to a placement test. *Language Testing, 7*(1), 13-30.

Brown, J. D. (1989). Improving ESL placement tests using two perspectives. *TESOL Quarterly, 23*(1), 65-83.

Brown, J. D. (1995). *The elements of language curriculum: A systematic approach to program development*. New York: Heinle & Heinle.

Buck, G. (1997). The testing of listening in a second language. In C. Clapham and D. Corson (eds.), *Encyclopedia of language and education, volume 7: Language testing and assessment* (pp. 65-74). Netherlands: Kluwer Academic Publishers.

Carroll, J. C. (1980). *Testing communicative performance: An interim study*. Oxford: Pergammon Press.

Chandavimol, M. (1988). *The placement test: A useful or harmful tool*. (ERIC Document Reproduction Service No. ED 298 769)

Courchêne, R. (1995). An alternative method for teaching and testing reading comprehension. *TESL Canada Journal, 12*(2), 50-58.

Cumming, A. (1997). The testing of writing in a second language. In C. Clapham and D. Corson (eds.), *Encyclopedia of language and education, volume 7: Language testing and assessment* (pp. 51-63). Netherlands: Kluwer Academic Publishers.

Davis, K. A. (1995). Qualitative theory and methods in applied linguistics research. *TESOL Quarterly, 29*(3), 427-453.

Elson, N. (1992). The failure of tests: Language tests and post-secondary admission of ESL Students. In B. Burnaby and A. Cumming (Eds.), *Socio-political aspects of ESL* (pp. 110-121). Toronto: OISE.

Fulcher, G. (1997). The testing of L2 speaking. In C. Clapham and D. Corson (eds.), *Encyclopedia of language and education, volume 7: Language testing and assessment* (pp. 75-85). Netherlands: Kluwer Academic Publishers.

Gaffney, J. & Mason, V. (1983). Rationalizing placement and promotion decisions in a major ELT program. *TESOL Quarterly, 17*(1), 97-108.

Harrison, A. (1983). *A language testing handbook*. London: Macmillan Press.

Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.

Ilyin, D. (1970). Developing a placement test for adults in English-second-language programs in California. (ERIC Document Reproduction Service No. ED 036 766)

Johnson, D. M. (1992). *Approaches to research in second language learning.* New York: Longman.

Kirschner, M., Spector-Cohen, E., & Wexler, C. (1996). A teacher education workshop on the construction of EFL tests and materials. *TESOL Quarterly, 30*(1), 85-111.

LeBlanc, R. & Painchaud, G. (1985). Self-assessment as a second language placement instrument. *TESOL Quarterly, 19*(4), 673-687.

Lynch, B. K. and Davidson, F. (1994). Criterion-referenced language test development: Linking curricula, teachers, and tests. *TESOL Quarterly, 28*(4), 727-743.

Malu, K. F. (1988). Entrance testing and course placement at the U.N. International School, New York City. *ELT Journal, 43*(3), 206-212.

O'Malley, J. M. & Pierce, L. V. (1996). *Authentic assessment for English language learners: Practical approaches for teachers.* Addison-Wesley.

Oscarson, M. (1997). Self-assessment of foreign and second language proficiency. In C. Clapham and D. Corson (eds.), *Encyclopedia of language and education, volume 7: Language testing and assessment* (pp. 175-187). Netherlands: Kluwer Academic Publishers.

Norton, B. (1997). Accountability in language assessment. In C. Clapham and D. Corson (eds.), *Encyclopedia of language and education, volume 7: Language testing and assessment* (pp. 313-322). Netherlands: Kluwer Academic Publishers

Peirce, B. N. & Stein, P. (1995). Why the "Monkeys Passage" bombed: Tests, genres, and teaching. *Harvard Educational Review, 65*(1), 50-65.

Progosh, D. (1996). Using video for listening assessment: Opinions of test-takers. *TESL Canada Journal, 14*(1), 34-44.

Rea-Dickens, P. (1997). The testing of grammar in a second language. In C. Clapham and D. Corson (eds.), *Encyclopedia of language and education, volume 7: Language testing and assessment* (pp. 87-97). Netherlands: Kluwer Academic Publishers.

Read, J. (1997). Assessing vocabulary in a second language. In C. Clapham and D. Corson (eds.), *Encyclopedia of language and education, volume 7: Language testing and assessment* (pp. 99-107). Netherlands: Kluwer Academic Publishers.

Rich, J. C. (1993). Can a writing sample improve placement in English courses? Research Report No. 93-13R. (ERIC Document Reproduction Service No. ED 366 400)

Richards, J. C. (1996). Teachers' maxims in language teaching. *TESOL Quarterly, 30*(2), 281-296.

Richards, J. C. & Lockhart, C. (1995). *Reflective teaching in second language classrooms.* New York: Cambridge University Press.

Schumacher, S. & McMillan, J. H. (1993). *Research in education: A conceptual introduction* (3rd ed.). New York: HarperCollins College Publishers.

Shohamy, E. (1993). The power of tests: The impact of language tests on teaching and learning. (ERIC Document Reproduction Service No. ED 362 040)

Tung, P. (1986). Computerized adaptive testing: Implications for language test developers. In C. W. Stansfield (ed.), *Technology and language testing* (pp. 13-28). Washington: TESOL Publications.

Underhill, N. (1991). *Testing spoken language: A handbook of oral testing techniques.* Cambridge: Cambridge University Press.

Vandrick, S. (1995). Privileged ESL university students. *TESOL Quarterly 29*(2), 375-381.

Wall, D. (1997). Impact and washback in language testing. In C. Clapham and D. Corson (eds.), *Encyclopedia of language and education, volume 7: Language testing and assessment* (pp. 99-107). Netherlands: Kluwer Academic Publishers.

Weir, C. J. (1990). *Communicative language testing*. New York: Prentice Hall.

Weir, C. J. (1997). The testing of reading in a second language. In C. Clapham and D. Corson (eds.), *Encyclopedia of language and education, volume 7: Language testing and assessment* (pp. 39-49). Netherlands: Kluwer Academic Publishers.

APPENDIX A

Questionnaire

Instructions: Put a check in the box, for example ☑, where appropriate.

1. I am ☐ an administrator.
   ☐ a teacher.

2. On registration day, also known as "interview day," I usually interview new students at the following level or levels (check all that apply).

   ☐ Level 1 ☐ Level 5
   ☐ Level 2 ☐ Level 6
   ☐ Level 3 ☐ Level 7
   ☐ Level 4 ☐ Level 8

3. In general, how useful is the Entrance Examination *alone* (that is, the 30-minute timed multiple-choice part of the entrance test containing 48 grammar, vocabulary, and reading comprehension questions) as a means of placing new students in the appropriate level?

   ☐ very useful
   ☐ somewhat useful
   ☐ neither useful nor useless
   ☐ somewhat useless
   ☐ useless

4. In general, how useful is the writing sample *alone* (that is, the 8-minute timed sample of writing found on the back of the multiple-choice answer sheet) as a means of placing new students in the appropriate level?

   ☐ very useful
   ☐ somewhat useful
   ☐ neither useful nor useless
   ☐ somewhat useless
   ☐ useless

5. In general, how useful is the oral interview *alone* as a means of placing new students in the appropriate level?

- ☐ very useful
- ☐ somewhat useful
- ☐ neither useful nor useless
- ☐ somewhat useless
- ☐ useless

6. (If you are an administrator, go to question 6b)

As a teacher, have you ever had students who you believed were misplaced into your class as a result of the entrance test?

☐ Yes (*go to question 6a*) ☐ No (*go to question 7*)

6a. (*TEACHERS ONLY*) *Approximately* how many students in an average month do you believe are misplaced into your class?

- ☐ 0 students
- ☐ 1 student
- ☐ 2 students
- ☐ 3 students
- ☐ 4 students
- ☐ 5 or more students

6b. (*ADMINISTRATORS ONLY*) *Approximately* how many students in an average month do you believe are misplaced *per class* at Pacific Language Institute?

- ☐ 0 students
- ☐ 1 student
- ☐ 2 students
- ☐ 3 students
- ☐ 4 students
- ☐ 5 or more students

7. In general, how effective is the ***entire*** placement process (that is, the 30-minute timed multiple-choice part of the entrance test containing 48 grammar, vocabulary, and reading comprehension questions; ***plus*** the 8-minute timed sample of writing found on the back of the multiple-choice answer sheet; ***plus*** the oral interview) as a means of placing new students in the appropriate level?

- ☐ very effective
- ☐ somewhat effective
- ☐ neither effective nor ineffective
- ☐ somewhat ineffective
- ☐ ineffective

Thank you for participating in this study!