# STUDENT AND TEACHER PERCEPTIONS OF THE AUTHENTICITY, VALIDITY AND EXPERIENTIAL GROUNDEDNESS OF ENGLISH AS A SECOND LANGUAGE (ESL) EVALUATION: A PROGRAM CASE STUDY

## Guohua Pan

### B.A. Southwest Petroleum Institute, 1982

A Thesis
Submitted to the Faculty of Education
of the University of Lethbridge
in Partial Fulfilment of the Requirements for the Degree

**MASTER OF EDUCATION**

LETHBRIDGE, ALBERTA
April, 1997

## Preface

This thesis is structured into three sections: I. auxiliary section (A); II. main text; and III. auxiliary section (B).

Auxiliary section (A) includes the thesis abstract which briefly outlines the focus of this study, the tools used for data collection and the conclusions drawn from the study. As well, a table of contents is included which gives the chronological order of the content of the paper.

The main text consists of six chapters. Chapter One is a general introduction, in which the researcher explains the background and rationale for this study; gives the definitions for some key terms used in the paper; focuses on some components that the researcher believes are essential in an efficient and effective examination or test; and points out the significance of the study.

Chapter Two is a general literature review on ESL evaluation, focusing on such topics as the purpose of ESL evaluation; proficiency testing; the Test of English as a Second Language (TOEFL); ESL test content validity; authenticity in ESL evaluation; experiential groundedness; teachers and test content; criterion-referenced tests (CRTs), norm-referenced tests (NRTs); an ESL evaluation model; formative evaluation and summative evaluation.

Chapter Three presents the methodology used for the study. Here the researcher puts forward the hypotheses and the methodology for the study; describes briefly the ESL program that he is researching; and introduces the tools he intends to use for data collection.

Chapter Four includes the interview data from the researcher's interviews with eight randomly selected students in the Advanced Level classes A and B, and two instructors who taught the two classes' reading course, including the researcher's interpretation of the data. The focus is on validity, authenticity, and experiential groundedness in ESL evaluation.

Chapter Five includes the survey data derived from 18 returned questionnaires that the researcher distributed to the students in the two Advanced Level classes. The focus is again on validity, authenticity, experiential groundedness and other factors that may affect an efficient and effective ESL evaluation. The researcher's interpretation of this data is also presented in this chapter.

Chapter Six presents the conclusion the researcher draws from the study he has conducted. It includes the merits, the strengths that he thinks the ESL program possesses, and some possible modifications which he thinks might be of some help to enhance the already successful program, and last but not least, the limitations of the study which has been conducted within two and half months.

Auxiliary section B is composed of two parts: references and appendixes.

# Abstract

This thesis explores and discusses the current practice of English as a Second Language (ESL) evaluation primarily in terms of validity, authenticity, and experiential groundedness.

The researcher endeavours to apply theories of ESL evaluation to actual practice, analyzing the differences or harmonization between them with regard to validity, authenticity, experiential groundedness and some other factors identified as being important in evaluation in language acquisition situations, particularly CRTs, NRTs, formative and summative evaluations.

Ethnographic inquiry is used for obtaining data. A program case study, a technique of qualitative research, is used for this study. The purpose is to gather some genuine data from the interviewees which are real and revealing, thus enabling the researcher to gain particular insights relative to the aims of his study.

The analysis aspires to give voice to interviewee's views on ESL evaluation in terms of validity, authenticity, and experiential groundedness, along with other factors noted above.

This report sets out to determine how or if validity, authenticity, and experiential groundedness, along with some other factors identified above, are integrated elements of an efficient and effective ESL evaluation. Such was found to be the case and it was also found that these can be reinforced by guaranteeing that an evaluation targets its population closely, adopting real-life approaches and, establishing linkages between evaluation and prior knowledge and skills.

v

# Acknowledgements

# Table of Contents

ix

# List of Tables

# Chapter One  Introduction

## 1.1 Background and Rationale for the Study

Canada has been a multicultural country since her birth, populated by people from all over the world. Owing to her unique characteristics and history, English (along with French) has been one of the official languages in the country. This study will focus on English in particular.

According to Statistics Canada (1995), there were 90 200 foreign students studying in Canada in the 1992-1993 academic year. In fact, each year, there are over 20 000 newly arrived foreign students who come to Canada for various levels of education. Because of their limited English language ability, many foreign students need first to advance their English at ESL training institutions in Canada so that they will have the necessary English language skill to further their academic pursuits. This is so since English is the only language for instruction and communication in most learning institutions outside Quebec. Various ESL test or evaluation results, in addition to other qualifications, decide whether the students are qualified for entering academic programs at Canadian institutions of higher learning. Many times students' ESL results are the decisive factor for their further academic endeavours since they have already been

evaluated in other fields before coming to Canada. Consequently, the ESL
evaluation determines, to a certain extent, a student's future.

Because of the importance of ESL evaluation, special attention has been given
to its design and delivery. There is much discussion and debate on the
improvement of this evaluation. In spite of this, examinees and examiners
complain that ESL evaluation often does not reflect examinees' real English
language ability. Moreover, because of the unreliable or invalid results they
obtain on from the tests or examinations, ESL students are either excluded the
opportunity to pursue their academic studies at an institution of higher learning
or have to stay in an ESL program longer, thereby delaying their further
academic pursuits.

How does one design and deliver ESL evaluation that reflects the true English
ability of the examinees? This is a question that has various answers and on
which this study will focus.

## 1.2 Definition of Terms

A number of specialized terms will be used throughout this study. Definitions for
some terms are provided in this section, which the researcher thinks are most

relevant in this particular research. These terms include: authenticity, criterion-referenced tests (CRTs), evaluation, formative evaluation, experiential groundedness, norm-referenced tests (NRTs), summative evaluation, test or examination and validity.

Authenticity -- The conformity between test content and the fact or reality, for example, what is presented in a test is what is actually used in real situations. As Carrol and Hall (1985) argue: "the tasks have been presented in an integrated way so that parts of the test pick up the multi-skill features of real life where reading leads to speaking, then to writing with possibly more speaking--the sort of sequence we would see in an executive meeting, for example" (p. 5).

Criterion-Referenced Tests (CRTs) -- CRTs relate students' scores directly to the performance of specific tasks, usually at a given number of levels of mastery (Carroll & Hall, 1985).

Norm-Referenced Tests (NRTs) -- NRTs aim at interpreting each student's performance by comparing it with that of other students on the same exam (Carrol & Hall, 1985). "An individual's performance is interpreted in terms of his/her relative position in a specified group (Ary, Jacobs, and Razavieh, 1985, p. 191).

Evaluation -- Burke (1994) defines evaluation as "the process of interpreting the evidence and making judgments and decisions based on it" (p. xvi). Aitken (1996) interprets evaluation as "the judgment of the outcome of the analysis and interpretation of the data gathered in the assessment" (Appendix C).

Experiential Groundedness -- Prior knowledge and actual experiences from learning, life, work, or other sources.

Formative Evaluation -- Formative evaluation is ongoing, aims to improve learning or performance rather than to grade. "Formative evaluation looks at a program during its early stages, while program elements can still be changed, if desired, in response to local conditions" (Jacobson, 1982, p. 288). Formative assessment also identifies students' strengths and weaknesses and provides feedback to help them improve.

Summative Evaluation -- Summative evaluation is usually conducted at the end of a study unit, activity, course, semester, or program so that the results from it, as well as from formative evaluation, can be used to determine students' achievement and the effectiveness of a course or program. (Board of Education for the City of Etobicoke, 1987)

Test or examination -- This study adopts Aitken's views that tests are "the measuring instrument or tool" (Appendix C). The Webster's Dictionary (1961) defines examination as an exercise or a series of exercises designed to check or find out writers' progress on problem solving skills, task performance and therefore to determine the quality of learning over a period of time. A test is "a set of stimuli presented to an individual in order to elicit responses on the basis of which a numerical score can be assigned." The score indicates the extent a student possesses the characteristics being measured (Ary, Jacobs & Razavieh, 1985).

Validity -- Validity is the degree of appropriateness and adequacy with regard to the original purpose of the test. "Validity," according to Burke (1993), "is the degree to which a measurement technique obtains the kind of evidence which its user intends to collect" (p. 17). It refers to the extent to which the test measures what it was intended to measure (Shohamy, 1985, p. 74).

## 1.3 Some Components of An Effective Test

As is the case with all tests, it is essential to conduct an ESL test effectively. But how can test makers guarantee that a test is effective, such as the test is to reflect objectively the students' real target language (i.e., English) abilities?

Test content validity is one of the important components that ensures that the test reflects the students' real target language (English) ability. A reading comprehension test, for example, should be adequate and appropriate in serving its goals and intentions. The students' score obtained from the test should be a meaningful indicator of the students' ability and be a measure of that ability only. In short, a valid, effective test should measure what it is supposed to measure, and nothing else.

Since the students will advance their study or are actually studying in a spoken English environment, the context of a test should also be authentic, e.g., the test context itself should be as close as possible to the English that is actually being used in real settings. Bachman (1991) notes that the characteristics of the test task need to be perceived as corresponding to the features of a target language use situation (situational authenticity), or an interaction that exists between the test takers and the test task, and an involvement of test takers' language ability in accomplishing a test task (interactional authenticity).

Practices have shown that a test still might not reflect test takers' real ability in the target language even though content validity and contextual authenticity are taken into account. The test takers' experience of, or familiarity with, or background knowledge relative to the test topic (i.e., experiential groundedness) also play an important role in reflecting their ability in the target

language.

In summary then, three key factors appear to contribute to ESL test effectiveness and efficiency: validity of the test content, authenticity of the test (both situational and interactional) and the experiential groundedness of the test (from the examinee's perspective). These could be supported by Principles of Fair Student Assessment Practice in Canada (1993).

Besides these three key factors, some other factors on which this study will focus also play a part in providing true, reliable data for efficient and effective evaluation. Those factors include formative evaluation, Criterion Referenced Tests(CRTs) and Norm Referenced Tests (NRTs).

Furthermore, reliability is one of the important issues in effective evaluation. Reliability means that the result or conclusion drawn from one test would be the same or similar on other tests of the same kind. The researcher of this study believes that once validity, authenticity, experiential groundedness and the other factors identified in this study are reached in an evaluation, the data should be reliable. Consequently, reliability ensues. Therefore, reliability is not discussed in this study and is not seen as being relevant to the researcher's objectives in this study.

## 1.4 Significance of the Study

In English-speaking Canada, non-English speaking students need to communicate with people in English in most situations, i.e., shopping or advancing their studies at learning institutions. The evaluation or assessment of their English ability becomes very important and, in many cases, decisive in determining their status.

An efficient and effective evaluation is needed and necessary for ESL students to smoothly transit themselves to the new and appropriate study situation. The Test Of English as a Foreign Language (TOEFL) has been widely used at universities across North America for non-English speaking applicants to undergraduate and graduate studies. In spite of its popularity, however, the reliability of its result varies with specific students and other factors. In fact, many applicants are better while conducting their undergraduate or graduate studies at a university in North America in terms of their English ability than they appeared in TOEFL tests. On the other hand, some students whose TOEFL scores are above the minimum admission requirements do not appear so competitive in English. The practice shows that there is discrepancy between TOEFL scores and the actual English abilities of the non-English speaking students.

It is possible that the students might be placed in a class that is inappropriate for them (either too difficult or too easy in terms of English level) because of their evaluation results, or they might be delayed, or even be excluded from further study in academic programs, if their results are a consequence of evaluation that is not well designed or delivered (rather than indicative of their poor level of competence and performance).

There were many factors that affect the accurate evaluation of the students' English ability through English tests. It is especially so when a test such as TOEFL is conducted in the test-takers' home countries where the test is called English as a Foreign Language (EFL). There are many differences between ESL and EFL evaluation, as the latter is conducted at a place where English is the first language.

This study proposes to present the importance of content validity, situational and interactional authenticity and experiential groundedness and other factors in terms of ESL evaluation as derived from the literature, and from the approaches used by the ESL program at the Language Centre affiliated with a western Canadian university. The purpose is to explore, on the basis of the Language Centre's specific conditions, the possibility of more authentic, valid, experientially-grounded test/evaluation contexts or approach(es) that will consequently reflect more accurately the ESL test takers' English ability. The

study aims to provide the parties concerned (i.e., students, decision makers, test makers, instructors, et cetera) with "real" and "effective" data that show the significance of valid content, authentic test contexts and experiential groundedness in facilitating whatever decisions or directions they may take. Thus, it also seeks to recommend the merits of the ESL instruction and learning and evaluation activities at the Language Centre, facilitating teachers' instruction and students' learning on its currently successful basis, thereby possibly ensuring that the Centre provides other departments with highly qualified students and enhances the overall reputation of the university.

# Chapter Two Literature Review

In lieu of an exhaustive literature review, the researcher has chosen to limit himself to a number of factors considered most pertinent for this particular study.

## 2.1 Purposes of ESL Evaluation

When talking about evaluation, tests, grades and measurement come to mind. Omaggio (1979) states: "Evaluation is often thought of in the very narrow sense of 'tests' and 'grades'" (p. 236). In reality, it is usually only lay people who think of evaluation exclusively in these terms. Many professionals, however, tend to have different views as to the definitions of evaluation based on their specific perspectives.

Language testing usually takes place at schools in which results are used to assess students' performance and improve ongoing instruction and learning activities. Both instructors and students are frequently and closely involved in assessment. Instructors are able to give feedback through tests and evaluation. They are able to identify the weaknesses in their curriculum delivery as well as areas for instructional improvement. The data from tests also give useful information to other stakeholders--parents, administrators, and education

11

departments or ministries. At schools, testing, assessment, and evaluation are used as part of the instructional process to assess and improve existing teaching and learning activities (Shohamy, 1992). This is the issue of accountability--the accountability of instruction and learning.

The tests also serve the purpose of diagnosis--identification of student or program strengths and weaknesses, to enhance learning. Courtland and Gambell (1994) note that assessment in schools provides teachers with daily feedback or response and plays a decisive role in deciding students' future. Concurrently they can provide diagnostic information about students to others: parents, school administrators, employers, or college and university registrars.

Outside school, the data from evaluation serve the purposes of motivation and certification regarding specific persons' futures: admission to institutions of higher learning, placement, promotion, and the granting of certificates (Shohamy, 1992). These encourage both teachers and students to work efficiently and effectively towards their goals--graduating qualified personnel, learning as much as possible within a given period of time, assessing or endorsing the program, consequently approving the teachers' curriculum delivery. The only difference of various tests lies in the size or scale of the testing world-wide or campus-wide, or even class-wide. According to Shohamy

(1992), language testing occurs in two key contexts:

> 1) the school context, in which tests and other assessment
> procedures are used as part of the instructional process
> to improve teaching and learning in the school; and

> 2) the external context, in which tests are used to make
> important decisions about the future of individuals, as in
> granting certificates, accepting candidates for programs,
> and placing students in appropriate programs. (p. 513)

Lynch (1990) believes that evaluation (of ESL) is "the systematic attempt to examine what happens in, and as a result of, language programs, [that] typically serves as the basis for judgments and decisions about these programs" (p. 23).

As the results from a test are so important for the students and teachers, it is very important to have the tests reflect the true capacity of the students' English language skill. Yet how can we as professionals ensure that a test really does what it is supposed to do; namely, give the parties concerned data that are reliable and very close to reality? Professionals offer various ideas in the literature. Test content validity and authenticity of the test context certainly serve very important roles in accurately reflecting a test taker's ESL skills (Gordon & Hanauer, 1995; Bachman, 1991). Therefore, they should be guaranteed in the design and delivery of ESL tests. It is easy for people to communicate on

familiar subjects. It is very difficult to read, or talk over subjects that we don't know much about even in our mother tongue. It is, logically, almost impossible for us to check ESL students' English ability with test content and context that they know little about in their first language. Prior knowledge or current experience (i.e., experiential groundedness) regarding test content and context is, therefore, another important factor that decides the effectiveness of an ESL evaluation. Furthermore, one snapshot test might provide misleading information of the students because of some unexpected event, emotional or physical. A cumulative record would therefore offer more reliable data because it was built on a series of tests.

## 2.2 Proficiency Testing

There are many types of ESL tests, such as placement tests, achievement tests, proficiency tests, and so forth. This study's focus is on proficiency testing because  the test takers are students who live in an English environment and most of them are going to pursue their further academic studies in degree programs after their successful completion of ESL study at the Language Centre.

The first and most important need of ESL learners in a foreign (English)

language environment is survival, that is, to make themselves understood in terms of daily needs, be they living, working, or participating in the academic environment. Proficiency testing has become more and more important in ESL instruction, learning and evaluation because it aims at evaluating how well or to what degree learners are meeting these needs (Clark, 1983).

ESL evaluation is conducted to serve different purposes for different people. Professionals (in ESL circles) have been focusing their attention on proficiency testing because such a test checks the test takers' performance or functional (English) ability under conditions as close as possible to actual daily life in an English speaking environment.

Direct proficiency testing is used more and more often in practice. Well known ESL tests, such as TOEFL or MELAB, are proficiency tests. Clark (1983) asserts that the (proficiency) test requires the test takers to perform functionally-oriented language tasks in situations that approximate as closely as possible the conditions under which these tasks are carried out in the real-life setting. As this researcher noted in the previous Section 1.4 and in Section 2.3 which follows, the overall reliability of the above tests, particularly TOEFL, has been questioned frequently in the literature. The content validity, contextual authenticity, and experiential groundedness of an ESL language test are decisive components in evaluating test takers' true capacity in English

language skills. An overall test and evaluation program might better get at these evaluation features than does any "one-shot" test, however popular it might be. The TOEFL test will be our next focus in this review.

## 2.3 TOEFL Test

The Test of English as a Foreign Language, or TOEFL, as it is popularly called, is a widely recognized English test whose scores are regarded as quite acceptable for non-English speaking applicants for undergraduate or graduate studies at universities across North America. Scanning the admission requirements of North American universities or colleges, we would read that the TOEFL score is one of the compulsory admission requirements that non-native-English-speaking applicants must meet.

In spite of its merits and popularity in reflecting test takers' English competency, however, TOEFL has its weaknesses. It is not sensitive to the "evolutionary changes in ESL instructional goals and practice" (Suomi, 1991, p.1). It is a snapshot test. The one time placement test involves too great a stake for test takers whose fate might be decided by the result from a single test. The accompanying anxiety is predictable and could consequently affect greatly the performance of the test takers.

Most TOEFL takers are applicants for university undergraduate or graduate studies. Yet, the test is not geared to these specific audiences. Educational Testing Service (1996), the American organization that oversees the design and delivery of TOEFL tests, points out: "TOEFL is not a test of academic aptitude or a teaching instrument, nor can it provide information about the various social and psychological factors that must be considered in making placement or other decisions" (p. 3). The reality has been that it is the sole data for decision making regarding university admissions in terms of the English proficiency of the applicants. Some supplementary measures have to be taken for more accurate reflection of the test takers' English abilities.

There is a distance between test givers and test takers, physically and emotionally. Physically, most test takers write the test in their home countries or areas where English is not their first language instead of America where the test has been designed. Emotionally, most writers are to study at universities or colleges. Yet, the test does not target specific audiences. It is a general English proficiency test. As Raimes (1986) notes, there are "the question of what the (TOEFL) test measures, the compatibility of topic types, the look of topic choice; the lack of distinction between [writing tasks for] graduate and undergraduate students; the scoring system" (p. 427).

The Educational Testing Service (1996) acknowledges that some necessary

changes are needed so that it reflects "the current understanding of communicative competence and performance-based language assessment and its results provide more information than current TOEFL scores about international students' ability to use English in academic settings" (p. 5).

This brings us to the matter of test validity.

## 2.4 ESL Test Content Validity

The validity of an ESL test plays a decisive role in deciding whether the test can reach its goal--to evaluate a test taker's English language ability. Validity is the "accuracy of educational assessment" (Popham, 1995, p. 39). Test content validity is the degree of appropriateness and adequacy with regard to the original purpose of the test. The Joint Advisory Committee (1993) defines validity as "the degree to which inferences drawn from assessment results are meaningful" (p. 5). The content validity of an ESL test is the degree to which a test is measuring what it claims to measure. Then, what is an ESL test to measure?

An ESL test aims to find out the real ability of the students' English language skill. To reach the goal, the measures used in the test should be effective, that

is, they are expected to collect the data that reflect the students' real English ability as precisely as possible in terms of the area being tested. Any measurement for the test that does not serve that purpose is otherwise misleading. For example, a multiple-choice test on the components of essay writing is not so authentic or "real" as actually writing an essay on the topics of the students' choice.

Students' results from an ESL reading comprehension, for example, are expected to reflect their ability in constructing their understanding from processing written text in English. Any and every effort in designing and implementing the test is toward the purpose of measuring as precisely as possible the students' ability in deriving meanings from the written materials they read, not from what they write, listen to, or speak in English, nor from what they read, write, listen to or speak in French or any other language. Thus, the scores from reading comprehension reflect the process and the product of meaning construction from print sources, which ensures consequently that inferences made about a student's reading ability and decisions made on the basis of these scores are valid (Gorden & Hanauer, 1995). The scores from the test, therefore, approximate the student's true level of English language skill. As Bachman (1991) contends, the scores from a valid test should be a "meaningful indicator of a particular individual's ability (in English) and measure that ability and very little else" (p. 688). This is, of course, a strong case for discrete-point

testing and such an argument would need some adaptation in the case of authentic assessment in ESL, in which the interplay of skills is perhaps more important.

The appropriateness and adequacy with regard to the original purpose of an ESL test or test items ensure the content validity of the test. The result from the test is, therefore, meaningful because it measures what it is expected to.

However, content validity alone does not guarantee the test reflects the real capacity of the students. The authenticity of the test context also counts.

## 2.5 Authenticity of ESL Test Context

Authenticity is the degree of worthiness, credibility, genuineness of a thing or view. Webster's Third New International Dictionary (1961) defines authentic as "worthy of acceptance or belief by reason of conformity to fact and reality; not contradicted by evidence" (p. 146).

In ESL language evaluation, the significance of authenticity in inferring a student's English ability can never be overstated. A good test should be able to determine the extent to which ESL students can function in an appropriate and

effective linguistic manner in the particular language-use conditions. The key here is that the language should be real, authentic, that is, should be what people in English-speaking countries actually use daily in life, work, or academic settings. (Of course, this itself varies greatly. Pennycook (1989) notes that many teachers tend to promote views and methods they prefer.) This way, and only in this way, the goal(s) of the testing can be met--to find out the students' ability to use English language appropriately and effectively. Clark (1983) argues that ESL testing will utilize "external-to-program" real-life-oriented measures of functional proficiency and focus more on the testing of the functional language used within the context of specific students' learning experiences (p. 436).

The authenticity of ESL tests is improved with the relevance of the test items to the features of the English language use situation. In designing and conducting an English language test, for example, the teachers should make the characteristics of the test method correspond to certain features of vocabulary (e.g., occasional, professional) and topics (e.g., academic, management). If the test takers are prospective university students who are interested in business management or economics, the use of terms and topics from those areas is likely to increase the authenticity of the test. However, the introduction and use of such terms and topics are not adhered to rigidly. Test designers and implementers should not simply take some passages and paragraphs directly

from books for native English speakers, without assuring their relevance.

Since the targeted users (students) are from non-English speaking countries and areas, editing and adaptation are necessary for the specific level of language efficiency of the non-native English speaking users (students), but the critical features of the test materials are retained as they are actually used in real-life situations. Thus, what the students are tested on are what they are actually exposed to in reality, be it in business management, economics, or computer science, et cetera, only more suitable to the specific background of the students and to their interests as well.

Relevance in ESL tests also means that what is tested should be closely connected with what has been taught. We can't expect the test takers of business management perspectives to write a test with questions on geology, or physics, or astronomy. The characteristics of test input should match the topics and kind of materials a teacher used in class. Bachman (1991) agrees that there should be a close connection between the test items and authentic, related vocabulary and topics. He points out that if the students are from the engineering field, then "inclusion of technical terms and topics from engineering would tend to increase the situational authenticity of the test" (p. 690).

Besides relevance of vocabulary and topics in the test items, there is also

the matter of the uniqueness of the individual student in terms of vocabulary and the topics to be introduced in test items. ESL students are from different countries and areas of the world. They vary in culture, values, education and socio-economic backgrounds. Accordingly, their involvement and response vary with their background, in spite of the fact that they are taught by the same instructor using the same materials in the same class. The test designers and implementers can design and implement a test as authentic and valid as practically possible. The authenticity, however, is only meaningful to a given test taker, or a given group of test takers, to a certain degree. Bachman (1991) points out that

> authenticity ··· is a function of the extent and type of involvement
> of test takers' language ability in accomplishing a test task. ··· We
> can do our best to design test tasks that will be authentic for a
> given group of test takers, but we need to realize that different
> test takers may process the same test in different ways, often in
> ways we may not anticipate. (p. 691)

The increasing level of the students' involvement in improving their English language ability will consequently raise the authenticity of an ESL test. The involvement means that students participate in selecting instructional materials, test materials, the format of the test, and so forth. This way, a combination of students' interests, their English language level and their familiarity with the

method of instruction ensues.

In children's education, for example, it is advocated that children's learning can be facilitated by increasing their interest (i.e., making them absorbed in what is being learned or done). There are many cases in which things are learned more quickly and better if the learners are interested in them. Many children, for example, are good at playing video games. An important factor is that they like it -- either watching others play or playing themselves, and as much as possible since it is a very interesting thing for them to do. They fully plunge themselves in learning and practising the game and they gradually get better and better at the skills involved in the game.

Similarly, making the test content interesting will certainly increase the students' interest and consequently their involvement in the language ability being evaluated. The word interest in Latin is inter-esse, which means "showing the connections between things." Instructors, therefore, should make students connect things by integrating learning and test activities closely and engaging students in participating in those activities. Bachman (1991) thinks that "involvement of language ability may be increased by making the test task interesting to test takers" (p. 695). Moreover, the closer involvement of the students in the language ability being assessed consequently increases the interactional authenticity of the test. The following practice may achieve or

increase the interactional authenticity and/or the relevance of the test task.

In a writing test, ESL students are asked to write a composition with a minimum number of words, for example, 200. They may choose the topic for the composition -- whatever they feel is most comfortable, familiar or interesting.

Then the time is allocated for the test. [After the assigned time, more detailed information is to be provided on grading criteria.] Using the scoring criteria, the students are encouraged to revise their writing. There may also be peer sharing and comment on each other's writing or peer-assessment. Further revision is permitted and encouraged based on the peer's sharing and comments. This will create a favourable condition and more possibility or opportunity for the students to display their real ability in English language skill, and consequently, the result from the test will more likely be closer to their true level of English ability. Thus, the purpose of the test -- validity, authenticity, both situational and interactional, is reached. Furthermore, the students improve themselves through a variety of activities during the test, as the proverb says "practice makes perfect" (Bachman, 1991).

The trend in second language assessment is that more authentic assessment is being advocated and steadily put into practice from the theoretical to the practical level of second language classrooms. For example, in the French

language, the Canadian Association of Second Language Teachers has launched a major assessment project aimed at developing contextually authentic French as a Second Language (FSL) tests. In Alberta, such development is exemplified in the 13 model tests developed by the FSL Evaluation Project (Heffernan, Caouette, Bourassa, and Colbourne, 1996).

Personal interest in picking topics for a test helps to improve the authenticity of an ESL evaluation. Students' interest in topics helps them to focus on the content of what they choose, therefore, their real ability in English language is more likely to be reflected through the test.

Bachman (1991) has suggested:

> the theory of interactional authenticity and situational authenticity in a language test ground practical considerations in test design, development, and use firmly on a theoretical framework of the nature of language ability and test tasks, and thus provides a principled basis for making practical test development decisions. ... enable us to specify and assess the relationship between language test performance and nontest language use, and thus provide a principled basis for addressing issues of validity and authenticity.
> (p. 698)

In addition to content validity and contextual authenticity, experiential

groundedness is also a significant factor in language acquisition evaluation.

## 2.6 Background or Current Knowledge: Experiential Groundedness

In practice, we often find that when students are asked something familiar, or something they know well, or something learned before, the students will be able to answer the questions more smoothly and clearly, even in the language that is not their mother tongue. They are able to use all the vocabulary available to answer, describe, narrate, or write familiar things and questions because of their background or current knowledge. Nevertheless, they can't carry out the task smoothly and successfully if the questions or subject matter are completely new to them. They may even be unable to answer, describe or narrate the questions or subject matter in their mother tongue because of the lack of background or current, experiential knowledge. Small wonder it is out of the question for them to answer, i.e., describe the unfamiliar/unknown in a foreign language -- English!

The goal of a language test is to assess the test takers' ability to understand the target language. The test takers' background and current knowledge and experience indeed play important roles in deciding the validity of an evaluation. Perkins and Brutten (1988) agree that readers' prior knowledge, in particular,

has much to do with their understanding of the test context.

Writing, describing, and listening in English, a second language, are not passive activities, but are critical components of the process of either description, or writing, or listening, like those in a first language. The process of writing a test is an active one -- the students construct meanings from their cultural and experimental frames -- their knowledge of language, text structure, concepts, and special fields. Students use their existing knowledge to connect the information from the texts of test and construct or infer meanings. Whether the test content is adequate and appropriate for the students in terms of their culture, profession and vocabulary background, depends to a certain extent on whether validity of the test context can be achieved. That explains the reality that in most tests, the texts of tests are chosen by selecting, adapting and revising some original readings with the purpose of making the texts relevant to the test takers' background; thus the test is believed to be valid.

Hill and Parry (1992) state that:

> It [the text of tests] has been expressively chosen, adapted, or constructed because it provides test makers with material around which they can build what they conceive to be valid tasks. Central to their notion of validity is that tasks be built around odd bits

of fact that readers do not ordinarily carry around
in their heads. .... In working through the tasks, test
takers must first determine what information is called
for, locate it in the passage, and then decide on
appropriate response by comparing the exact
wording of the passage and task. (p. 437)

Schmidt-Rinehart (1994) also notes the relation between the familiarity with

testing topics and the evaluation score/results in his observation from one of his

studies:

The Chinese EFL students' comprehension was
measured by a multiple-choice test that contained
both passage-dependent and passage-independent
items. Regarding topic familiarity, the subjects
scored higher on the familiar - topic than the
unfamiliar-topic lecture. A significant effect was
found, however, only on the passage-independent
items this interaction between prior knowledge
and test type may show that passage-independent
items provide a measure of background knowledge,
but effect of background knowledge itself on
comprehension of information from the passage
remains unclear. (p. 180)

Schmidt-Rinehart (1994) believes that the results of the between-within analysis

of variance clearly indicate that topic familiarity affected the scores of the recall

measure. He argues that it is important for teachers to recognize that students'

existing knowledge contributes significantly to their comprehension and that listening is not a passive activity. He states that taking time to assess the conceptual base the listeners bring to the text will enable teachers to go beyond dealing with the linguistic information in order to help students understand and make their learning more meaningful. The result of his study indicates that helping students make connections to their previous knowledge in order to build a mental framework with which to link the new information might facilitate comprehension.

As context is so critical in ESL evaluation, it appears teachers' connection to the context of evaluation is important.

## 2.7 Teachers and Test Context

Except for a few ESL tests such as TOEFL or MELAB, most tests are designed and actually written by instructors who at the same time teach. The advantage of instructor, test giver, designer or grader being one and the same is that instructors usually know the students well. They have the most and closest contact with them through their daily teaching and learning activities. Instructors know exactly their target with respect to the students' reality -- their motivation, needs, and background. The instructors' adequate and appropriate design and

delivery of a test in accordance with the actual linguistic level of the test takers usually facilitate the students' display of their true competence in English language skills.

Lynch and Davidson (1994) advocate the teachers' involvement in test design and delivery and hold that there is a close link between the test context, the students' reality, and the curriculum goals. This linkage thus enables the test context to be appropriate, effective, and therefore valid from their test design and delivery activities. They argue that the Criterion-Referenced Language Test Development (CRLTD) process can help instructors better articulate their understanding of their curriculum objectives and help them to link those objectives to the testing mechanisms used to evaluate student achievement.

In spite of their solid knowledge about the students, teachers still need to advance their knowledge and skills in designing and conducting tests. ESL speaking students have different goals in writing an ESL test. The differences are not to be ignored, but rather, they are to be recognized and considered accordingly in the test preparation. This also requires that teachers have a knowledge and understanding of their differences. It requires that teachers set out to learn about the students, appreciate their culture, values, and interests and reflect this in the test context. Thus, the content validity and authenticity of he test context and experiential groundedness of the test can be attained and

accordingly, the scores from the test are more reliable and authentic, and reflect the true capacity of the students' English language skill.

All of the above suggests no need to "throw out the baby with the bath water". Indeed, in ESL evaluation, the literature suggests a balanced approach, combining the use of CRTs and NRTs.

## 2.8 Combined Use of Criterion-Referenced Tests (CRTs) and Norm-referenced Tests (NRTs)

To an extent, CRTs are regarded as effective tools for ESL evaluation. CRTs, according to Brown and Yule (1983), are usually created to measure "well-defined" and "fairly specific" instructional objectives. They argue that the objectives are often unique to a particular program and serve as the basis for the curriculum. Hence, it is important for the instructors and students to know exactly what those objectives are so that appropriate time and attention can be focused on teaching and learning them. The purpose of CRTs, then, is to measure the degree to which students have developed knowledge and skill on a specific objective or set of objectives.

It is recommended that CRTs be used together with NRTs so that the evaluation

for a student is more comprehensive, inclusive, and objective. NRTs aim at measuring the general language skills or abilities (such as reading comprehension, conversation, listening, et cetera). A student's English language ability is judged not only by his scores from the specific test but also by reference to the scores of other students in the same test. Brown and Yule (1983) point out that each student's score on NRTs is interpreted relatively to the scores of all other students in the same test so that each of them is spread out along a continuum of scores and their placement along the continuum immediately tells their abilities relative to the norms in ESL. The NRT takers usually have an idea about the general form of the questions that will appear in the test papers but do not know the specific content in the papers.

In way of illustrative example of a balanced approach in ESL evaluation, this review cites Lynch's (1990) model.

## 2.9 A Model for ESL Evaluation

Lynch (1990) builds a context-adaptive model for ESL program evaluation. The model consists of a series of general steps:

1. Establish the audience(s) and goals for the evaluation.

2. Develop a context inventory and determine
   which dimensions are important in light of
   the goals and audience for the evaluation.

3. Develop a preliminary thematic framework
   based on the issues that are central to the
   particular context.

4. Develop a data collection design/system
   based on the audience and goals and on
   the context inventory, and that is focused
   by the thematic framework.

5. Collect the data and revise Steps 3 and
   4 as necessary; possibly elaborate Step 2.

6. Analyze the data and revise Steps 3 and 4
   as necessary.

7. Formulate the evaluation report. (p. 24)

Lynch (1990) claims that the model is meant to be interative, with the results of
certain steps necessitating a return to earlier ones for changes in
conceptualization. He recommends that the strongest approach to evaluation is
one that combines as many methods, qualitative and quantitative, as are
appropriate to the particular evaluation context. He stresses:

"The context-adaptive model provides a framework that encourages the

multiple-strategy approach. This iterative framework leads program evaluators through a set of considerations that can adapt the evaluation to a variety of specific program settings" (p. 39).

In this study, Lynch's model of evaluation has been assessed to assist the researcher in the design of his interview instrument.

Two other significant features of evaluation which remain to be reviewed are formative and summative evaluation.

## 2.10 Formative Evaluation

According to Jacobson (1982), there are mainly two kinds of evaluation that have different foci (purposes), formative and summative evaluation. In this section, the object of our discussion is formative evaluation.

"Formative evaluation looks at a program during its early stages, while program elements can still be changed, if desired, in response to local conditions" (p. 288).

Formative evaluation is ongoing. It provides valuable diagnostic feedback for

teachers as well as students, tells them the weaknesses that need to be improved on as well as the progress that has been made in their English study. The evaluation can and should be conducted at the initial stage of an ESL program development, and carry on until the end of the program; there are always things to be changed and improved. The feedback helps teachers with their curriculum design and delivery, and their test design and delivery. Thus, the instruction, as well as test, is able to focus more on the key issues, more appropriate for the students; thereby the data from tests are more reliable and closer to reality.

Jacobson (1982) asserts that formative evaluation data are sorely needed by, and lacking in, most foreign language programs. He points out that such data can be used to:

> examine the effectiveness of current instructional materials towards meeting the program's instructional goals;
>
> examine the match between program goals and learner activities;
>
> look at the match between testing content/ strategies and learner instructional activities;
>
> develop criteria for program self-assessment; investigate the effectiveness of various teaching methodologies in achieving program goals; examine the relationship between teaching

> method and program conditions (e.g., which
> teaching methodologies are best suited to the
> various types of programs?);
>
> identify the most effective components to the
> "foreign language teacher style";
>
> look at "innovative" programs in their early
> years of implementation in order to identify the
> most promising elements and to increase their
> effectiveness. (p. 288)

Ongoing formative evaluation procedures and feedback, according to Jacobson

(1982), "will provide defensible evidence leading to increased credibility for the

profession's efforts to improve instructional programs" (p. 289). In spite of the

fact that these observations were made a decade and a half ago, they still have

considerable currency.

Summative evaluation must be looked at as well with respect to formative
evaluation in an ESL evaluation.

## 2.11  Summative  Evaluation

Summative evaluation judges an operational program's worthiness or a

student's achievement. The conclusion drawn from the evaluation decides the

fate of the program or the student - continuation or termination, graduation or

continuation in a program. The result from the evaluation decides whether an

ESL student can graduate from a program and pursue his/her further study

either in a higher level of an ESL program or programs other than ESL, or stay

in the same program for one more semester. Summative evaluation plays a

rather important role in this affair (Jarvis & Adams, 1979, p. 6). Summative

evaluation focuses on determining the merits of a program at its completion.

Jacobson (1982 ) further summarizes it as the following:

> Summative evaluation is terminal evaluation of a program
> that is already operational. Its purpose is to make
> judgment about a program's worth. Ultimately, summative
> evaluation is tied to decisions about support and
> continuation of a program. ….Summative evaluation data
> are the most authoritative and defensible information the
> profession can provide to any interested parties. Such
> data can be used to:
>
> provide replicable outcome data for all types of foreign
> language programs;
>
> provide comparative program outcome data for
> competing program types;
> determine effective instructional materials and teaching
> methodologies;
>
> identify exemplary programs worthy of dissemination
> and/or replication;
> provide program cost effectiveness data. (p. 289)

It will be of interest in this study to determine how formative and summative

evaluation approaches are balanced in an ESL evaluation.

## 2.12 Summary

This review of the pertinent literature suggests that ESL evaluation is used to either get feedback for program/instructor/student improvement, or gather data for decision making on the continuation or termination of a program, students' graduation from, or continuation in the program. Also it is used internationally to judge non-English students' ability in English as the dominant international language.

TOEFL is a popular EFL test whose results are widely recognized throughout North American universities and colleges. It is delivered worldwide several times a year. It was organized over 30 years ago. The places where it is held, the audience who writes the test, all these and other, related factors affect its effectiveness in reflecting the test takers' English abilities.

The content validity, authenticity of the test context and experiential groundedness of the test are essential to reflect the real capacity of the students' English language skill. The validity of a test refers to the appropriateness and adequacy of its content. The result from a test with good content validity serves as a meaningful indicator that, used together with other components, students' true ability in English language is being demonstrated. The test is, therefore, up to what it is expected to do -- measures what it is

supposed to measure.

Authenticity of the test context is another decisive component for the evaluation of students' real strengths or weaknesses. Authenticity means the "real thing." It means that there is little gap, if any, between test and the "actual stuff." It should be seamless. Authenticity also means relevance, e.g., relevance of the test items to the actual features of the English language use situation, that is, what is being taught and tested is what is actually being done in real life. This helps to attain the requisite ESL test context authenticity.

The involvement of the students in the test task-the situation in which the students care for and are interested in what they do also improves the ESL test context authenticity.

Test authenticity is further decided by the students' familiarity with, and background or current knowledge of the test content (i.e., experiential groundedness).

The feedback from an ongoing evaluation helps teachers and students identify weaknesses as well as strengths in their instruction and learning, and ultimately improve the authenticity of the instruction, learning and test context. Summative evaluation, usually used more authoritatively for the assessment of an

operational program's relative worthiness, or a student's achievement, is recommended to be used together with formative evaluation for a more balanced, weighted conclusion. Similarly, people are recommended to use NRTs and CRTs jointly for more effective ESL evaluation that not only offers particular students' scores in a test but also their placement along a continuum, e.g., their position in a group of students as well as their knowledge and skill on a specific objective or set of objectives.

Guided by this review of the literature on what constitutes effective ESL evaluation and his own curiosity about how this might apply in a particular situation, the researcher framed particular research questions (hypotheses) and developed a methodological approach to allow him to find answers to those questions (test those hypothese).

## Chapter Three  Methodology for This Study

### 3.1 Hypotheses

As has been indicated earlier in the literature review, professionals have been exploring, discussing and experimenting with new theories on ESL proficiency testing. The foci on which this researcher has chosen to orient his study are primarily: content validity, authenticity of the test context, and the experiential groundedness of the test vis-a-vis the life experience of the particular test takers. Other evaluation factors (formative and summative evaluation and CRTs and NRTs) were also identified and will be the secondary object of this study.

The researcher proposes that a study be conducted testing the following hypotheses:

A. Integration (in terms of validity, contextual authenticity and
   experiential groundedness) of the content of tests with
   students' future academic pursuits, will be discernible in
   the Language Centre's practice and in interviewees'
   testimonial about that practice ; and

42

B. A combination of formative evaluation with summative evaluation,
and of CRTs and NRTs, which helps teachers and students to
gain a more objective picture of the students' overall proficiency,
will be discernible in the Language Centre's practice and in the
interviewees' testimonial about that practice.

## 3.2 Methodology

This research has been conducted in the form of observations (informal) and
interviews (the results of which constitute the main focus of this study). The
researcher attended the ESL classes offered at the Language Centre affiliated
with a western Canadian university on a regular basis (for an eight-week
period) in fall, 1996. A close and relaxed, though objective, relationship was
established between the researcher and the ESL instructors (who were also
test givers) and students. This not only helped him with his observations but
also facilitated his subsequent interviews. In particular, this reduced the
interviewees' feeling of being intruded upon, and at the same time, enabled the
interviewer to get an inside view of what was being studied, through the
participants' perspectives, while also helping him maintain his objectivity as a
researcher (Bogden & Biklen, 1992).

For the informal observation component of the study, the researcher attended two advanced (ESL) classes once at each class per week from the class of the second week in September until mid-November in the fall of 1996. He kept notes during his observations, communicated with students, learned the students' backgrounds and their ideas about ESL learning and evaluation, and their intent regarding further academic pursuits after completing their course of study. The observation notes, which the researcher has retained and which are incorporated in the body of this report, have been compared with what they said in formal interviews, as well as in casual conversations.

The researcher also exchanged views between classes with the two instructors who taught the classes. He got to know the instructors' ideas regarding instruction and evaluation through these informal conversations. This has been written down and used to compare with the instructors' views articulated in the formal interviews (See Appendix A) which were actually conducted three times in four successive weeks and completed by the end of October, 1996.

Interviews were conducted with eight ESL students and two instructors (at the Language Centre). There were 31  students altogether in two advanced classes in the semester of September-November, 1996. The researcher interviewed four students in each class. This sample represents 25% of all ESL students in the two classes.

The students were randomly chosen. However, the researcher intentionally chose four male and four female students as interviewees. Interviewees of different nationalities were also deliberately chosen. This way, the researcher believed the results of the interviews would be more inclusive, based on wider perspectives of both male and female students, and more representative of these multilingually and multiculturally diverse classes.

The interviews usually started with the explanation of their purpose and an expression of thanks for the students' and instructors' participation and cooperation.

Some interviews were saved in the form of cassette recordings, and transcribed in summary form on paper, with others done by telephone, and note taken. For reasons of confidentiality, the tape will be erased shortly after final completion of this study. However, for reasons of any possible future verification or follow-up regarding this study, the researcher will retain his notes. The interviews were structured and semi-structured and adapted after consulting with the instructors and on the basis of pilot interviews of two students.

The interviews were conducted in the way of conversations which focused on issues in ESL instruction and learning and evaluation [including formative, summative, Norm-referrenced tests (NRTs) and Criterion-referenced tests

(CRTs)], on validity, authenticity and experiential groundedness issues in particular, and on an open-ended question (any aspects of ESL learning and instruction and evaluation that the subjects wished to discuss).

Under the guidance of the supervisor and the researcher's thesis committee members, the interview protocol was drawn up and administered to the ESL students at the Language Centre from mid-October to mid-November, 1996 (on the basis of the protocol approved by committee members in early September and in light of the pilot interviews carried out in early October).

The interview protocol included 30 items (see Appendix B), plus a space for students' concerns, observations, recommendations regarding ESL teaching and evaluation. The protocol was divided into six groups of questions for the interviewees to answer. Each group, consisting of five items, raised questions in the same category, such as the test takers' satisfaction with the test topic, connection with instruction, connection with the subjects to be pursued, understanding of questions, the evaluation tools, format, suggestions for improvement, and so forth. They were categorized under the broad rubrics of validity, authenticity, experiential groundedness, formative and summative evaluation, NRTs and CRTs, and general concerns and recommendations.

The items were generated from the literature review and also out of the

researcher's own experience as an EFL instructor, a test maker, an EFL (TOEFL) test taker in the People's Republic of China, and one conducting EFL tests and piloting the EFL (TOEFL) test (for Chinese TOEFL test takers).

Each question had options for the subjects' structured responses ranging from strongly agree, agree, undecided, disagree, and strongly disagree; or yes, does not apply, no; and open-ended questions for the subjects' (semi-structured) responses. As the wider range of answers, from strongly agree to strongly disagree, offered more options for the interviewees to respond to the questions, there were five groups of such question-response types. There was one group of open-ended questions that allowed interviewees to air their opinions freely (based on their English ability, one group of open-ended questions was sufficient for the purpose, the researcher believed), while one group of question-response types was for those points that were relatively easier to agree or disagree with.

The interviews and returned interview protocol (Survey, Appendix C) were carefully sorted and analyzed according to the question groups. Since each group focused on a specific field of the ESL test, the division facilitated the researcher's identification of strengths and weaknesses in the test or in the program. The analysis of the returned interview protocol was combined with the researcher's notes from informal observations and conversations with the

students and teachers. The conclusions drawn from this broader base of analysis, therefore, has been more comprehensive and potentially less biased.

A consent form was signed by both the interviewer and the interviewees regarding the study to be conducted. The consent form assured that the students' confidentiality regarding their academic standing and status at the university and other institutions in Canada, as a result of their participation in the interviews and observation, would be respected. It was also made clear to all that participation in the study was on a volunteer basis; no one was pressed to participate.

A proposal for informal classroom observations, together with the purpose, frequency and duration of this activity, was first submitted late in August, 1996 for approval by the director of the Language Centre as well as the head instructor who took care of the daily academic affairs at the Centre. The observation of, and interview with, the instructors were again matters of volunteer participation. A consent form was signed representing an agreement between the interviewer (the researcher) and interviewees (the instructors) with regard to the instructors' classroom teaching and evaluation. The data gathered from the observation of and interview with the instructors were to be kept confidential. The consent was to assure the instructors that their teaching and evaluation would by no means be affected by their cooperation with the study.

That was, the data regarding the research location, the participants would be kept confidential and deleted upon the complement of the research. The research would provide general feedback which might be useful locally and which might also have some broader, generic implications for similar ESL teaching/learning/evaluation contexts.

## 3.3 The ESL program

The ESL Language Centre was a teaching unit affiliated with a university in western Canada. The Centre ran an ESL training program from Intermediate to Advanced levels to TOEFL preparation classes. Among them, Advanced classes aimed at graduating qualified students whose first language was not English for academic undergraduate studies at the university. The students' qualification was decided by their final scores from tests based on their language training (in addition to other qualifications, which they had met before coming to Canada or possessed when entering the program). One of the benefits of learning English at the Centre, claims the Centre (1996) in its brochure, is that "a student who successfully completes the Advanced level with an average of 70% or above meets the University's English language requirement and will not be required to submit a TOEFL score for university admission" (p. 2). This policy is in line with the university's general English

writing competence requirement for native English speakers, which also allows them a number of different ways to demonstrate that competence (University Calender).

## Outline for the Advanced Class (Reading)

The researcher was able to obtain a copy of the reading course outline for the Advanced level class. The objective for the course was to "help students, through a "process approach" to read, understand and critically respond to a variety of articles" (ADVANCED ESL READING COURSE OUTLINE, FALL'96).

The outline set more specific objectives for the students to achieve. It indicated (Fall, 1996) that the students should be able to:

1. do prereading using different techniques
2. guess at word meaning using context
3. find, understand and remember main ideas and important supporting points
4. handle difficult readings and exam questions using "intensive reading" techniques
5. adjust reading speeds appropriately
6. critically respond to readings

7. apply suitable reading strategies depending on the text and on the
   reader's needs

Generally, language tests serve two purposes:

a) to assess and improve the existing programs, learning and teaching
activities; and b) to decide the learners' future such as admissions to institutions
of learning, the granting of certificates, promotion, placement, etc. (Shohamy,
1992, p. 513). The ESL program (Advanced level classes) belonged to the
second category since the qualified students were to be admitted to the
undergraduate programs at the university.

There were 31 students altogether learning English in the two Advanced level
classes which the researcher observed at the ESL training centre. The program
lasted about three months (from the first week of September to the last week of
November). The students were from all over the world -- Asia, Europe, Central
and South America. They were in Advanced level classes because of their high
TOEFL scores or scores from the entrance test (held at the beginning of the
training program). They were divided into Class A and Class B randomly, as
the number of students was big enough for two classes (Interview notes,
October, 1996). The division of the classes was based on students' "racial,
cultural, gender, country differences" (Interview notes, October, 1996).

There were reading, writing, communication, and grammar courses for the two classes. They were taught by different instructors who each taught one course. The students worked hard because they were going to study in academic programs at the university. They had to write four tests altogether, plus two projects and some take-home assignments. Their final scores represented the accumulation of their scores from the tests, projects, and their classroom participation. Their final scores determined whether they would qualify for studying degree programs at the university, stay on for one more semester's training or graduate from the program for good.

According to the regulation approved by the university, those ESL students in the Advanced level classes whose final scores were 70 percent or above were qualified for undergraduate studies at the university without providing TOEFL scores. The students who got lower than 70 percent could either keep learning until their final scores were 70 percent or above if they wanted to study degree programs at the university, or graduate from the ESL program and do whatever they wished outside the university context. There were both pressures and incentives for students to study hard since the majority of them expressed their intent to go into degree programs at the university after they finished their programs (the Centre brochure, 1996).

## 3.4 Classroom observation

The researcher attended the two Advanced level classes every week. In order to avoid putting pressure on the students, there was no formal introduction for the researcher's presence at the classes. It was as if he were one of the students who attended classes. The only difference was that he went to each class once every week and did the exercises and reading along with the rest of the class.

At the same time, he kept notes on how the instructors organized classroom learning activities as well as on the strategies and tactics they used to deliver the curriculum. The students' reactions to teachers' strategies and tactics, the materials used for classroom learning, the subjects the materials were in and the source of the materials were noted.

The observations began in the second week of September and ended on November 13th, when the survey questionnaires were distributed.

## 3.5 Interviews

The researcher interviewed the two instructors and eight students. To ensure

the fairness and inclusiveness of the interview, the (student) interviewees were randomly selected. However, the researcher deliberately set out to choose four male and four female students (two each in class A and class B). As well, efforts were taken to make sure that they were all from different countries and areas of the world. The interviewees were from the former Yugoslavia (female), Hong Kong (female), Japan (male), Korea (female), Mexico (female), Pakistan (male), Taiwan (male) and Venezuela (male). The interviews were conducted either face-to-face, or by telephone, depending on the preference and convenience of the interviewees. Each interview lasted on average about 40 minutes, with the exception of the interviews with the two instructors. There were three interviews with each of them, respectively, and each of them spent altogether over two hours on answering interview questions. It took about one month to complete the interviewing, starting in early October, and ending in early November. Then, a questionnaire was distributed to the students (on the 13th of November), asking them questions under the categories of validity, authenticity, and experiential groundedness and other factors in their ESL tests. At the same time, the researcher kept attending the classes regularly for his observations until mid-November, which was one and a half weeks away from the end of the program, when he distributed the survey.

## 3.6 Informal conversations

Having attended the classes for observation for a couple of weeks, the researcher tried to talk with the students in class. The contact was very informal. The topics for conversation would be about their names, country of origin, future plans, the length of time involved in English study, and so forth. The conversation was conducted not only in the classroom, but also at the food court on campus, in the library, on the bus, wherever they met and could talk.

No matter where the conversation took place, however, sensitive topics would be avoided. The researcher's intent was to get to know the students and get along with them to pave the way for later interviews and survey.

The relationship between the researcher and the students improved greatly along with their familiarity regarding each other. There was a sense of companionship between them because the researcher himself was indeed a graduate student. This helped greatly with the researcher in his later interviews and survey, enabling him to collect some in-depth data that were revealing.

The relationship was such that some students would tell the researcher about such things as their impression of a certain course, their scores from the final test, their departure date for home, their plans for the following semester, or their mailing address.

## 3.7 Questionnaire survey

The Centre's training semester began in the first week of September and ended in the last week of November. The researcher distributed 31 copies of a questionnaire to the students on November 13, 1996, following the students' completion of their last project in class and one and half weeks away from their final exam. They were relatively available at that time and could, therefore, afford time to read and fill in the questionnaire survey.

By November 20, 1996, 18 of the questionnaires had been returned. The return rate was 58%. The survey consisted of 30 items ranging from multiple-choice to open-ended questions. Considering the subjects' English level and time pressures, the researcher tried to make the questions easy to read and answer. The questions were developed under five categories: validity, authenticity, experiential groundedness, formative/summative evaluation, CRTs and NRTs, and general. The intention was to include as wide a range of questions as the limited scope of language ability could permit so that answers might result in some revealing data.

## 3.8 Summary

Preparatory work had been done before the researcher set out to collect data on site. The work included consent forms for the parties involved--the director of the Language Centre, the two instructors, the students in two Advanced level classes. Besides that, a letter to the Human Subjects Research Committee had been drawn and submitted to the committee for the approval of the planned data collection. There was also the design of interview protocols for both the instructors and the students.

Actual data collecting, lasting over two months, consisted of classroom observation, interviews, casual or informal conversations, and questionnaire survey. The classroom observation enabled a casual, friendly relationship to develop between the researcher and the interviewees, which in turn paved the way for revealing interviews, survey, or casual conversations.

All the questions and conversations focused on validity, authenticity, experiential groundedness, formative and summative evaluation and NRTs and CRTs in the context of ESL evaluation.

The two instructors who taught Advanced A and B, respectively, were interviewed and 25.8% of 31 students (8 of them altogether) were interviewed.

Thirty-one copies of the survey were distributed and 58% (18) of them were returned.

Interpretive results of this study will be reported on in the two sections which follow, focusing particularly on the interviews and questionnaire surveys, while also providing reference to informal observations and casual conversation notes, both also forming part of and contributing to the outcome of this study.

# Chapter Four  Interview data and interpretation

The interviews lasted over one month. The interview protocols had been designed for the instructors and students, respectively. The questions used were exactly the same, in spite of the interviewees' different background and gender. The protocol (Appendix B) contained questions under the categories of validity, authenticity, experiential groundedness, formative and summative evaluations, CRTs and NRTs and general questions. The intent was to compare the data from interviews with those gleaned from the later survey in order to make the inferences derived from the study more accurate and convincing.

## 4.1 Validity

The tests (in the Advanced level classes) in reading, writing, and communication were designed and conducted with a primary concern for accountability. In the reading test, for example, the subject matter being tested focused mostly on those subjects that had been covered in classroom learning and teaching. Students were asked to analyze articles, describe their main ideas, point out the topic sentence in each paragraph and find synonyms or antonyms for key (new) words and expressions. (Interview notes, October, 1996)

There were also two projects, the evaluation of which formed part of the students' final score. A project was actually an independent reading task in which, according to the instructors, the students were required to find an article of their choice from English newspapers, magazines, or other sources and read the articles, then write one to two pages about them in terms of main ideas, paragraph topic sentences, et cetera. It was, claimed the instructors (Interview notes, October, 1996), the verification of students' application of what had been learned (in class) to the real situation. It was appropriate since students were to choose their own articles, based on their English level and interests. The instructors believed that it was, in a sense, better than multiple-choice or true/false questions in that it was real application of learned knowledge and skills to actual problem solving. Moreover, the project followed closely the materials and skills learned in class. The instructors expressed the opinion that students' processing of the articles would display the degree of their comprehension, consequently their English ability in reading. As it was not a formal test; it put less pressure on them, yet the results were no less a meaningful indicator of their English ability. The students took it seriously since it represented 15% of their final scores. Besides, no one wanted to appear inferior to his/her peers. In addition, the materials for the project were interesting and adequate because they were chosen by the students themselves and therefore were suitable for them in terms of difficulty level. The researcher noticed once, in his classroom observation, that a few students were actively

asking the instructors questions about or seeking advice on their project after the project had been assigned, a sure sign of the project's wherewithal to motivate them.

The instructors recognized and attempted to make a close connection between learning and application (test). They (October, 1996) said:

> We usually try to test skills that are being taught. We want the students to demonstrate their understanding of what is being taught. This will closely engage them in learning. We tie things together. We use in exams or tests over 60 percent of materials that have been discussed or learned in class. Only one -third in the tests is new material that students did not learn or cover in class. Even the one-third is from the materials that students take home for after-school work. You have to have relevance.
>
> If there is relevance, the data are valid. (Interview notes)

For the communication class, students were asked to do classroom presentations. The goal was to prepare students for upcoming academic degree studies. Students were to prepare their presentations and deliver them in class by themselves. They were evaluated through observation: how they presented the topics, their choice of words, expressions and their oral delivery.

All presented in front of the class with their strengths or weaknesses pointed out by both peer students and instructors. The presentations were not only for instructors to evaluate but also for peers to learn from. As two Japanese students (who were not interviewees but volunteered to talk to the researcher) put it, presentations were very good; they learned most from them--by presenting and by watching others present. They improved not only from doing it themselves but from learning how others did it, too.

For the writing component, students were asked to write essays on subjects of their choice. Again, the goal was to prepare students for their further academic studies in which writing of essays would play an important part. Since writing requires a student to utilize all the knowledge and skills in vocabulary, structure, meaning construction and writing skills, it was a challenge for them. Nevertheless, most students felt it both a challenge and an opportunity for them to write essays. They thought it a good opportunity to display their comprehensive English ability and to find their weaknesses for improvement through essay writing. It encouraged and directed students to develop their problem-solving ability as it required that they did all the writing themselves. With improved writing skills, they would be better prepared for their later academic (degree) studies.

Because of the close connection between their ESL training and future

academic pursuits, students enjoyed the test format and content delivered at the
Centre. The materials used for tests were mostly covered in classroom learning
or outside-class assignments so that what appeared in tests served mostly as a
check on what had been learned before. The connection was obvious. It
seemed that they enjoyed very much the instruction and evaluation practice.
And the scores from tests served the purpose for which they were intended, that
is, as a meaningful indication of the students' English ability. As Bachman
(1990) states, "validity    is the extent to which the inference or decisions we
make on the basis of test scores are meaningful, appropriate, and useful" (p.
25).

## 4.2 Authenticity

Most of the students were going to study academic degree programs in Canada
(particularly at the university). They had to familiarize themselves with the
language (terms) used in university studies, as well as the social, economical,
cultural setting they were in. They had to learn to read, write, listen to and speak
the language that people used in actual settings. The classroom instruction
aimed at real-life language use certainly facilitated students' awareness and
acquisition of "authentic language." The close connection between learning and
assessment engage students in learning, consequently in learning and using

real, authentic English.

Language learning and instruction and language testing are not contradictory, competing against each other. Rather, they are different parts of an integrated whole that serve the roles of engaging students in learning and learning well. What is learned determines what is to be tested, for example, the content of an test. On the other hand, tests not only provide information on how well students are learning (strengths and weaknesses of students' learning) but also how well the curriculum is delivered (the strengths or weaknesses of teachers' organization of learning activities). They tell students the area on which to focus their learning, they also tell teachers the areas on which to focus their instruction, thus enabling more efficient and effective learning and instruction. Shohamy (1984) notes that language learning theories can bring the input component into language testing, defining appropriate testing content and language behaviour while measurement/testing theories can bring out the output component of tests, such as, how language performance and proficiency can turn into tests of these constructs.

Both teachers and students pointed out that articles in their original were used for reading class because students had learned English before (based on the interview with the students, their average English learning time had been about seven years before coming into this program)(Interview notes, October, 1996).

Most of them had been either college or university graduates or students in their home countries/areas. Of course, some modifications were made to the original articles so that students would not feel frustrated when reading them. The teachers highlighted some key words and expressions that they thought were new and difficult for students to understand. Vocabulary sheets were prepared ahead of class with definitions and explanations of the key words and expressions. Apart form those, the rest were left untouched: they were original; they talked about North American life, work, economics, politics, culture, science and education, et cetera. The students were exposed to the authentic language environment. It was believed that they would gradually pick up and use what they had been learning in real settings.

There was over 60% of testing content that was taken directly from teaching materials (e.g., the materials discussed, questioned and answered in class). The teaching materials (and consequently the testing materials) were selected according to the students' interests and preferences, based on the survey conducted at the beginning of the training. Conflicts of interest and preferences might arise because of the diversity of the students' background, in country of origin, and cultural, economical, religious, social and educational differences. In order to minimize the possible conflicts, instructors had the final say in deciding the use of teaching materials, and the testing materials accordingly; in terms of their readability, depth, et cetera. (Interview notes, October, 1996). They were

also based on the teachers' perception of what was expected for first year university students in terms of English language, since they themselves had gone through university education and had taught at different institutions of learning.

The students felt quite positive about the practice of learning English and being tested at the Language Centre. They (all the eight interviewees) mentioned that a great advantage of learning English at the Centre was that what they learned and consequently on what they had been tested was real English used in actual settings (Interview notes, October, 1996). Therefore, the knowledge and skills checked in tests were useful and applicable in real life. The tests really checked their ability in using learned English language knowledge and skills. They all said that the tests were more relevant (to their learning) because of close links between classroom learning and tests. They also felt less pressure when writing a test.

As Jones (1985) contends, an applied performance test focuses on the examination of the students' ability in applying learned knowledge or skills to actual or simulated settings. Either the test stimulus or the desired response or both are intended to lend a high degree of realism to the test situation.

On the other hand, however, five out of the eight interviewees said that they

were uncertain for particular subjects when they were asked whether the scores from their tests reflected their English ability. They were afraid that knowing what was to be tested in advance would possibly allow students to focus on the targeted materials and therefore, get through the test with relative ease. Nevertheless, it was not necessary that they could get the same high scores when writing a test of similar difficulty level, yet the content was not disclosed ahead of the test. This was so because there was no discussion, neither were there answers in a non-informed test which required students to spend more time on reviewing what had been learned but not everyone would spend that much time on reviewing for one reason or another. One supporting evidence two students provided was that some students had studied at the training program for at least two successive semesters; however, their scores were not any better than those of some new comers, for example, themselves (Interview notes, October, 1996).

It seemed that a transfer was necessary. The purpose of testing is to check the students' acquisition of covered knowledge and skills. The aim of acquiring the knowledge and skills is to be able to use them in reality. Indeed, what had been learned should be consistent with what was to be tested. Yet, it was not judged as appropriate to mechanically use the classroom learning materials in tests. Some changes might be necessary or something new might also be introduced into the test so the "exam/test stimulus or desired response" (Jones, 1985)

might be reached. After all, things change. We cannot find in a textbook or teaching materials exactly the same as what we encounter in life. Dewey (1916/1966) believes that learning is best by doing, something few people doubt. We may also learn through testing by applying learned knowledge or skills to problems that are a little different from what we read in textbooks or other learning materials.

As Swain (1985) suggests, well-balanced test content consists of a substantial ratio of new to known information so it is motivating, substantive, integrated, and interactive. It was probably based on this viewpoint that the students were required to do projects in which they might encounter many things that were new to them, including vocabulary, grammar, culture, and other things. In addition to the (two) projects, about "one-third in the tests is new material that students did not learn or cover in class" (October, 1996, Interview notes).

The projects and the tests with about one-third new applications certainly would encourage or force students to learn and to apply what they had learned to real situations. The projects and tests were therefore motivating to most if they wanted to perform well in them. As well, the projects, tests were integrated and substantive in that they were well-balanced, and contained the necessary knowledge and skills that check not only the their problem-solving ability. The process of problem-solving was an interactive one in that it required the

students to figure out what the problems were and how they could be solved by applying the knowledge and skills they possessed.

## 4.3 **Experiential groundedness**

Students' prior and new, real-life experiences and knowledge come to mind when experiential groundedness is mentioned. Indeed, it is essential for teachers to think of students when they design and conduct a test because the result from a test is valid and authentic only if the knowledge and skills in it are what the students learned or experienced before or are now experiencing in a real-life context. Therefore, a test not considering students' prior and current knowledge or experiences is not a good one. In fact, the results from such a test would be invalid therefore misleading, since it is irrelevant to the reality of the students and something other than language ability might account for the better or poorer test results.

However, not only students need prior and current experiences and knowledge and skills to display their ability in tests; teachers, too, need many experiences and a range of knowledge and skills in designing a test.

Sometimes, teachers' experiences, knowledge and skills may be more

important than anything else in carrying out efficient and effective evaluations.

The teachers at the Language Centre had good, broad-based ESL teaching experiences. One teacher at the Advanced class had taught ESL in different places (colleges in BC, NWT, for about nine years). She had the experience of teaching ESL for college and university applicants as well as for new immigrants. Another used to teach at high school, then went to business, then back to ESL teaching. She had been teaching, at varying intervals, for ten years. Another had just come back to Canada for a half year from EFL teaching of two years at a college and a university, respectively, in China. (interview notes, October, 1996)

The (head) teacher interviewed (October, 1996) pointed out that her previous experiences in teaching new immigrants English, her English language teaching experiences at schools and colleges in the provinces of Alberta, British Columbia and NWT, and her perception of what a first year student was required for studies at Canadian universities in terms of English language all helped her understand the ESL instruction and evaluation (Interview notes). Consequently, she was able to design and conduct curriculum and evaluation with the  consideration of all possible factors involved, allowing more efficient and effective activities in learning and evaluation. An example was that while designing and delivering curriculum, the teachers insisted on a balance

between students' interests, their preferences and the teachers' perception of requirements for undergraduates in terms of English. Furthermore, their experiences in teaching, particularly in ESL, EFL teaching, could certainly be of great help for the students to adjust and transfer more smoothly to the new environment and were essential to the design and delivery of the ESL curriculum. As Clarke (1994) stresses: "The experience of teachers is central to the process of developing and applying theory" (p. 14).

For the students, their answers were mostly positive when the questions in the interview related to experiential groundedness. For example, all eight interviewees (Interview notes, October, 1996) said that the tests were more relevant and better in terms of links between the test and their learning and living experiences than the tests they had written (in their home countries) before because most knowledge and skills in the tests had been either experienced or learned by them before taking the tests. They believed that the results from such tests reflected more closely their real English ability than those that did not take into consideration students' experiential backgrounds (Interview notes, October, 1996).

Experiential groundedness is particularly important for ESL students because their English knowledge and skills are usually acquired through learning at schools, colleges or universities, with limited time and space for practice, in

spite of the fact that they may live in an English speaking country or area. This certainly affects their performance in academic studies and tests in which they are able to derive meanings within their limited specialty. Alderson (1985) notes that ESL students, when using English texts, are "much narrower than native speakers, and the skills they have acquired may be rather limited, enabling them to extract a certain information from a specialized text" (p. 27).

In real life, it is common wisdom that people have to spend a considerable amount of time if they want to know something that is not in their area of study. For example, we may not expect native English speaking liberal arts students to know much about math, or physics, or chemistry. Many a time it is likely these persons cannot even recognize some terms (in English) in math, or physics, or chemistry unless they consult a dictionary. It is small wonder then that many ESL students would fail if they were examined in English for something that they had not learned or experienced before. In a sense, they are like young school children who are learning to speak, read, write, and listen. Their test content should also focus on what they learn, not on something foreign, which would be totally beyond them experientially.

## 4.4 Summary

Over one month's interviews of eight students and two teachers focused on

validity, authenticity and experiential groundedness, as well as on the other identified factors in this study: NRTs and CRTs and formative and summative evaluation.

Both teachers and students in the interviews agreed that a close connection between classroom teaching and learning and tests and projects enhanced the validity of their ESL evaluations as experienced in tests, projects and homework assignments.

Authenticity is reached because of the introduction of original English materials about North American life, work, education and other areas in teaching and learning and tests based on the students' interests, preferences (from the survey conducted at the beginning of the program). What was tested was real English used in actual settings. Consequently, the tests were relevant, reflected the students' real English ability. Yet, how to balance teaching and learning and tests, how to determine the ratio of new things in a test, remained a question.

ESL students learning English are like young children learning to speak, read, write and listen in a school. An effective test is to check what they (ESL students as well as children) have learned (in class or wherever) before, not something that they have not experienced in some way. The students believed that experiential groundedness was achieved because of the consideration of their

experiential backgrounds before determining the content of tests.

Having interviewed the students and instructors, as reported on in this chapter, the researcher approached his research questions using an alternative methodology, a survey questionnaire, addressed to his student sample only. It was anticipated that this methodological triangulation would serve as a check on his interview findings, further validating them, as the case might be.

# Chapter Five   Survey data interpretation

As has been mentioned earlier in this study, 31 surveys (questionnaires) were distributed to the students on November 13, 1996. The survey were given just after their last (second) project, and one and a half weeks before their final exam. The students were relatively available at that time and could, therefore, afford time to read and fill in the questionnaire survey.

The researcher organized an informal social gathering for the students to show his appreciation for their participation in and cooperation with the research. The get-together was held after the survey was distributed to the students. It also was organized with a view to ensuring that there would be a higher return ratio of the survey. Both the researcher and the teachers invited the students formally to the informal gathering.

Most students were expected to be present at this event where enough food and beverages were prepared for them. However, only five of them showed up, besides the teachers. The researcher asked some of them later why they did not turn up at the party. All those he asked said they felt shy to be at the party because the researcher was neither a student nor a teacher at the Centre.

Though he talked to most of them and developed a rather casual relationship with them, they still didn't feel comfortable to come to his social event. Additionally, an accidental death had unfortunately occurred to one of the students and this, it is believed, also impeded much student attendance at this event.

In spite of the episode, however, the return on the survey was considered adequate for the purposes of this study. By November 20, 1996, 18 of them had been returned. The return rate was then 58%.

The survey consisted of 30 items ranging from multiple choices to open-ended questions. Considering the subjects' English level and time pressures, the researcher tried to make the questions easy to read and answer. The questions were developed under five categories: validity, authenticity, experiential groundedness, formative and summative evaluation and CRTs and NRTs, with an additional, general category for open-ended responses (please see Appendix C). The intention was to include as wide a range of questions as the limited scope could permit so that answers might result in some revealing data.

## 5.1 Validity

Generally speaking, there was a close correlation between the students' efforts and their scores from tests. However, students did not tend to agree with each other on many specific questions. This was partly due to their different backgrounds, their understanding of the evaluation from their specific perspectives. It is common for people to have different rather than uniform views on many things. It is especially so when young people of roughly the same age, yet diversified backgrounds, are asked about their views on specific things.

As has been mentioned in the interview section of this report, the students vary greatly in terms of their backgrounds--country of origin, religion, culture, education, et cetera. In spite of the fact that they either took an entrance exam or had written a TOEFL to be admitted to the Advanced class, their English level varied with their previous education and their backgrounds. Moreover, when it came to the matter of their consensus on the tests, there were diversified views regarding ESL evaluation.

Table 1 includes the data collected from the returned questionnaire survey regarding the validity of their ESL evaluation. As might be noticed, The students had similar or close views on certain items while different ones on others.

Table 1. Validity of ESL Evaluation (18 respondents)

| Q\A | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 | 0 | 8 | 2 | 6 | 2 |
| 2 | 2 | 6 | 5 | 4 | 1 |
| 3 | 0 | 2 | 4 | 10 | 2 |
| 4 | 4 | 11 | 1 | 2 | 0 |
| 5 | 0 | 11 | 3 | 3 | 0 |

Note 1: Horizontally, A (strongly agree) <------------->E (strongly disagree);
2: Vertically: A=answer; Q=question.

The students in the Advanced classes could not agree with the adequacy of difficulty in tests. Some (eight of them) believed that the difficulty level of the tests was in conformity with classroom instruction while others (eight of them) thought the tests were more difficult than classroom materials. Considering the diversity of the students, the controversy was natural and predictable. The remaining 2 had no comments (Survey item 1, Nov., 1996). This showed that different education backgrounds might lead to the students' different perception of the same thing. For those who learned English longer, and probably more competitive, they would think it not very difficult to perform well in the tests. It was a different story for those who had not learned English very long. They might have more difficulty in completing tests.

Most students (eight of them) thought that there was still considerable distance

between what was being learned and taught in class and checked in tests, and what was being used in reality. One of the reasons they cited was they did not think that they had made much progress, especially in speaking, and maybe in listening as well. A few of them said that they still could not use English to communicate well with Canadians (Interview NOTES, OCTOBER, 1996). However, some students (five) disagreed with the eight and thought that what they were learning was what was actually used in real life situations, though they sometimes, too, had difficulty in making themselves understood. (Survey item 2, Nov., 1996)

Nevertheless, about 67% (12) students did not think there was a distance between classroom instruction and learning and test content (Survey item 3). Only five thought the distance existed. Furthermore, the majority (83%) of the students, 15 of them agreed that they got good scores when they worked hard on test material (Survey item 4). Although it was contrary to the item 2, it seemed that the tests met their intended purpose--to measure what they were supposed to measure and the scores from them were meaningful indicators of the individuals' ability in English and measured that ability (Bachman, 1991).

Despite the close connection between learning and instruction and tests, many students (11 of them) still thought it necessary for the program to narrow its focus (in tests) (Survey item 5). This, as some students wrote on the returned

survey, depended on courses, or on the individual student. Some students who were relatively lower in level might like to have the focus narrowed further so that they could concentrate more on fewer materials and get marks good enough for them to graduate and enter into degree programs, while those who were relatively higher would not mind if the learning and instruction activities and tests remained unchanged. Placement procedures might also warrant further review.

Validity means a harmony in classroom instruction and learning and testing in terms of difficulty level. Tests with materials that are beyond what students have learned in class are not good because they don't reflect students' real language ability, though they may have used "real," "authentic" language. A conformity between classroom learning and instruction, and tests , plus use of real, authentic English in both, would add to the validity of a test.

As has been described earlier in this study, the students were from all over the world-Asia, Europe, North America and South America. Not only did they vary in country/race backgrounds but also they differed from each other in education experiences, with some being university graduates and others undergraduates, college graduates or high school graduates. What added more to the students' diversity was their English learning history--some of them had learned English for a couple of years; others learned it at college or university; still others

learned quite a few years of English; and a few of them went to high schools in Canada and got their Canadian high school diploma where they completed their studies in English. For university graduates, or undergraduates, there was the difference in the subjects they learned or were learning at the university or college, with some who took science or engineering subjects and others humanities or liberal arts. All these factors, plus their age, culture, and other differences, could explain their differing views on the tests' relative difficulty.

It might also be suggested that, given the students' varying competency levels as a result of their differing backgrounds, not all of them might have been appropriately placed at the Advanced level classes. There might be an indicator here for the Language Centre to review relative to its placement procedures.

It is further noted that the training program lasted less than three months from the first week of September to the last week of November. It was only about two and a half months into the program when the researcher distributed his survey. One could not expect students to have achieved a lot within such a short period of time. It takes longer, many times much longer, for people to improve their English substantially. In fact, there were at least three students in the ESL (Advanced) classes who had spent two years or more in Canada for their high school diploma studies before entering the program. Yet, they were there taking more ESL courses for university admission. The researcher asked two of them

why they should take ESL. One from Japan said his English grammar and writing were not competent enough for university study. A student from Hong Kong said neither his English reading nor his grammar was good enough for university admissions (Interview notes, Oct. 1996).

Summary

Both the teachers and students indicated that the teaching and evaluation should be closely integrated. The teachers (interviewed) and the students (from both their interviews and their returned surveys) agreed that there had indeed been a close connection between the instruction and test.

Almost the same number of students thought the tests either a little bit too difficult or relatively easy. Considering the variety of the students in the classes in terms of their background and their future pursuits and many people's tendency of wanting to have things easier than they are, differing perceptions regarding the difficulty level of the tests were natural and predictable.

The harmony between learning and tests, as had been implemented by the instructors and recognized by the students, meant that the tests were relevant and that the data from them were valid because they reflected what the students

had acquired and mastered in English. The tests, therefore, met their goal--to measure what they were supposed to measure. This finding corroborates the earlier findings derived from the interview data.

The survey questionnaire similarly focused next on authenticity.

## 5.2 Authenticity

Authenticity means that what is being learned and tested should be what is actually used in reality. Real-life is one of the approaches in authenticity theory. Real-life is the opposite of isolation from life, e.g., the kind of learning confined to very formal, academic, many times bookish materials that are not commonly used in daily situations. It means doing things the way it would be done in real life. A real-life approach requires teachers to use English of real, authentic form in teaching or examinations. For example, one would not give listening comprehension test on greeting cards, as they are written and read in real life. One would hesitate, if not refuse, to use English materials that are from non-English speaking sources.

However, real-life approaches should not be implemented mechanically. A real-life approach does not mean we have to have students learn and be tested on

site only. It is indeed effective if students could all learn and be tested on site, for example, in working, daily-life settings. By doing so, students are learning what people are using. Since there is a close connection between learning and evaluation, what they are tested on is what they have been learning, e.g., what people are using. It is, therefore, very real and authentic. But the real and the authentic are relative. Things are real and authentic only for those who are going to work or live in the same or similar settings after their ESL training. It is not so real or authentic, however, for those who are going on with other academic studies at institutions of learning after ESL training, because the setting will not be that of living or working, but rather academic. Besides, on-site learning and tests are usually time- and resources-consuming -- you need to have more time and more personnel to supervise them.

Authentic or not, a learning or evaluation activity, therefore, does not just depend on whether a learning or evaluation activity is conducted on site, in a real-life setting, not just on whether the materials for learning or evaluation are from original English (speaking) resources. What counts equally, or more, are the adequacy and appropriateness of them for the audiences, regarding their background, English level, and future pursuits.

Table 2 was the data concerned with authenticity. It was collected from questionnaire survey.

Table 2: Authenticity of ESL Evaluation  (18 respondents)

| Q\A | A | B | C | D | E |
|---|---|---|---|---|---|
| 6 | 0 | 2 | 4 | 8 | 4 |
| 7 | 2 | 4 | 6 | 6 | 0 |
| 8 | 0 | 8 | 3 | 7 | 0 |
| 9 | 1 | 0 | 4 | 11 | 2 |
| 10 | 0 | 4 | 5 | 9 | 0 |

Note 1: Horizontally, A (strongly agree) <--------->E (strongly disagree);

2: Vertically, A=answer;  Q=question.

From item 6 we could see that most (12) students were against the idea that the tests had a too narrow focus on daily-life English language because there had been little focus on that (Nov., 1996). Only two of them thought so. It might be because of their advanced English level that they thought the occasional appearance of daily English language in tests was not necessary.

At the same time, there was the same number of students (6) who either thought there was a difference between classroom learning and actual tests in terms of content difficulty (Survey item 7, November, 1996). It was natural that the students felt a gap between learning and application. It was especially so in communication, such as listening and speaking since they could not

communicate with native speakers freely in English. They had been working hard with the hope to make great progress as a result of their having taken the program. The majority of them came to Canada from their home countries directly, many of them from places as far away as Asia. It was not so easy even for Canadians to go this far just for language training. They certainly had great expectations and dreams for being able to learn English in an English speaking country. Many of them expressed that their major goal was to improve their listening and speaking abilities. They might hope to be able to speak as fluently as native speakers since they had the advantage of learning it in an authentic environment.

However, the reality was that they didn't feel so easy to talk with Canadians in English, though they had been reading, writing, listening to and speaking in English in (and perhaps outside) class most of, if not all, the time.

There was indeed some distance between classroom learning and the real-life situation. In spite of the fact that original materials from North American sources were used in the classroom, they were not as active and alive as they (the words, expressions, dialogues) were in real situations, although the former was usually based on the latter. Here we saw the need for adaptation and flexibility.

Teachers need to modify teaching materials to adapt to the students' actual

ability. Learners, too, have to adjust and adapt to the real-life situation for which it is very difficult to find exact examples, solutions from books, or classroom learning materials.

In fact, the different views among the students on the gap between learning and tests showed from another perspective that the students' diversified backgrounds decided greatly their perception of the program. More than two-thirds of the test content was previously discussed and covered. The remaining one third was also from materials such as take-home assignments, exercises, et cetera. Yet, for those who were less competitive, the one-third still appeared too much for them.

Bachman (1990) notes that there is always some distance between what is used in actual life settings and what is tested in the classroom. Accordingly, we should

> 1) accept 'real life' as a criterion or authenticity and modify our testing
>
> methods so that they do not impinge on the language use observed, or
>
> 2) recognize that a language test is different from real-life language,
>
> and attempt to define what constitutes 'authentic' test language. (p. 314)

Authenticity does not limit itself to the introduction of real-life language into tests.

The authenticity in a test also lies in the students' familiarity with the topics. But with the students so diversified in their background and their future pursuits, this familiarity could only be introduced in a general way. That is, tests could use the materials that are interesting and readable to most people of similar English level. An article of too narrow a focus might limit its readers to a certain group, which was neither the intention of the teachers nor of the students.

Interestingly and strangely enough, many students (eight out of the 18 returned surveys) (Survey item 8, November, 1996) claimed that there were topics in tests that they knew little about, and 13 of them thought interests had nothing to do with their test scores (Survey item 9), although most students believed there was a close connection between classroom teaching and tests (as teachers from the interview and students from both the interview and survey acknowledged that most test materials had been discussed in class, or previewed ahead of each test).

Nevertheless, it was not strange at all once we knew that the topics in classroom learning (and consequently in tests) covered a wide range of topics, from entertainment, to new drugs and new medical experiments, to new technology, et cetera. Though discussed in class, the wide range of fields (of the materials) could in most cases give students some impression because of the limited time possibly spent on them and the special knowledge

necessary for further and deeper understanding. There was not, nor could there be any in-depth understanding because of the wide range of topics and limited length of time devoted to each topic (of course, it might not be necessary for students to know the content exactly as long as they could understand what was being talked about in principle). An important factor was that the topics might have little connection with what they were going to study in academic programs. Despite the fact that they did learn some knowledge and skills from doing such extensive exercises, they thought they learned little or they had little connection with their tests.

The students thought interest of little importance because they were going to study in different subjects. Their interests were therefore diversified. What they cared about most was that the materials used for both learning and tests be selected from original English sources and based on their preferences (from the survey conducted at the beginning of the training). Most importantly, their understanding of the English materials was solid, the skills they used in exercises, projects, tests were appropriate. This not only helped them pass the tests so they could go on with their degree studies at the university, but also prepare them for the academic studies at the university.

It was based on this consideration that the teachers conducted a survey at the beginning of the training program, asking students to list their interests and their

preferences in selecting materials for classroom learning. Indeed, most teaching materials were selected this way (of course, the teachers had the final say in deciding whether to use a certain material or not based on their perception of the students' English level and its adequacy in terms of difficulty, sensitivity, and other factors).

The ESL students (nine of them) at the Advanced level classes disagreed with the statement in the survey "More exams should be conducted on site, e.g., at library, computer lab, registrar's, bookstore, shopping mall or post office, et cetera" (Survey item 10, Nov., 1996). Only four out of 18 supported the idea.

Living in an English-speaking country, the priority for a non-English speaking person is to survive--to know how to do shopping, mailing, greeting people, et cetera. They were indeed necessary for the beginners, or those whose English could not allow them to survive in daily needs. It was indeed a real-life approach to either teach or evaluate students in real life shopping, and so on.

However, it was inadequate and inappropriate for students at the Advanced level classes. Consequently, 50 percent of them were against it because they, having learned English for an average time of seven years when they were admitted to the training program and having lived in English speaking environments for some time when the survey was conducted, had already

proven their ability with deeds, or performance-based assessment activities. For them, these were a too simple, unnecessary repetition. It was simply redundant and a waste of time for them, too.

And the teachers knew that, so they planned and delivered their curriculum based on students' needs and the length of the program. Since the majority of the students were to further their studies in academic (degree) programs at the university, classroom teaching and learning, as well as tests, were relatively more adequate (and probably more efficient and effective) for the purpose because they were similar to what the students were to be exposed to in academic study programs.

The students also agreed with this. Based on the same item (10), only four out of the 18 thought that more tests should be conducted on site--library, computer lab, registrar's office, bookstore, shopping mall, post office, et cetera, while the remaining nine of them said no. Most of the students who had been living and learning in Canada for about two and half months at least, had all experienced those things and could carry out the majority of them independently. It simply made no sense for them to do the tests on their proven abilities.

## Summary

Authenticity consists of real-life approaches, topic familiarity and personal interests.

Efficient and effective evaluation requires that the tests introduced into class are materials that people are actually using in real life and work situations, yet not necessarily on site. Adequacy and appropriateness also count, and that, it seems on the basis of this study, many times more.

A close connection between classroom learning and tests promotes the conformity between the two activities. The students, because of their diversity in background and the lack of direct link between on-going learning and test and future pursuits, disagreed with each other with respect to the conformity of learning and test. However, they actually agreed with each other on conformity between their learning and test and further academic studies when they thought that there was a close connection between learning and the test because most of the skills and knowledge were of great use in their later studies. Again, particular interests were not a concern to them because of their diversified future plans, and because of the fact that most of the materials had been or were to be covered in class before tests.

Essentially, the results on authenticity in ESL evaluation derived from the survey questionnaire corroborates once again the findings from the interviews.

The survey dealt next with the issue of experiential groundedness.

## 5.3 Experiential groundedness

The survey data indicate that students' familiarity with the format and methods used in tests would help them perform better in such tests. Their prior knowledge about the topics used in tests could also lead to their improved performance. The data in Table 3 provide us with some interesting ideas about students' perception of their ESL evaluation from the perspective of experiential groundedness.

Table 3 : Experiential Groundedness in ESL Evaluation (18 respondents)

| Q A | A | B | C | D | E |
|------|---|---|---|---|---|
| 1 1 | 5 | 8 | 2 | 2 | 1 |
| 1 2 | 1 | 9 | 2 | 5 | 1 |
| 1 3 | 1 | 3 | 6 | 8 | 0 |
| 1 4 | 4 | 12 | 1 | 1 | 0 |
| 1 5 | 2 | 12 | 2 | 2 | 0 |

Note 1: Horizontally, A (strongly agree) <------->E (strongly disagree);

    2: Vertically, A=answer  Q=question

The majority of the students (13) indicated that there were differences in terms of test format between the Language Centre in Canada and that of their home countries (Survey item 11, November, 1996). Multiple-choice test questions were the main forms used in the test papers at the Language Centre.

Tracing back to the record (Interview and conversation notes, October, 1996), the researcher found many students said in interviews and outside class conversations that multiple-choice test questions had also been used in the test papers in their home countries. The difference was that usually English was examined as a course consisting of reading comprehension, grammar, writing with listening and speaking being gradually introduced there, instead of the individual course and testing for reading, writing, communication, grammar here at the Language Centre in Canada. Another difference was that in their home countries there had been lots of translation from their mother tongues into English and vice versa, while here at the Language Centre there had been none.

Nevertheless, those differences were superficial, that is, they could not affect the students' performance substantially as the core of English learning and

evaluation--grammar, vocabulary, reading comprehension, writing, listening were carried out in their home countries. The requirements for knowledge and skills remained close. That explained why students experienced little shock when they commenced their training in Canada with the new activities.

There had been writing of essays for the writing course and test, classroom presentations on chosen topics for communication course and test at the Language Centre. They were not familiar to most students when they entered the ESL training program at the Centre.

Though many of them had not experienced the kind of tests before, eight students said that they knew something about the test topics and subjects in their mother tongue (Survey item 13, November, 1996), and ten of them claimed that they enjoyed the format of tests at the Language Centre and had benefited a lot from them (item 12). One advantage, according to them, was that the test formats prepared them for the future studies -- the presentation and essay writing, for example, that were both common practice in academic studies. As well, and perhaps more importantly was that the requirement for grammar, vocabulary, writing skills, comprehension skills remained unchanged at the Centre. Because of this, they did not feel very much difference.

Human beings' history is a history of exploring and uncovering unknowns, from

land into oceans, into sky and space. It has been human beings' nature that we want to have control of our fate. We want to know our surroundings and we want to know what is going to happen to us. We don't like the unknown. The unknown means uncertainty, which consequently makes us feel insecure. We want to know and deal with what is going to happen to us. We want to have firm control of ourselves and our surroundings. The same was true for the students in dealing with tests. They didn't want the tests unknown to them or they could fail. They wanted not only to know what would be in the tests, but know that in as detailed a way as possible. Thus they could prepare and pass them (possibly with flying colours!). The situation was exactly what surfaced in the returned survey: they (16 students) were overwhelmingly for the idea that they could have done much better if they had known the test content better (item 14, November, 1996).

It appeared that most students talked more easily when they were out of class. Many of those the researcher talked with either in interviews, or casual conversations, appeared to be better, some of them much better, more talkative than they were in class presentations or tests. Moreover, they knew it. The reason they cited was that it was less pressured--one was not to be judged by one's peers and teachers;one was not to be scored. This certainly had something to do with exam anxiety.

Exam anxiety has been one of the important factors that prevents students from performing their best or to their potential in tests. Anxiety reduction before a test has been receiving much attention from professional circles. With anxiety reduced, students could perform better, and the scores from the test would be more informative to reflect the strengths and weaknesses of both the students' mastery of targeted knowledge and skills and teachers' design and delivery of curriculum.

One way of reducing exam anxiety is to have students do more exercises that have similar requirements to those of in formal tests. More pilot tests could also be conducted so the students become gradually familiar with them.

The familiarity with the format of, the knowledge about content of, and the time for preparing for an upcoming test enabled students to feel safer and less worried since they made the test less unknown and less uncertain. This was recognized by both teachers and students and practised in the Centre's program. They all showed a quite positive attitude toward the view and practice in the interviews and returned survey. fourteen students were in favour of the idea (and the Centre's practice) that some introduction about the coming test was made ahead of the test implementation (Survey item 15, November, 1996).

On the other hand, this familiarization is relative. You don't want to, nor can you

let students know exactly everything that is to be tested. We try to familiarize students with the format, the topics of tests in the same way we train people to use tools for production. It is students' responsibility to apply what they have learned to solve problems in tests, as well as in their life or work.

It is true that the more you know something, the better you may do it. Yet there should be a balance. Over two thirds of each test content had been or was to be discussed in class learning, and some others would be from the students' assignments outside class. You could not expect more than that. The researcher was wondering whether it was wise to let the students know exactly everything that was to be checked in tests. The teachers' practice concurred with the researcher's appraisal of this situation. Therefore, students' requirement for more information about test content was reasonable, nevertheless unjustifiable.

One thing many students pointed out repeatedly was that they still could not speak fluently, write good essays after about two and a half months' training; they still had difficulty in making themselves understood either orally or in writing, or both. As well, many a time they would find themselves unable to use proper words and expressions to communicate effectively with others. Because of that, ten of them claimed in the returned survey that the skills and knowledge checked in the tests had been little help for their future academic studies, while

only six thought otherwise.

It was reasonable that they wanted to achieve a lot from their training at the Language Centre. However, the length of the program, their entrance level, the length of time necessary for English fluency (in writing, speaking, or other areas) and other factors all determined that it was unlikely that great, dramatic achievement could be made within such a short training period, and that their expectations were not very realistic.

The students made visible progress; some of them were rather impressive. Yet, not everyone felt so because it was built up on a cumulative, gradual basis. There was always a comparison between themselves with people around them, especially with native speakers and writers, which made them feel inferior and far behind, and sometimes frustrated.

If those students had known that even native speakers and writers had to learn and improve themselves constantly (even as university students), they might feel better for what they had achieved in a relatively short period of time. As well, they might not have said that there was no relation between their learning and later academic studies in terms of English language knowledge and skills, because those presentations and essay writings they had been practising in classes are common practices in university studies. Therefore, the practices at

the Language Centre better prepared them for their further academic studies.

## Summary

Experiential groundedness means that we have to pay attention to students' prior and current knowledge and experiences and the knowledge, skills and topics which appear in tests. The form of questions and answers could all affect their performance in tests. Want of knowing and uncertainty may result in insecurity, which in turn leads to anxiety. Reduction of exam anxiety enables students to perform better and closer to their potential. The scores from the tests, therefore, reflect more closely the students' real English ability, which is the purpose of tests. The results from such tests are naturally authentic. The teachers tried to familiarize the students with test format and content as much as they could while the students thought they benefited a lot from the teachers' such efforts. As a result, their anxiety was reduced and the evaluation results were closer to their real English abilities.

A balance, however, is necessary when we try to relieve students from feeling uncertain and insecure in the tests. The relief should not be overdone. An informative test is one that neither puts too much pressure on students nor allows them to feel too easy or comfortable.

Students' expectations regarding what they might achieve overall in terms of their language abilities in so short a course of study also appear to have been very ambitious or overly optimistic for most.

In the next section, the survey instrument dealt with other evaluative factors considered important in this study.

## 5.4 Other factors (formative, summative, NRTs, and CRTs)

A test aims either at checking students' mastery of targeted knowledge and skills for decisions to be made on placement, admissions, et cetera, or at finding out their strengths and weaknesses from a unit of study for further modification of curriculum design and delivery.

Sometimes the researcher heard some students complain that they did not perform well in a certain test. Still, the result from it would be used for some important decision making. They felt this was not fair. Indeed, it was not very informative if one student happened to be sick either physically or emotionally during an test. There were simply too many odds that would result in the inaccuracy of a test result.

To reflect more accurately the students' targeted knowledge and skills ability, both teachers and students recognized the necessity for cumulative records of the students from tests, assignments, and learning activities. This way, even if one, or some students could not perform well in one, or a couple of tests, assignments, activities, the final results for their study would not be affected too much as long as they performed normally in the rest of tests, assignments, activities. Of course, one's performance may be atypical at some occasions. However, it can't be so all the time. If cumulative records were the basis of the final score, one's final result would be derived from the accumulated records of that overall performance. Table 4 is based on the data collected on other factors that may affect efficient and effective evaluation.

Table 4 : Other Factors in ESL Evaluation  (18 respondents)

| Q A | A | B | C | D | E |
|---|---|---|---|---|---|
| 1 6 | 0 | 1 | 2 | 9 | 6 |
| 1 7 | 2 | 1 4 | 2 | 0 | 0 |
| 1 8 | 0 | 1 3 | 4 | 1 | 0 |
| 1 9 | 7 | 1 1 | 0 | 0 | 0 |
| 2 0 | 0 | 7 | 6 | 5 | 0 |

Note 1: Horizontally, A (strongly agree) <------->E (strongly disagree);

2: Vertically, A=answer   Q=question

The students were uniformly in favour of the cumulative records as their final scores for studies (Survey item 19, November, 1996). The teachers and students (in the interviews) believed that only through the students' performance in tests, projects, exercises, assignments and classroom activities could their real English ability be closely reflected. All 18 students agreed their final scores should be drawn from the series of tests, assignments, projects, conducted from the beginning to the completion of the program.

To them, the merits of ongoing tests and projects were quite obvious: it was diagnostic in that the scores from each test and project could provide valuable feedback for teachers who could accordingly make adjustments and modifications in their consequent curriculum delivery; for students, their strengths and weaknesses would be displayed and they could adjust their study and narrow their focus on key area(s) that needed improving.

The cumulative, ongoing evaluation could, as has been contended by Jacobson (1982), examine the match between program and learner activities and provide feedback for both teachers and students in terms of their curriculum delivery, their performance in learning and evaluation. It is indeed more accurate and closer to the students' real ability since it illustrates the points of performance at the different phases of the learning spectrum, which could avoid the bias from the atypical performance of students in just one or two

tests. Logically, 15 students rejected overwhelmingly the idea that their final score from ESL should be the one they got from their final exam only (item 16). Only one student was in favour of one snapshot exam. However, there were five of them in favour one-two tests over five-six tests during their training while seven students were against this (item 20). The students might have misunderstood this item as it was almost the same question as item 16, only with the extra statement "This helped me to focus on learning". Otherwise it was difficult to explain their response to item 16.

A student would feel very happy if he/she got 90 or above from a test. What if most of the students got 90 or above in the same test? Similarly, he/she would feel depressed from his/her result in a test being 55. By comparison, it was too low. However, what would he/she feel if his/her peers all got scores below 60 in the same test? A comparison with others in the same condition might help one to see more objectively his/her standings in a test.

A NRT is helpful in telling people more objectively their English abilities. Citing the merits of NRTs, Brown and Yule (1983) note that each student's score on NRTs is interpreted relative to the scores of all other students in the same test so that each of them is spread along a continuum of scores and their location along the continuum immediately tells their abilities relative to the norms in ESL.

The students saw the merits of NRTs. The researcher (November, 1996) suggested "after each unit test, the instructor should let all students know their standings in class based on the exam or test" in order to have a more objective view of their standings with relation to others (survey item 18). The majority of them (13) thought it a good idea that they be informed of their evaluation standings in relation to each other. Only one student disagreed. Generally, it was felt that a student would not feel discouraged even if his/her score from a test was not very high if he/she knew most of his/her peers were not any better in the same test.

But one weakness of NRTs is that it gives people relative, instead of absolute information of performance of the students on specific tasks. One may be the best in class, yet he/she may not achieve what he/she is supposed to because of factors other than himself/herself or his/her performance. For example, students were asked to write ten sentences with past, present, present continuous, present perfect, future and future perfect tenses. One might be able to write sentences with only two tenses, yet he/she might pass the test because many of his/her peers did worse than him/her.

Suppose stakeholders need to know how well the students can perform on specific tasks so that they will be in a better position to determine the students' mastery of English. To accomplish this task, we have CRTs. Instead of relating

one student's score to his/her peers, CRTs "relate the scores directly to the performance of specific tasks, usually at a given number of levels of mastery" (Carroll & Hall, 1985, p. 5). A group of students might be able to write ten sentences with all the required tenses in a test except one who could only write nine sentences with required tenses. He/she was at the bottom of the group according to NRTs. However, he/she passed or should pass the test because his/her knowledge and skills of grammar and writing in English had no relation to how his/her peers performed but to the objective only. Shohamy (1985) notes that in CRTs "success is measured according to defined objectives" (p. 23).

Brown and Yule (1983) suggest that unit tests, projects and assignments (in CRTs) serve to measure "well-defined" and "fairly specific" instructional objectives: the results from each test, project and assignment are informative to the teachers who would accordingly adjust and modify their teaching and tests, allocating appropriate time and materials to the key areas in order to enable students to develop the knowledge and skills on specific objective, or a set of objectives. The data from the research proved the theory. Sixteen students supported the idea of unit test or assignment and thought it helpful for adjusting their focus (Survey item 17, November, 1996).

Summary

The merits of formative, on-going tests were obvious: they were diagnostic and informative. The final scores of the students accumulated from each test, project or assignment enables more accurate inference regarding the students' English language ability, even if they might have performed abnormally once, or twice previously.

The results from cumulative tests present a spectrum of students' study which is less biased and closer to their real capability in English.

On the basis of the results of our survey, it was noted that the series of tests, projects, assignments and other activities in NRTs provide the knowledge about individual students' standings on each of the activities in relation to others in class, give students' an objective view of their own studies and confidence and incentive to carry on and to catch up. The different ability of the students, their personal growth in study are shown through the cumulative record of the activities.

At the same time they (NRTs) are able to present teachers information on the adequacy and appropriateness of their learning and test, enabling them to

either stick to, or modify their curriculum.

On the other hand, CRTs are able to present a spectrum of students' study which is less biased, more informative than a single test by measuring students according to specific objectives.

The final area of the survey questionnaire, to which this study turns next, provided some opportunity for the students to respond to open-ended questions about ESL evaluation.

## 5.5 General questions

The researcher put forward some statements on things that the students might concern about in ESL evaluation. They included such things as involvement of students in test material selection, the combination of ESL training and academic studies, and additional help after class, et cetera. The data in Table 5 gives us an idea about the students' views on those things.

Table 5 : General Questions Regarding ESL Evaluation (18 respondents)

| Q\A | A | B | C | D | E |
|-----|-----|-----|-----|-----|-----|
| 2 1 | 1 0 | 3 | 5 | 0 | 0 |
| 2 2 | 5 | 5 | 8 | 0 | 0 |
| 2 3 | 1 1 | 2 | 5 | 0 | 0 |
| 2 4 | 1 1 | 0 | 7 | 0 | 0 |
| 2 5 | 1 2 | 2 | 4 | 0 | 0 |

Note 1: Horizontally, A=yes; B=no; C=I don't know;

2: Vertically, A=answer; Q=question

Many people like to know the process of our personal growth. We like to know how much we have achieved after a certain period of time. We especially like to compare our current status with that of the past. The comparison gives us a clear picture of our progress and may add to our confidence and incentive.

The same was true for the students at the Centre. They liked to know their personal growth, too. They liked to know how much they had achieved in English through their ESL training program. Ten students believed that a test at the beginning of their learning was good because the results could be compared with those of from the last test (Survey item 21, November, 1996). Only three did not want the pretest.

Because of the diversity of the students, they always had some different views on many things. For example, five students thought the test material selection

should involve them while the same number of students did not want the selection to involve them (item 22).

The students at the training program had some striking characteristics. One was that most of them wanted to study at the university's undergraduate program upon their successful ESL training at the Centre. As they had little idea of how academic study was carried out at Canadian universities, they were certainly eager to learn it, and that as early as possible. Consequently, 11 of them were in favour of the suggestion that one or two academic (undergraduate) courses should be conducted simultaneously with their ESL training (Survey item 23, Nov., 1996), and 12 of them agreed that some relation should be there in the test content between ESL training and their future academic pursuits (item 25). In each case, only two students disagreed.

It is a common practice that we, as learners, would not mind to receive extra help or consultancy in the area we are learning. In fact, many times we would like to receive as much help as we possibly can. Additional help does no harm to our learning. On the contrary, it sometimes is valuable and may facilitate our learning greatly.

The students at the Centre had the same point of view. Many of them (11) would like to receive some outside class help to reinforce their classroom learning

(item 24). According to some students, they especially enjoyed some native speakers as their listening and speaking companions.

# Chapter Six   Conclusion

In the concluding section of this study, the researcher will report on the key conclusions of the study, some recommendations for future, related research and the limitations of this study.

## 6.1 Conclusions drawn from the study

In the hypotheses formulated earlier in this research, it was put forward that integration of the content of tests with students' future academic pursuits will be discernible in the Language Centre's practice and in interviewees' testimonial about that practice and; a combination of formative evaluation with summative evaluation and of CRTs and NRTs, which helps teachers and students to gain a more objective picture of the students' overall proficiency, will be discernible in the Language Centre's practice and in the interviewees' testimonial about that practice. (See pp. 42-43)

Indeed, it was. Eighty-nine percent (16) of the students had positive views of the Centre's ESL program when they were asked what general impression they had had of that program. The words they used to describe their impression

included "good", "very good", "helpful," "useful," "very helpful," "so helpful".
There was no negative comment at all. The only two students who did not use
positive words had no comment at all. (Survey Item 26, November, 1996) Most
of them (13) thought the program helped them better prepare for the planned
undergraduate studies (Item 27).

It was extraordinary to have such a high approval rating of the program with
respect to the variety of the students' backgrounds and their diversified further
pursuits. The consensus only proves that the strategy and tactics the teachers
adopted at the Language Centre were correct and working, and convinced the
students in spite of their differences in country of origin, education experience,
culture and English level. This subjective, holistic response is all the more
impressive vis-a-vis the program , as many of the students also made it clear in
their survey responses that they had had higher expectations regarding their
language competence than they had been able to reach as a result of the
program. They did not appear to hold this against the program or teachers. One
can only conclude that they recoganized their own expectations were too high.

The Centre's experience showed that integration did not link rigidly the
students' specific subjects of future studies with on-going ESL training. Rather,
the teachers focused on the knowledge and skills that had and have been
widely used in university studies. This was flexible use of the concept of

connection between ESL training and students' future academic pursuits.

Based on the students' backgrounds and their future pursuits, it was impossible for the teachers' to connect ESL training with each students' specific future academic pursuit since they were so diversified. As well, it was unnecessary to do so since the students' main purpose for the time being was their English language ability improvement. It was additionally unrealistic to do so, as the teachers were trained for ESL teaching, not management, computers, liberal arts and the teaching of other subjects.

The teachers' focus on the teaching, especially their focus on validity, authenticity, experiential groundedness in the evaluation, proved successful and had been recognized by the students (Interview notes, October, '96). The feedback showed that the teachers' curriculum delivery was very appropriate and effective.

## 6.2 Some recommendations for future, related research

In spite of this generally positive impression of the program, however, students were not so positive when it came to the specific items about validity, authenticity, experiential groundedness, and other factors of an effective

evaluation. Students, consciously or unconsciously, would look at the items and answer them from their specific perspectives, with the intent of tailoring the learning and evaluation activities most suitable and adequate for them, or for the students of similar background and English level. Through the observations, casual conversations, interviews and survey, it was noticed that there were a couple of things on which the majority of the students (regardless of their differences) held the same or similar point(s) of view that needed to receive more attention from teachers and curriculum developers. For example, except for one (item 19), all the items in the survey showed that the students' had different views regarding validity, authenticity, experiential groundedness, and other factors in ESL evaluation. In fact, sometimes the students were matching in numbers in holding opposite points of view regarding specific items (items 1, 7, 8, 22). Those split views dealt respectively with validity (item 1), authenticity (items 7, 8), and student involvement in test material selection (item 22).

Some students expressed as their first priority for English training in Canada to improve their listening and speaking ability. The advantage they cited to learning English in Canada was that they were exposed to an English environment all the time in all places. Accordingly, they held great expectations from their English training in Canada.

They expressed their desire for more hours for communication class, which

allowed them more opportunities to practise the two (interviews and casual conversation notes, fall 1996). Therefore, it might be helpful to allocate more hours for communication class. Of course this had something to do with specific students' English levels, or their future endeavours.

Another thing was the balance within tests with regard to validity, authenticity and experiential groundedness. There was indeed a close connection between what was learned and what was tested. It was important to reduce the students' exam anxiety. A well-ballanced test, however, would have both this close connection, would reduce anxiety, and would have some challenge for students to apply the principle to actual problem solving as well. Too much pressure in a test was not good for reflecting the students' real English ability. No challenge was not good either for checking the students' English ability. Some students mentioned, as a matter of fact, that they liked essay writing and classroom presentations because they integrated closely classroom learning and actual application. They had to put something of their own into the activity, not just memorization. Moreover, such activities were useful in their later undergraduate or related studies.

It might be suggested also that a future research in this area investigate the subsequent academic and linguistic success of graduates of this program, who have gone on to study other subject matter as undergraduates.

## 6.3 Limitations of the study

There is always something to learn about the students. As Reid's (1987) survey of ESL students' learning style preferences has shown, ESL students from different language/educational/cultural backgrounds sometimes differ significantly in various ways from each other as well as from native speakers.

It is difficult to know a person well within three months. It was more difficult to know a person by talking, or watching him/her once a week in two and half months. The uniqueness each person brought to the ESL Advanced classes required more time for the researcher to learn, to appreciate and to have an in-depth understanding of that person.

Due to the length of time and scope of this study, the conclusions and thoughts that have been drawn may be limited to the same or similar context(s) and situation(s). The limited number of students interviewed, the time allocated for interviews, the questions used for the interviews, the researcher's specific background, may all have had some impact on his analysis, interpretation, and conclusions and thoughts drawn from the study. More time and more frequent attendance at class, longer interview times, closer relations with interviewees (both instructors and students) might lead to better understanding of students, more in-depth thoughts, more in-depth interpretation and conclusions.

## 6.4 Concluding Statement

As with all research, this study has enabled the researcher to find tentative responses to the questions raised and hypotheses tested. It has also opened up new avenues for further discovery and reflection relative to ESL evaluation.

# REFERENCES

Aitken, E.N. (1996). Gage Active Mathematics Assessment Pack. Toronto, ON: Gage Educational Publishing Company.

Alderson, J. & Urquhart, A. (1985). This test is unfair: I'm not an economist. Second Language Performance Testing . Hanptman, P.C., LeNlanc, R., Wesche, M. B. Ottawa, ON.: University of Ottawa Press.

Anderson, N. J. (1991). Individual differences in strategy in second language reading and testing. The Modern Language Journal, 75,(4), 460-472.

Ary, D., Jacobs, L. C. & Razavieh, A. (1985). Introduction to research in education. (3rd ed.). New York: CBS College Publishing.

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

Board of Education for the City of Etobicoke. (1987). Making the Grade: Evaluating Student Progress. Scarborough, ON: Prentice - Hall Canada.

119

Bogden, R. C., & Biklen, S. K. (1992). Qualitative research in education: An introduction to theory and methods. Toronto, ON: Allyn & Bacon.

Brown, G., & Yule, G. (1983). Discourse analysis. Cambridge, England: Cambridge University Press.

Burke, K. (1994). The mindful school: How to assess authentic learning. Palatine, IL: IRI/Skylight Publishing, Inc.

Carroll, B. J., & Hall, P. J. (1985). Make your own language tests: A practical guide to writing language performance tests. Oxford, England: Pergamon Press Ltd.

Clark, M. D. (1994). Dysfunctions of the theory/practice discourse. TESOL QUARTERLY, 28, (1), 9-26.

Clark, J. L. D. (1983). Language testing: Past and current status-- direction for the future. The Modern Language Journal, 67,(4), 431-443.

Courtland, M. C., & Gambell, T. J. (1994). Curriculum planning in the language arts K-12: An holistic perspective. North York, ON: Captus Press Inc.

Dewey, J. (1916/1966). Democracy and education. Miller & Seller. Curriculum Perspectives and Practices. Mississauga, ON: Copp Clark Pitman Ltd.

Gorden, C. M. & Hanauer, D. (1995). The interaction between task and meaning construction in EFL reading comprehension tests. TESOL QUARTERLY, 29,(2), 299-324.

Heffernan, P. J., Caouette, L., Bourassa, D. and Colbourne, L. (1996). FSL Evaluation Project: 13 Model Tests. Lethbridge, Alberta: University of Lethbridge/Edmonton, Alberta: Alberta Education.

Hill, C. & Parry, K. (1992). The test at the gate: Models of literacy in reading assessment. TESOL QUARTERLY, 26,(3), 433-461.

Jacobson, P. H. (1982) Using evaluation to improve foreign language education. The Modern Language Journal, 66,(3), 284-291.

Jarvis, G. A., & Adams, S. J. (1979). Evaluating a second language. Arlington, VA: The Centre for Applied Linguistics.

The Joint Advisory Committee. (1993). Principles for Fair Student Assessment Practices for Education in Canada. Edmonton, AB: Centre for Research in Applied Measurement and Evaluation.

Jones, R. L. (1985). Some basic considerations in testing and proficiency. (Eds.) New Directions in Language Testing. Lee, Y.P. Oxford, U.K: Pergamon Press.

Lynch, B. K. (1990). A context-adaptive model for program evaluation. TESOL QUARTERLY, 24, (1), 23-42.

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

Omaggio, A. (Ed.). (1979). Looking at the results. Building on Experience-Building for success, Skokie, IL: National Textbook Co.

Pennycook, A. (1989). The concept of method, interested knowledge, and the politics of language teaching. TESOL QUARTERLY, 23,(4), 589-618.

Perkins, K, & Brutten, S. R. (1988). A behavioural anchoring analysis of three ESL reading comprehension tests. TESOL QUARTERLY, 22, (4), 607-622.

Popham, W. J. (1995). Classroom assessment: what teachers need to know. Needham Hights, MA: Allyn & Bacon.

Raimes, A. (1990). The TOEFL test of written English: Causes for concern. TESOL QUARTERLY, 24, (3), 427-442.

Reid, J. M. (1987). The learning style preferences of ESL students. TESOL QUARTERLY, 21,(1), 87-115.

Schmidt-Rinehart, B. A. (1994). The effects of topic familiarity on second language listening comprehension. The Modern Language Journal, 78, (2), 179-189.

Shohamy, E. (1992). Beyond proficiency testing: A diagnostic feedback testing model for assessing foreign language learning. The Modern Language Journal, 76, (4), 513-521.

Shohamy, E. (1985). A Practical Handbook in Language Testing for the Second Language Teacher. Tel-Aviv, Israel: Shaked.

Shohamy, E. (1984). Input and output in language testing. Practice and problems in Language Testing, Culhane,T. (Ed). Essex, U.K: Department of Language Linguistics, University of Essex.

Statistics Canada. (1995). Education in Canada, 1995. Ottawa: Analytic Outputs and Marketing Section, Education, Culture, and Tourism Division.

Suomi, B. (Feb., 1991). TOEFL - 2000: An overview. Report presented to TOEFL Research-Committee. Princeton, N.J: Educational Testing Service.

Swain, M. (1985). Large-scale communicative language testing: A case study. (Eds.). New Directions in Language Testing. Oxford, U.K: Pergamon Press.

Educational Testing Service. (1996). TOEFL 1996 - 1997 Products & services catalogue. Princeton, N.J: Test of English as a Foreign Language P.O. Box 6155.

Webster's Third New International Dictionary. (1961).

# Appendix A

## Interview protocol (with instructors)

1. What are the sources most often used for (classroom) instruction?
   What is the norm/standard for selection of instruction materials?

2. Do you make any modifications for the materials used in class if
   they are adapted from books, journals, magazines, newspapers?
   Why (not)?

3. What do you think the students' (English) level?

4. How did you divide the classes into two (e.g., Class A, Class B,
   or intermediate, advanced)?

5. How long have you taught ESL?
   Do you see any affect of you teaching experiences on your
   current practice of teaching/evaluation?
   How?

6. I have noticed that there are two (students) missing today. Is it because they pay more attention to exam/test, or because of whetever other reason?

7. What do you think the students' involvement in classroom learning/teaching activities?

8. How do you musually design/develop your exam/test?

9. Do you have any idea about students' future plan (after their ESL learning)?
   If yes, what specific plans and your response (in organising learning and consequently evaluation)?

10. Is there any connection between their (students') future plan (subject area) and the present ESL learning/teaching at the two advanced level classes?

11. Some people (in ESL circle) argue that the closer link between instruction/learning and exam/test, the more accurate/reliable the data out of exam/test.

    What do you think?

12. What do you think of the connection between your course and other courses in the program (e.g., is there any discussion between you and your peers who teach other subjects of the same classes in terms of curriculum and evaluation development and/or delivery? If yes, how? If not, why not?)?

13. How do you see the relation between facts search and problem solving in terms of instruction and exam/test)?

    What is you solution to/recommendation for it (the relation)?

14. You mentioned that most articles were discussed and sought after based on the survey conducted at the beginning of semester.

    How do you see a teacher's role in ESL training and/or evaluation?

15. How do you balance your curriculum (e.g., how do you design and deliver your curriculum without favouring certain students and discouraging/discriminating others?)?

16. What does a project usually consist of?

    What is the percentage of each project in the final score?

    Why have you included

    projects as part of evaluation?

17. How many exams/tests altogether are there?

    What is the percentage of each exam/test in the students'/ examinees' final score?

18. What do you think the variety  of the students' background in country origin, education and/or other fields?

    Did/does the variety have any affect on the students'/ examinees' performance in learning/ examinations?

    How do you cope with the variety and possible consequence?

**Appendix   B**

**Interview   Protocol**   (with   students)

Which country are you from?

What was your education before you came to Canada (senior high school, college, university, or other)?

How long had you learned English before coming to Canada?

What is the difference in English instruction between your country and here in Canada?

What are the advantage(s) and/or disadvantage(s) of learning English here?

How was English evaluation conducted in your home country?

What is the difference in English evaluation between your country and Canada?

Which (English) evaluation is better?

Why?

What are you going to study after ESL study?

Are there any conflicts between the ESL exam and your future study?

How?

When do you think of the questions in ESL exams?

Are the exam forms familiar to you? (specify)

What do you think of the instruction in class?

Are there any connections between classroom instruction and exams?

Do you think there should be connection(s) or not?

Why?

Do you think the exams reflect your real English ability or not?

What would you do if you were to design and carry out an ESL exam?

What suggestions do you have to make?

**Appendix C**

**Questionnaire Survey**

The following are some questions. After each question, there are 5 answers ranging from A to E. Please use your pen or pencil to mark the answer that you think best reflects your thinking. Please be advised that your answers will be used for the purpose of the researcher's study only. Your academic status will in no way be affected by your participation in this activity. The information you provide is confidential and will be destroyed immediately after the study is completed.

1. I found that the contents of ESL exams/tests were easy to understand. I could answer questions without much difficulty.
   A. strongly agree          B. agree          C. no comment
   D. disagree                                E. strongly disagree

2. I still am poor in real-life English performance (reading, writing, listening, or speaking), though I got good scores in ESL exams/tests.
   A. strongly agree          B. agree          C. no comment
   D. disagree                                E. strongly disagree

3. The exams/tests (results) showed that I was always studying the wrong thing, though I focused on the text.

    A. strongly agree        B. agree        C. no comment

    D. disagree                E. strongly disagree

4. The exams/tests were good. My scores were good when I worked hard. They were not good when I didn't work hard.

    A. strongly agree        B. agree        C. no comment

    D. disagree                E. strongly disagree

5. The exam/test content should be prepared with more focus (e.g., have clear goal, target).

    A. strongly agree        B. agree        C. no comment

    D. disagree                E. strongly disagree

6. The exams (communication, grammar, reading, writing, etc.) focused too much on daily life (English) language such as shopping, greeting, etc. that I already knew before my ESL study here.

    A. strongly agree        B. agree        C. no comment

    D. disagree                E. strongly disagree

7. The sentences and/or paragraphs for the exams/tests were longer and more complicated/difficult than I usually met in classroom learning.

   A. strongly agree          B. agree              C. no comment

   C. disagree                           E. strongly disagree

8. There were topics in the exams that I knew little about.

   A. strongly agree          B. agree              C. no comment

   D. disagree                           E. strongly disagree

9. I did not do my exams well because they were not interesting to me.

   A. strongly agree          B. agree              C. no comment

   D. disagree                           E. strongly disagree

10. More exams should be conducted on site--library, computer lab, registrar's, book store, shopping mall, post office, et cetera.

    A. strongly agree          B. agree              C. no comment

    D. disagree                           E. strongly disagree

11. The forms/format of exams here were different from those in my home country.

    A. strongly agree          B. agree              C. no comment

    D. disagree                           E. strongly disagree

12. The exams/tests had little knowledge/skills that I will be able to use in my future academic study.

    A. strongly agree         B. agree         C. no comment

    D. disagree                E. strongly disagree

13. I did not even know some of the exam/test topics/subjects in my mother tongue.

    A. strongly agree         B. agree         C. no comment

    D. disagree                E. strongly disagree

14. I could have done much better if the test/exam content were familiar to me.

    A. strongly agree         B. agree         C. no comment

    D. disagree                E. strongly disagree

15. Before each exam/test, there should be some introduction about the exam/test, or some group discussion about it and preparation for it.

    A. strongly agree         B. agree         C. no comment

    D. disagree                E. strongly disagree

16. My final score for ESL study should be the score from the final exam.

    A. strongly agree         B. agree         C. no comment

    D. disagree                E. strongly disagree

17. The test after each study unit was very good. It told me about

    my strengths and weaknesses. It helped me to focus on my

    learning.

    A. strongly agree          B. agree              C. no comment

    D. disagree                          E. strongly disagree

18. After each unit test, the instructor should let me know my

    standing in class based on the test. This enables me to have a

    better picture of my learning compared with that of other

    students.

    A. strongly agree          B. agree              C. no comment

    D. disagree                          E. strongly disagree

19. The score from each unit exam/test should be part of my final

    score for ESL learning.

    A. strongly agree          B. agree              C. no comment

    D. disagree                          E. strongly disagree

20. I preferred one or two exams/tests over having five or six during

    the ESL learning. This helped me to focus more on learning.

    A. strongly agree          B. agree              C. no comment

    D. disagree                          E. strongly disagree

The followings are some statements. There are three responses after each statement. Please mark the response that is closest to your opinion.

21. At the beginning of my learning, there should be an exam/test. Each student's score from the pre-test should be used to compare with his/her final examination results in the course.

    A. Yes.                B. No.                C. I Don't know.

22. The materials for exams/tests should be selected by both teacher/examiner and students/examinees.

    A. Yes.                B. No.                C. I Don't know.

23. My ESL learning should be carried out simultaneously with one or two courses in my further academic (degree) program.

    A. Yes.                B. No.                C. I Don't know.

24. A resource person outside class would be more helpful in my ESL learning (like a tutor on regular hours).

    A. Yes.                B. No.                C. I Don't know.

25. There should a relation/tie between the eaxm/test content and the academic areas that I might pursue in the future.

    A. Yes.                B. No.                C. I Don't know.

**Please briefly answer the following questions:**

26. What is your general impression of the ESL program at the Center?

27. What are your future plans after the ESL program? How might the ESL program prepare you for your future studies?

28. What is your nationality (Which country are you from)? Is there any relation/tie between your nationality and your study here in Canada? How?

29. How is the ESL program at the Language Center compared with ESL in your home country?

30. What is your age (i.e. 15-19, 20-24, 25-29, 30-34, etc.)? What is your sex (i.e. female, male)? Do you believe they have had any impact on your ESL learning here?

Finally, you are more than welcome to raise any concerns or constructive observations that have something to do with your ESL learning and/or evaluation.

**Appendix   D**

**Questionnaire   Survey** (with categories)

The following are some questions. After each question, there are 5
answers ranging from A to E. Please use your pen or pencil to mark
the answer that you think best reflects your thinking. Please be
advised that your answers will be used for the purpose of the
researcher's study only. Your academic status will in no way be
affected by your participation in this activity. The information you
provide is confidential and will be destroyed immediately after the
study is completed.

**I.  Content  Validity  of  the  ESL  test**

1.  I found that the contents of ESL exams/tests were easy to
    understand. I could answer questions without much difficulty.
    A. strongly agree          B. agree              C. no comment
    D. disagree                          E. strongly disagree

2. The exams/tests were not meaningful because I still could say little, write little in English, though I got good scores in the exams/tests.

   A. strongly agree          B. agree          C. no comment

   D. disagree                          E. strongly disagree

3. The exams/tests showed that I was always studying the wrong thing, though I focused on the text.

   A. strongly agree          B. agree          C. no comment

   D. disagree                          E. strongly disagree

4. The exams/tests were good. My scores were good when I worked hard. They were not good when I didn't work hard.

   A. strongly agree          B. agree          C. no comment

   D. disagree                          E. strongly disagree

5. The exams/tests content should be further integrated.

   A. strongly agree          B. agree          C. no comment

   D. disagree                          E. strongly disagree

## II. Authenticity

6.  The exams focused adequately on daily life (English) language
    such as shopping, greeting, etc. that I had learned before my ESL
    study here.

    A. strongly agree          B. agree          C. no comment

    D. disagree                          E. strongly disagree

7.  The sentences and/or paragraphs were longer and more
    complicated than I usually met in classroom learning.

    A. strongly agree          B. agree          C. no comment

    C. disagree                          E. strongly disagree

8.  There were topics in the exams that I knew little about.

    A. strongly agree          B. agree          C. no comment

    D. disagree                          E. strongly disagree

9.  I did not do my exams well because they were not interesting to
    me.

    A. strongly agree          B. agree          C. no comment

    D. disagree                          E. strongly disagree

10. More exams should be conducted on site--library, computer lab,
    registrar's, book store, shopping mall, post office, et cetera.

    A. strongly agree          B. agree          C. no comment

    D. disagree                          E. strongly disagree

## III. Experiential Groundedness

11. The forms/format of exams here were different from those in my home country.

    A. strongly agree      B. agree      C. no comment

    D. disagree      E. strongly disagree

12. The exams had little connection with my future academic study.

    A. strongly agree      B. agree      C. no comment

    D. disagree      E. strongly disagree

13. I did not even know the exam topics in my mother tongue.

    A. strongly agree      B. agree      C. no comment

    D. disagree      E. strongly disagree

14. I could have done much better if the test/exam content were familiar to me.

    A. strongly agree      B. agree      C. no comment

    D. disagree      E. strongly disagree

15. Before each exam/test, there should be some introduction about the exam/test, or some group discussion about it and preparation for it.

    A. strongly agree      B. agree      C. no comment

    D. disagree      E. strongly disagree

## IV. Formative, Summative, CRTs, and NRTs

16. My final score for ESL study should be the score from the final exam.

   A. strongly agree        B. agree        C. no comment

   D. disagree        E. strongly disagree

17. The test after each study unit was very good. It told me about my strengths and weaknesses. It helped me to focus on my learning.

   A. strongly agree        B. agree        C. no comment

   D. disagree        E. strongly disagree

18. After each unit test, the instructor should let me know my standing in class based on the test. This enables me to have a better picture of my learning compared with that of other students.

   A. strongly agree        B. agree        C. no comment

   D. disagree        E. strongly disagree

19. The score from each unit exam/test should be part of my final score for ESL learning.

   A. strongly agree        B. agree        C. no comment

   D. disagree        E. strongly disagree

20. I preferred one or two exams during the ESL learning. This

helped me to focus more on learning.

A. strongly agree        B. agree                C. no comment

D. disagree                            E. strongly disagree


V. **The followings are some statements. There are three**

**responses after each statement. Please mark the**

**response that is closest to your opinion.**


21. At the beginning of my learning, there should be an exam/test.

Each student's score from the pre-test should be used to

compare with his/her final examination results inthe course.

A. Yes.                B. No.                C. I Don't know.

22. The materials for exams/tests should be selected by both

teacher/examiner and students/examinees.

A. Yes.                B. No.                C. I Don't know.

23. My ESL learning should be carried out simultaneously with one

or two courses in my further academic (degree) program.

A. Yes.                B. No.                C. I Don't know.

24. An advisor outside class would be more helpful in my ESL learning.

    A. Yes.                B. No.                C. I Don't know.

25.

    A. Yes.                B. No.                C. I Don't know.

VI. **Please briefly answer the following questions:**

26. What is your general impression of the ESL program at the Center?

27. What are your future plans after the ESL program? How is the ESL program connected with your future plans?

28. What is your nationality? How does your nationality affect your ESL learning?

29. How is the ESL program at the Language Center compared with ESL in your home country?

30. What is your age (i.e. 15-19, 20-24, 25-29, 30-34, etc.)? What is your sex (i.e. female, male)? Do you believe they have had any impact on your ESL learning here?

Finally, you are more than welcome to raise any concerns or constructive observations that have something to do with your ESL learning and/or evaluation.

**Appendix E**

**Consent Letters**

**Letter I** (to the director of the Language Centre)

Dear        , Director:

I am conducting a study of how to improve the effectiveness of ESL evaluation building on your Centre's current sound base. The purpose of this study is to investigate the possibility of increasing the validity, authenticity, and experiential groundedness of the ESL evaluation used so that the results from tests will reflect even more closely the students' English language ability. I anticipate the students (and the instructor who is to teach and test the students) at your Centre will benefit from participation in this study by having their English capacity reflected as objectively as possible in tests. I would like your permission for your employee (ESL instructor) and the students to participate in this study.

As part of this study, the instructor and a sample of 10% of (as well as two-three students for pilot interview protocol) the students will be asked to talk and write about what they think regarding the classroom instruction, test design and delivery. There will be some (informal) classroom observations and interviews based on a questionnaire survey for both instructor and students. Therefore, some extra time will be needed from them for participating in conversations and

interview (answering the questionnaire). Please note that all information will be handled in a confidential and professional manner. When responses are released, they will be reported in summary form only. Further, all names, locations and any other identifying information will not be included in any discussion of the results. You also have the right to withdraw the instructor and students at the Centre from the study without prejudice at any time. Participants' rights will always be considered first. They may withdraw their participation at any time without penalty. Their right to confidentiality will be guaranteed.

If you choose to do so, please indicate your willingness to allow the instructors and students (in one class) to participate by signing this letter in the space provided below.

I very much appreciate your assistance in this study. If you have any questions please feel free to call me at (403)329-2018 (O) or (403)329-6242 (H). Also feel free to contact the supervisor of my study Dr. Heffernan, Associate Dean, Faculty of Education at (403)329-2424 and/or any member of the Faculty of Education Human Subjects Research Committee, if you wish additional information. The chairperson of the committee is Dr. Robert Runte Faculty of Education.

Yours sincerely,

Guohua Pan    M.Ed. Candidate, Graduate Studies, Faculty of Education, the University of Lethbridge (403) 329-2018

(Please detach and forward the signed portion)

---

The validity, authenticity, and experiential groundedness of
English as A Second Language (ESL) Evaluation: A Case Study

I agree to allow the designated instructor,             , and
students of            (class), to participate in this study.

Name                 Signature

Date

**Letter II** (to the instructors)

Dear Ms./Mr, the Instructor:

I am conducting a study of how to improve the effectiveness of ESL evaluation building on your Centre's current sound base. The purpose of this study is to investigate the possibility of increasing the validity, authenticity, and experiential groundedness of the ESL evaluation used so that the results from tests will reflect even more closely the students' English language ability. I anticipate you will benefit from participation in this study by having the students' English capacity reflected as objectively as possible in tests. I would like your permission for your participation in and cooperation with this study.

As part of this study, you as instructor will be asked to talk and write about what you think regarding the classroom instruction, test design and delivery. There will be some (informal) classroom observations and interviews based on a questionnaire survey for both you and students. Therefore, some extra time will be needed from you for participating in conversations and interview (answering the questionnaire). Please note that all information will be handled in a confidential and professional manner. When responses are released, they will be reported in summary form only. Further, all names, locations and any other identifying information will not be included in any discussion of the results. You also have the right to withdraw from the study without prejudice at any time. Your rights will always be considered first. You may withdraw your participation at any time without penalty. Your right to confidentiality will be guaranteed.

If you choose to do so, please indicate your willingness to participate by signing this letter in the space provided below.

I very much appreciate your assistance in this study. If you have any questions please feel free to call me at (403)329-2018 (O) or (403)329-6242 (H). Also feel free to contact the supervisor of my study Dr. Heffernan, Associate Dean, Faculty of Education at (403)329-2424 and any member of the Faculty of Education Human Subjects Research Committee, if you wish additional information. The chairperson of the committee is Dr. Robert Runte Faculty of Education.

Yours sincerely,


Guohua Pan    M.Ed. Candidate, Graduate Studies, Faculty of Education, the University of Lethbridge (403) 329-2018


(Please detach and forward the signed portion)

-------------------------------------------------------------------------------------

The validity, authenticity, and experiential groundedness of
English as A Second Language (ESL) Evaluation: A Case Study

I,                                 , as the designated instructor, agree to participate in this study.

Name                          Signature

Date

**Letter III** (to the students)

Dear ESL Student:

I am conducting a study at the Language Centre. It is about ESL evaluation. The ESL program has been quite successful. The study will see if it can be even more successful. I will investigate the possibility of increasing the validity, authenticity, and experiential groundedness of the ESL evaluation. As an ESL student, you will benefit from participation in this study. You will know whether your examination results can better show your real English level. I would like your permission for your participation in and cooperation with this study.

As part of this study, I will ask you some questions. I will also ask you to write some of your ideas. They are all about ESL instruction, test design and delivery. I will talk to you, attend your class. And I will ask you to fill up questionnaire survey at the end of your study. Please note that all information will be kept confidential. Other people will not know anything that I ask you to say or write. When responses are released, they will be reported in summary form only. Your name, class, or this Language Centre will not appear in summary or any other documents about this study. The documents and summary will not mention where this study takes place, either. You can stop doing what I ask you at any time. Your stop will not have any affect on you and your study, your test scores. Your rights will always be considered first. You may withdraw your participation at any time without penalty. Your right to confidentiality will be guaranteed.

If you like to participate in this study, please sign this letter in the space below. And your signature shows that you are willing to participate in this study.

Your participation will be helpful for this study. I will appreciate your help very much. If you have any questions please feel free to call me at (403)329-2018 (O) or (403)329-6242 (H). Also feel free to contact the supervisor of my study Dr. Heffernan, Associate Dean, Faculty of Education at (403)329-2424. You are also welcome to contact any member of the Faculty of Education Human Subjects Research Committee, if you wish additional information. The chairperson of the committee is Dr. Robert Runte Faculty of Education.

Yours sincerely,

Guohua Pan     M.Ed. Candidate, Graduate Studies, Faculty of Education, the University of Lethbridge (403) 329-2018

(Please detach and forward the signed portion)

---

The validity, authenticity, and experiential groundedness of
English as A Second Language (ESL) Evaluation: A Case Study

I, _____ , as an ESL student of
(class), agree to participate in this study.


Name                          Signature

                              Date
sample