

**ADULT ESL STUDENT ATTITUDES TOWARDS  
PERFORMANCE-BASED ASSESSMENT**

**by**

**Lindsay Ann Brooks**

**A thesis submitted in conformity with the requirements  
for the degree of Master of Arts  
Department of Curriculum, Teaching and Learning  
Ontario Institute for Studies in Education of the  
University of Toronto**

**© Copyright by Lindsay Ann Brooks (1999)**



National Library  
of Canada

Acquisitions and  
Bibliographic Services

395 Wellington Street  
Ottawa ON K1A 0N4  
Canada

Bibliothèque nationale  
du Canada

Acquisitions et  
services bibliographiques

395, rue Wellington  
Ottawa ON K1A 0N4  
Canada

*Your file* *Votre référence*

*Our file* *Notre référence*

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

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

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

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

0-612-45956-X

Canada

**ADULT ESL STUDENT ATTITUDES TOWARDS  
PERFORMANCE-BASED ASSESSMENT**

by

**Lindsay Ann Brooks**

**A thesis submitted in conformity with the requirements  
for the degree of Master of Arts, 1999  
Department of Curriculum, Teaching and Learning  
Ontario Institute for Studies in Education of the  
University of Toronto**

**Abstract**

Performance-based assessment, such as portfolios, presentations, and participation, is currently being used in many second language programs. A review of the literature prior to this research revealed that although there have been a number of studies and papers on performance-based assessment, including alternative assessment and authentic assessment, few have reported student reactions to this wave of assessment techniques. Therefore, the attitudes of adult English as a Second Language (ESL) students (N=127) to performance-based assessment (portfolios, presentations, and participation) versus more traditional types of tests were surveyed by means of a questionnaire and semi-structured interviews. Quantitative and qualitative results suggest that the participants in the study perceived all four types of assessment positively. Analysis of background variables suggests that there were interaction effects for level of language proficiency and home country with regard to attitude toward assessment type. Other biographical variables showed little or no relationship to attitudes.

## Acknowledgments

Writing this thesis can be considered to be performance-based in that it has involved a long, stimulating process, culminating in this bound product. Many people helped me through the process and I would like to take this opportunity to thank them.

First and foremost, I wish to thank my supervisor, Dr. Merrill Swain, whose dedication and unwavering enthusiasm never cease to amaze me. With her expertise and insightful comments, she has patiently guided me right from the beginning. In addition to providing support throughout my thesis, she has also encouraged my development and growth in other academic pursuits and for everything she has done, I express my sincerest thanks and appreciation.

I would also like to thank Dr. Tony Lam, the other member of my thesis committee, for the guidance, fresh perspectives, attention to detail, and insightful comments that he brought to this process. It was in one of his classes that I picked up his enthusiasm and academic interest in performance-based assessment and so for that, I am also grateful.

I would like to express my gratitude to those at the two institutions in my study who granted me permission to conduct my research (both the pilot and my main study). To both of you, thank you. Also, I would like to take this opportunity to express my appreciation to all of the teacher and student participants at both institutions, without whom this research would not have been possible. I thank you all for your insights, comments, support and encouragement and moreover, the time you took to help me.



My appreciation also goes to Sue Elgie for her statistical help and guidance, to Hameed Esmaeili for his assistance in coding a sample of my data and for his overall helpfulness throughout my thesis, and to my classmates and colleagues for their ongoing encouragement and interest.

Finally, I would like to thank all of my friends and family from afar who provided support and encouragement even though they endured long lapses between phone calls and e-mail correspondence while I was in the throes of this process. To friends geographically closer, I thank you for being understanding of my general unavailability over the past months. To Kut and Dimitri, who have witnessed this performance first hand on a day to day basis, I would like to say a simple thanks and to make a promise to tidy up all those journal articles and books that have decorated the living room.

## Table of Contents

Abstract / ii
Acknowledgments / iii
Table of Contents / v
List of Tables / viii
List of Figures / x
1. Introduction / 1
Statement of the Research Problem / 1
General Description of Performance-based Assessment / 4
The Research Questions / 6
Rationale for the Study / 7
2. Literature Review / 9
A Brief History of Performance-based Testing / 9
The Precommunicative Era / 9
The Communicative Era to the Present / 13
Definitions of Performance-based Assessment / 14
Benefits of Performance-based Assessment / 16
Issues in Performance-based Assessment / 18
Fairness / 18
Validity and Reliability / 18
Student Attitudes to Different Forms of Assessment / 21
3. Design and Methodology / 30
The Origin of the Study – the Pre-Pilot / 30
The Pilot Study – the Evolution of the Idea / 30
The Context / 30
The Participants / 31
Instruments Used / 31
Data Collection and Analysis / 34
The Main Study / 35
The Context / 35
The Participants / 36
Student Participants / 36
Teacher Participants / 40
Instruments Used / 41
Procedures Followed / 42
Data Collection and Analysis / 44

4. **Quantitative Results / 48**
  - Results of the Factor Analysis of the Scale / 48**
  - Subscale Reliability / 49**
  - General Response to the Four Assessment Types / 49**
    - Student Attitudes / 49**
    - Teachers' Perceptions of Student Attitudes / 52**
  - Background Variable Results / 54**
    - Testing the Assumptions of Normality and Sphericity / 54**
    - Background Variable of Level of Language Proficiency / 55**
    - The Language Background Variable / 58**
    - The Background Variable of Home Country / 59**
    - The Level of Education Variable / 62**
  - Results of the Time Item / 63**
  - Results of the Secondary Factor Analysis / 64**
  - Summary of the Quantitative Results / 66**
  
5. **Qualitative Results / 70**
  - The Qualitative Data from the Questionnaires / 70**
  - The Qualitative Data from the Interviews/ 76**
    - Experience with the Four Types of Assessment / 76**
      - The Students' Responses / 76**
      - The Teachers' Responses / 77**
    - Definitions of the Different Types of Assessment / 77**
      - Tests / 77**
      - Portfolios / 79**
      - Presentations / 80**
      - Participation / 81**
      - Summary of Teachers' Definitions / 82**
    - Likes and Dislikes about the Four Types of Assessment - the Student Responses / 83**
      - Tests / 83**
      - Portfolios / 85**
      - Presentations / 86**
      - Participation / 87**
    - Student Likes and Dislikes about the Four Types of Assessment - the Teacher Responses / 89**
    - Ranking of the Assessment Methods / 90**
      - Students' Responses / 90**
      - Teachers' Perceptions / 92**
    - Other Ways of Being Assessed / 92**
    - Issues Arising in the Interviews / 93**
      - Marks in a Relatively Low Stakes Environment / 93**
      - Marking Criteria / 94**
    - Summary of the Qualitative Results / 95**

- 6. Discussion / 98
  - Attitudes to Performance-based Assessment / 98
  - Resistance to Performance-based Assessment? / 99
  - Attitudes to Traditional Tests versus Performance-based Assessment / 100
  - Background Variables / 101
    - Language Proficiency Variable / 101
    - Language and Country Variables / 103
    - Level of Education Variable / 104
    - Other Background Variables / 105
  - Teachers' Perceptions of Student Attitudes to Types of Assessment / 106
  - Issues in Performance-based Assessment / 107
    - Fairness / 107
    - Validity / 108
  - Pedagogy vs. Assessment / 110
  - Limitations of the Study / 112
    - Scale Reliability and Validity / 112
    - Preference vs. Attitude / 113
    - Individuality of Preferences / 113
    - The Connection Between Attitude, Marks and Performance / 114
  - Implications of the Findings / 114
- 7. References / 116
- 8. Appendices / 124
  - Appendix A: Student Questionnaire / 124
  - Appendix B: Teacher Questionnaire / 128
  - Appendix C: Interview Questions for Student Interviews / 132
  - Appendix D: Interview Questions for Teacher Interviews / 133
  - Appendix E: Procedures for Administering the Questionnaires / 134
  - Appendix F: Letters of Informed Consent / 135
  - Appendix G: Letters Requesting an Interview / 137
  - Appendix H: Letters of Informed Consent for the Interviews / 139
  - Appendix I: Typical Comments for Each of the Codes for the Four Types Of Assessment / 141
  - Appendix J: Student Definitions of a Test / 149
  - Appendix K: Student Definitions of a Portfolio / 151
  - Appendix L: Student Definitions of a Presentation / 153
  - Appendix M: Student Definitions of Participation / 156

## List of Tables

- 3.1 Categories Used to Code the Qualitative Data from the Questionnaires / 46
- 4.1 Factor Loadings / 48
- 4.2 Descriptive Statistics for the Four Assessment Types / 50
- 4.3 Descriptive Statistics for the Teachers' Perceptions of Student Attitudes to the Four Assessment Types / 52
- 4.4 Comparison of the Students' and the Teachers' Ranking of the Four Assessment Types based on the Mean Scores / 53
- 4.5 Tests of Normality for each of the Four Subconstructs / 54
- 4.6 Descriptive Statistics for the Variable Level of Language Proficiency / 56
- 4.7 Results of the Tests of Between-Subjects Effects for the Variable Language Background / 58
- 4.8 Descriptive Statistics for the Variable Language Background / 59
- 4.9 Results of a Repeated Measures ANOVA on Student Preference Ratings for Assessment Type and the Independent Variable of Home Country / 60
- 4.10 Descriptive Statistics for the Variable Home Country / 62
- 4.11 Results of the Tests of Between-Subjects Effects for the Variable Level of Education / 63
- 4.12 Descriptive Statistics for Item 7 for the Three Levels of Language Proficiency / 64
- 4.13 Rotated Factor Loadings in the Secondary Factor Analysis Conducted on each Subconstruct / 65
- 5.1 Tally of the Type and Number of Comments from the Qualitative Data / 70
- 5.2 Comments Reflecting the Idea of the Assessment Being an Opportunity to Learn / 71
- 5.3 Comments Reflecting the Idea of the Assessment being Motivational / 72

- 5.4 **Results of Coding the Questionnaire Comments for the Content of the Response / 73**
- 5.5 **Comments Suggesting that Marks are of Little Consequence to the Participants / 74**
- 5.6 **Comments about Marking Criteria and Equity / 75**
- 5.7 **Student Definitions of a Test / 78**
- 5.8 **Student Definitions of a Portfolio / 80**
- 5.9 **Student Definitions of a Presentation / 81**
- 5.10 **Student Definitions of Participation / 82**
- 5.11 **Results of Coding the Interview Responses about the four Assessment Types / 84**
- 5.12 **Comments on Concerns about Marking Participation / 88**
- 5.13 **Comments Indicating and Ambivalence Towards Marks / 93**
- 5.14 **Comments about Marking and Marking Criteria / 95**

## List of Figures

- 3.1 Countries represented by the participants / 38
- 4.1 The results of a tally of frequencies of each participant's highest mean rating of the four assessment types / 51
- 4.2 Interaction plot of the estimated marginal means for level of language proficiency and the four types of assessment / 57
- 4.3 Interaction plot of the estimated marginal means for attitudes towards assessment type for home country / 61
- 5.1 Ranking of assessment preferences in the interviews / 90

## CHAPTER 1

### Introduction

#### Statement of the Research Problem

In examining the educational literature, it does not take long before the terms performance-based assessment, alternative assessment or authentic assessment crop up. In most of this literature, the focus is on the perceived advantages of these assessment methods (e.g. Baker & O'Neil, 1996; Herman, Aschbacher, & Winters, 1992; O'Malley & Pierce, 1996; Wiggins, 1993). While it is not difficult to find advocates, as well as a few skeptics, of these assessment practices, what appears to be strikingly absent from the current research is the perspective of the students. If the students *are* mentioned in the literature, it is often in the context that performance-based assessment may be unfamiliar or culturally inappropriate for some groups and it may be difficult to convince students, particularly adults, of the pedagogical value of such assessment methods (Burt & Keenan, 1995; Wrigley & Guth, 1992). Furthermore, the focus of much of the literature on performance-based assessment seems to be on equity issues of second language (L2) learners in a first language (L1) classroom in an elementary or high school setting (Cummins, 1984; Damico, 1991; Darling-Hammond, 1994) or in a college or university writing program (Hamp-Lyons, 1996; Murray, 1994). One form of performance-based *testing* that has received some attention in the literature has been oral proficiency interviews and students' reactions to them (e.g. Shohamy, 1982, 1983a, 1983b) but little research has been conducted on other forms of performance-based assessment in a second language or an adult English as a Second Language (ESL) context. Although in the past



few years many publications have extensively discussed or debated the theoretical basis and practical implementation of performance-based assessment in the context of second language learning and teaching (e.g. McNamara, 1996; Norris, Brown, Hudson, & Yoshioka, 1998; Shohamy, 1995), there is a dearth of literature on the responses of adult ESL students to performance-based assessment. The use of performance-based assessment in second language programs is not new but what is new is the increasingly higher stakes of these assessment practices.

English as a Second Language programs associated with colleges or universities increasingly act as points of entry for international students into the post-secondary institutions. As these programs gain more recognition and are viewed not as ancillary but rather as integral to the internationalization of an institution, the onus is on these programs to provide transparent methods of assessment that have validity according to their host institution. This need is due in large part to the intent to have completion of such English programs satisfy the admissions language requirements in lieu of standardized tests such as the Test of English as a Foreign Language (TOEFL) or the Michigan English Language Assessment Battery (MELAB), which are currently being used for admission purposes despite contrary instructions from the publishers that the test scores not be the only basis for such decisions. Furthermore, as arguably most in the field of second language education would concur, the active, process-oriented assessment practices possible at the classroom level can provide a much richer picture of students' language proficiency than can standardized tests administered *en masse*. Having said that, it behooves English programs to be able to demonstrate that their assessment practices have validity while at the same time continuing to assess students using

methods reflective of the types of communicative teaching and learning that are currently accepted as best practice. Even the Educational Testing Service, publisher of the TOEFL, the most widely used standardized English language test, is moving towards using more performance assessment in the test so that students can be assessed in “a manner that more closely resembles the tasks they would be required to perform in an academic setting, while also improving washback,” (Carey, 1996, p. 1).

If, for reasons such as those stated above, English language programs do implement summative forms of assessment, it seems likely that performance-based assessment would be essential, in addition to more traditional types of testing, so that the methods of assessment more closely resemble the pedagogy and teaching methodology currently prevalent in the language classroom. These assessment methods reflect a balance of approaches that teachers have generally always used to assess their students in a formative way but now because the stakes are potentially higher, there is a need to provide more accountability and transparency for both the post-secondary institutions and the learners themselves, the biggest stakeholders in any assessment process. Even for those students who do not intend to pursue further studies at a college or university, the assessment practices constitute part of the whole learning experience and if the learners react adversely to being assessed, they will likely find other programs that better match their needs.

Although individual teachers in the classroom setting can get feedback at the classroom level on student reactions to different types of assessment practices, there is little in the literature on general learner attitudes to performance-based assessment on a larger scale, despite the claims that it may be unfamiliar or culturally inappropriate (Burt

& Keenan, 1995; Wrigley & Guth, 1992). Therefore, in light of what seems to be a gap in the current research, this study was conducted with the purpose of asking adult students in an ESL program their attitudes towards performance-based assessment.

### General Description of Performance-based Assessment

As the literature review in the next chapter will highlight, there seems to be no consensus as to an accepted definition of performance-based assessment. In the literature, different terms are used either synonymously with or in conjunction with performance-based assessment. Among those, alternative assessment or authentic assessment seem to be the most common. For the purposes of this study, the term performance-based assessment is preferable since it makes explicit the notion of the assessment task being to elicit a performance or production from the learners in written or oral form or in the form of a demonstrated behaviour. The key notion is that the learners demonstrate this output. What is implicit in performance-based assessment is the idea of a process approach to learning and assessment - not an emphasis on a one-off, on demand testing situation. The distinction between tests and assessment is particularly salient in this thesis, as student attitudes towards performance-based assessment, in particular, portfolios, presentations, and participation, will be compared to those that they hold towards tests.

The main distinction between performance-based assessment and tests is that in the former the process is key whereas in the latter the product or result is more the focus although performance-based assessment can involve both products and processes, as will be outlined in the next chapter. Portfolios entail the students' producing, revising, and editing their writing over the course of a program and then possibly discussing the

contents of the portfolio collaboratively with their teachers and their peers. In doing presentations, students prepare, rehearse, and finally present their work to their classmates and teachers. The presentations themselves can allow for multiple opportunities for students to interact and communicate with others. Participation, because it involves the in-class performance of students over an extended period of time, gives students many opportunities to perform, whether it be through class discussions, asking questions in class, cooperating in group work or even just attending class. What is common to all three forms of performance-based assessment in this study is that they require interaction between learners, their classmates, and their teachers so in a sense, they can be considered to encompass both input and output strategies. Performance-based assessment has been described as part of a constructivist philosophy as it involves a two-way interaction between learners and their environment (Yawkey, Gonzalez, & Juan, 1994). Therefore, although the emphasis of performance-based assessment is on *demonstration*, the final product tends to be but one of the outcomes as there are multiple opportunities for interaction and learning in the process of the assessment (see Swain, 1984).

Just as there is no universally accepted definition of performance-based assessment, so there is no agreement on a definition of a test, although some versions will be proffered in the next chapter. For the purposes of this study, a test is loosely characterized as focusing on the product, as having either correct or incorrect answers, and as involving a limited time to respond. Tests of this nature typically include multiple-choice or fill in the blank formats. Perhaps the most defining feature of tests of this sort is that although students could be said to be performing in a sense in that they

have to write something on the test paper, they are not interacting with the assessment instrument. Such a testing situation involves an on-demand, at the moment response to test questions within the time constraints of the class or the testing period. This definition was adopted in the present study, perhaps unfairly, so that a process versus product continuum between performance-based assessment and tests could be delineated.

### The Research Questions

Because performance-based assessment is such a broad construct, it was decided to limit the questions to three areas of performance-based assessment: portfolios, presentations, and participation and compare learner attitudes to these three assessment types with their attitudes towards more traditional tests. The questions posed for this study, therefore, are:

How do adult ESL students feel about three different forms of performance-based assessment - portfolios, presentations, and participation?

How true are claims that students from some cultures will resist these three forms of performance-based assessment methods?

How do adult ESL students feel about traditional tests as compared to performance-based assessment?

What is the relationship between adult ESL students' biographical variables and their attitudes towards performance-based assessment and tests?

What do teachers think about students' attitudes to the four types of assessment: tests, portfolios, presentations, and participation?

### Rationale for the Study

Through exploring the answers to these questions about learners' attitudes to the different forms of assessment, it may be possible for teachers, administrators and researchers to better address student needs and alleviate any possible tension between pedagogical intention and student perception. Because assessment practices should reflect curricular objectives, and since performance-based assessment lends itself to potentially positive washback effects, it seems that this method of assessment will continue to play an increasingly important role in many English language programs. However, it is important to know how students feel about getting marked on performance-based assessment because doing so can possibly help in the implementation of such assessment procedures or the decision whether to implement them at all.

Being aware of students' attitudes and perceptions toward different assessment methods is particularly important for those involved in English programs that have implemented or are thinking of adopting formalized assessment whether it be high stakes or otherwise. The reality for such English language programs is that the learners, in order to be satisfied with all aspects of the program, need to be assured that the assessment procedures are transparent, fair, and relevant to their learning and the post-secondary institutions need to be assured that the assessment procedures represent valid and reliable indicators of language proficiency.

Another reason for the import of surveying student attitudes is to give learners a voice in the assessment process. As Cray and Currie (1996) note in the context of including learners in teacher education, "Despite the emergence of learner-centred

approaches, we still tend to do what we deem best for learners. Even when these learners are adults, we tend to act for them rather than ask what they think,” (p. 117). The literature on performance-based assessment is replete with the views of teachers, administrators, and researchers but the learners, in the literature at least, seem to have been in large part overlooked.

## CHAPTER 2

### Literature Review

#### A Brief History of Performance-based Testing

Although the literature seems to be increasingly focused on performance-based assessment, the concept has existed in the context of second language testing for more than 100 years (Spolsky, 1995) but has been used more consistently in the field for the past 40 years (McNamara, 1997). Two major forces spurred on this development of performance-based assessment. The first was the need to test the language proficiency of foreign students prior to their studies in the North America or Britain. The second was the need to adopt testing procedures that would more closely mirror the changes in teaching practice that arose in response to theories of communicative competence (McNamara, 1996). However, prior to the first theories of communicative competence (e.g. Hymes, 1967), performance-based assessment had been used in the context of testing second languages for two decades (McNamara, 1996) although often in tandem or as a complement to discrete-point testing.

#### The Precommunicative Era

Predating performance-based testing, in the precommunicative era, testing of discrete-point skills predominated in the context of second language testing. The prevailing theory at the time, represented in the works of linguists such as Lado (1964), was one of a structuralist approach to viewing language – any language was simply the sum of its component parts and that language could be learned, and thus tested, according to a behaviourist paradigm. Behaviourists thought that repetition and practice of sounds



and structures would result in the formation of new language habits. Measuring the acquisition of such habits, therefore, involved creating discrete-point items that would test each of these sounds and structures. Discrete-point items could be scored objectively and without examiner judgment as either correct or incorrect and the sum of the items, or really the sum of what were thought to be the component parts of language, was used as a measure of proficiency. Lado (1960) felt that one of the promises of discrete-point testing was that potentially subjective judgments could be eliminated from the testing process and that it was possible "... to break away from having to ask the student to speak when we test his ability to speak, since this process is inaccurate and uneconomical," (in Barnwell, 1996). Fortunately, Lado's narrow view of language testing, with its concomitant dismissal of the need to ask candidates to speak in order to assess their speaking proficiency, was not the only approach to language testing at that time.

In the 1950s, although the discrete-point approach was still the norm, some performance-based methods were making their way into the testing field. At that time the Foreign Service Institute in the United States started to use an oral format to test the productive language skills of potential personnel for overseas postings. This test, the Foreign Service Interview (FSI), was performance-based in that candidates were rated on their ability to perform and demonstrate their language proficiency in an interview conducted by two people (Barnwell, 1996). However, because the focus of the rating was on discrete skills, such as grammar and pronunciation, the FSI was within the bounds of the structuralist theories of the time (McNamara, 1996). Some of the tests developed in the 1960s to deal with the increasing number of English as a second language students

applying for admission to universities in Britain and North America included a performance component in the form of an essay (McNamara, 1997).

In the early 1960s, the call for using more integrative tests in addition to the discrete-point approach came in J.B. Carroll's seminal recommendation for, "an approach requiring an integrated, facile performance on the part of the examinee ... [and] tests in which there is less attention paid to specific structure points or lexicon than to the total communicative effect of an utterance," (1961 [1972, p. 318]). Because Carroll's comments came in the context of testing for admission to English medium institutions, they reflected the pragmatic need for the test to reflect the purpose. Thus, test purpose driving the test format rather than linguistic theory being the determining factor became established since at that time, a theoretical basis of performance did not yet exist (McNamara, 1996). It was not until the development of theories of communicative competence in the late 1960s that performance-based assessment had a theoretical grounding.

In the 1970s with both the communicative movement in the classrooms and the articulation of theories of communicative competence (Hymes, 1967; Savignon, 1972), performance-based testing became entrenched as being reflective of the practice at the time. Savignon (1972) dismissed the discrete-point approach that had predated the communicative approach she proposed, namely to assess language skills, in "an act of communication" (p. 11). However, McNamara (1996) claims that Savignon's approach, although grounded in Hymes's theory, is atheoretical in defining what constitutes performance. The relevance of a theoretical basis of communicative competence in the

context of second language testing did not appear until Canale and Swain's (1980) framework in which communicative competence was divided into three sub-competencies consisting of grammatical competence, sociolinguistic competence, and strategic competence. A fourth component, discourse competence was included in later publications (e.g. Canale, 1983). Canale and Swain argued that communicative language testing must tap into the extent to which learners are able to demonstrate competence in a meaningful communicative performance (Weir, 1990), involving multiple interacting factors. Jones (1985) highlights this notion of demonstration of several facets of language ability:

With regard to second language performance testing it must be kept in mind that language is only one of several factors being evaluated. The overall criterion is the successful completion of a task in which the use of language is essential. A performance test is more than a basic proficiency test of communicative competence in that it is related to some kind of performance task. It is entirely possible for some examinees to compensate for low language proficiency by astuteness in other areas... (p. 20)

Since the influential Canale and Swain framework of communicative competence, there have been subsequent attempts to build on that model (e.g. Bachman, 1990) but as yet no comprehensive framework has been developed that accounts for the performance dimension of performance-based language tests (McNamara, 1996; Shohamy, 1995). McNamara (1996) characterizes this problem of this conceptualization of performance as a Pandora's Box that has yet to be sorted out. Although communicative competence

theory provided a justification for performance-based assessment, the exact nature of a theory of performance seems to be still open to debate.

### The Communicative Era to the Present

The 1980s saw the beginning of the communicative era of language assessment in which language tests were developed that would elicit the production of language in either an oral or written form and that would as much as possible simulate real-world tasks. Such tests included role-plays, group discussions, and oral reports. It was during this time that the distinction between proficiency testing and achievement testing was delineated, with the former embodying the notion of communicative competence and the latter measuring knowledge of language usually in the context of a particular course of study (Shohamy, 1997a). Swain (1984) outlined four principles of communicative language testing for assessing communicative competence: starting from a theoretical framework, concentrating on both the content and the task-type in assessment design, eliciting learners' best performance, and working for washback in involving teachers in the design and assessment of the communicative language tests. These principles seem eminently suited for performance-based assessment.

When viewed at the classroom level, the current trend in using performance-based assessment may or may not be new. With the rise of communicative language teaching in the 1980s, and before the current proliferation of literature about performance-based assessment, teachers might have routinely chosen assessment methods more reflective of the communicative language classroom. The only difference might have been in terminology rather than in actual assessment practices. For almost two decades, teachers

may have used performance-based methods such as presentations, participation, and more recently portfolios, in addition to more traditional tests, to assess their learners. As Barnwell (1996) points out, in order to write a complete history of language testing, it would be necessary to find out exactly what teachers were doing in the language classrooms but as he also points out, detailed information of this type is not available.

### Definitions of Performance-based Assessment

The term performance-based assessment is often used when referring to a broad spectrum of assessment types. Among those are alternative assessments and authentic assessments. However, some important distinctions in terminology between these terms should be noted. Authentic assessment can be defined as a special kind of performance assessment conducted in an authentic context as part of regular classroom learning rather than as contrived, intrusive assessment tasks (Gipps, 1994). Alternative assessment, as characterized by Aschbacher (1991) requires problem solving and higher level thinking, involves tasks that are worthwhile as instructional activities, uses real-world contexts or simulations, focuses on processes as well as products, and encourages disclosure of standards and criteria. What is important to note is that although performance-based assessment can be authentic, it is not necessarily so (Meyer, 1992) and performance-based assessment does not inherently include all the characteristics of alternative assessment. Performance-based assessment then, is a general term encompassing many aspects of both authentic assessment and alternative assessment.

Although a general description of performance-based assessment has already been provided in the first chapter, it is worthwhile to note the multitude of definitions in the

literature. Frechtling (1991) offers perhaps the broadest definition in describing performance-based assessment as “anything that is not a multiple-choice paper and pencil test,” (p. 24). One definition given by Fitzpatrick and Morrison (1971) is that a performance test is “one in which some criterion situation is simulated to a much greater degree than is represented by the usual paper-and-pencil test” (p. 238). Because performance-based assessment can include and often does include writing, the paper-and-pencil versus performance dichotomy is not particularly accurate. However, even Fitzpatrick and Morrison go on to state that performance tests and other kinds of tests do not necessarily have a clear boundary although the former does have some unique qualities. They characterize performance-based assessment as including both processes and products but caution that processes should only be assessed if the steps in the process have been explicitly taught. Conversely, if the procedures to be followed have not been taught, the products should be assessed rather than the processes. Similarly, Messick (1994) distinguishes between two traditions of performance-based assessment: one in which the performance, or the product, is the target of the assessment, and one in which the performance, or the process, is the vehicle of assessment.

Mehrens (1992) defines performance tests as requiring “heavy reliance on observation and professional judgment in the evaluation of the response” (p. 3). This notion of rater judgment is key to many of the definitions of performance-based assessment (e.g. McNamara, 1996; Norris et al., 1998; Pierce & O’Malley, 1992). Stiggins (1987) simply describes performance assessment as involving judgment and observation. Another key element of any definition of performance-based assessment is on demonstration of ability. Jones (1985) offers the following definition, “an applied

performance test measures performance on tasks requiring the application of learning in an actual or simulated setting,” (p. 16). Shohamy (1983b), in the context of oral proficiency assessment, also highlights that in performance-based assessment, the performance involves application of knowledge in a communicative activity. She contrasts performance assessment with assessment of knowledge, the former involving a behavioural component and the latter focusing on linguistic accuracy. Yet another definition in a similar vein is that performance-based assessment is the rating of a behaviour or classroom activity that would occur even if assessment were not the purpose (Haertel, 1992). In the context of the present study, this definition seems particularly relevant as all three forms of performance-based assessment, portfolios, presentations, and participation, would likely occur in some form or another in most second language classrooms.

### Benefits of Performance-based Assessment

A common theme in the current literature is to praise the benefits of performance-based assessment. The literature contains many references claiming that performance-based assessment holds great promise for L2 students (Tannenbaum, 1996; Yap, 1993). Teachers, administrators, and other stakeholders have embraced these assessment measures but little evidence exists to indicate whether students are equally as enthusiastic about this trend. In fact, there are few empirical studies of the attitudes of *any* of the stakeholders in the assessment context (Hamp-Lyons, 1996, p. 151). Many authors caution that using performance-based assessment with L2 students may be difficult due to cultural differences and expectations (Burt & Keenan, 1995; Nunan, 1995; Wrigley & Guth, 1992) but there is a paucity of studies to support these claims.

In the literature, one of the rationales provided as to why performance-based assessment should be implemented is that it can result in positive washback (Miller & Legg, 1993; Norris et al., 1998; Shohamy, 1995) as the assessment method more closely matches the type of instruction in a communicative language classroom. The view that assessment plays a fundamental role in any language curriculum is widely held (e.g. Bachman, 1990; Brown & Hudson, 1998) and some propose that performance-based assessment should be an integral consideration in curricular decisions (Shohamy, 1982; Short, 1993). Short (1993) suggests that L2 students may respond more positively to alternative forms of assessment such as performance-based assessment than to traditional types of testing. She further states that alternative assessment, including performance-based assessment, provides a more accurate demonstration of student ability than traditional assessment. Other benefits of using performance assessments are summarized in Norris et al. (1998).

Hamp-Lyons (1996) outlines why portfolio assessment can be positive for ESL students. Among the reasons for using portfolio assessment in an ESL context, she cites that the process approach to portfolios allows ESL writers time to notice and correct their grammar mistakes. Other benefits of this form of assessment are that learners can notice their own progress through their portfolios and also learners are able to be more involved in the assessment process through interaction with their teachers and classmates. However, Hamp-Lyons also highlights the need for studies in which student attitudes are surveyed for, as she notes, "We still have almost no data on students' views of or responses to portfolio assessment of writing," (1996, p. 161). In one study, although not with ESL students, Baker (1993) found that it made no difference in terms of students'



attitudes to writing, whether their teachers used traditional writing assessment methods such as in-class essays or portfolios.

### Issues in Performance-based Assessment

#### Fairness

Many of the proponents of performance-based assessment cite that, particularly in a mixed L1/L2 context, these assessment methods are fairer or more equitable for second language learners. However, others caution that fairness is not necessarily intrinsic in performance-based assessment (Darling-Hammond, 1994; Hamp-Lyons, 1996; Lam, 1995; Linn, Baker, & Dunbar, 1991; Mehrens, 1992; Supovitz & Brennan, 1997). Frechtling (1991) in discussing this issue cautions that “simply shifting from a product-oriented assessment to a process-oriented assessment does not guarantee an unbiased test,” (p. 24). In looking specifically at portfolio assessment, Hamp-Lyons (1996) states that little support exists for the claims that portfolio assessment is any fairer to ESL students than more traditional timed-writing tests. In one study in an L1 context on whether alternative assessment methods would result in greater equity than standardized-tests, Supovitz and Brennan (1997) found they did not. Inequities with regard to gender, socioeconomic status, and race and ethnicity were found in both alternative assessment and standardized tests.

#### Validity and Reliability

In the literature, alongside the benefits of performance-based assessment, are drawbacks or concerns about these methods of assessment, particularly when viewed from a psychometric perspective. Performance assessment is not inherently problem-free

(Hambleton & Murphy, 1992). In the field of second language testing, some of the psychometric concerns are compounded because of the complexity of having to describe language ability, let alone putting a score to it. As Spolsky (1968) states, “the central problem of foreign-language testing, as of all testing, is validity,” (p. 68). One of the criticisms that plagues performance-based assessment concerns this issue of validity, particularly in generalizing from a language performance to a criterion behaviour. Because performance-based assessment reflects an actual performance, it is often erroneously assumed to have high validity, especially when compared with other testing formats such as multiple-choice (Linn, Baker, & Dunbar, 1991). However, many theorists are of the view that performance-based assessment must be held to the same validity criteria and psychometric rigor as other assessments (Carey, 1996; Messick, 1994).

When scrutinized under the same psychometric confines as standardized-testing, performance-based assessment, such as oral proficiency testing, is typically characterized as being subjective and unreliable (Shohamy, 1983b). Swain (1993) proposes that given the complex, almost predictable variability of performance on different language tasks, the traditional notions of reliability are perhaps not appropriate and that alternative criteria should be applied to judge the quality of communicative assessments. Hamp-Lyons (1996, 1997) echoes this view that the difficulties of using the traditional psychometric notions of reliability and validity become apparent when applied to performance-based assessment, such as portfolios. Others concur that traditional measurement principles such as validity, or at least the criteria that are typically used to establish such traditional measures, do not necessarily have a place in performance

assessment (Linn, Baker, & Dunbar, 1991; Miller & Legg, 1993; Moss, 1994; Moss et al., 1992; Wolf, Bixby, Glenn, & Gardner, 1991) or in language assessment (Weir, 1990). As an alternative to traditional psychometric notions, Moss (1994) calls for a hermeneutic approach to assessment that would:

...involve holistic, integrative interpretations of collected performances that seek to understand the whole in light of its parts, that privilege readers who are most knowledgeable about the context in which the assessment occurs, and that ground those interpretations not only in the textual and contextual evidence available, but also in a rational debate among the community of interpreters. (p. 7)

That is to say, a hermeneutic approach would differ from a psychometric approach in that instead of examining the parts of an assessment and holding them to the rigors of traditional notions of validity, the whole is judged and the validity would stem from this well-informed judgment by the rater (in most cases the teachers). However, the issues of interrater and intrarater reliability tend to be problematic in performance-based assessment (Houston, Raymond, & Svec, 1991; McNamara, 1996; Yen, 1997) although rater training and the use of multiple raters can improve reliability (Bachman & Palmer, 1996) as can ensuring the rating scale has been validated from a theoretical perspective of the construct being measured (Brindley, 1998). In fact, often rater reliability has been found to be higher than score reliability (Dunbar, Koretz, & Hoover, 1991). As for the general issue of reliability, Moss (1994) does not advocate its abandonment but rather proposes that it is possible to have validity without reliability depending on the “context

and purposes for assessment” (p. 10). This notion seems particularly relevant in the context of second language assessment.

When viewed in terms of validity, perhaps the easiest form of validity to apply to performance-based assessment, such as portfolios, is face validity (Hamp-Lyons, 1996). Although according to Messick (1989), and as outlined in his validity framework, face validity is not a form of validity in the technical sense, he does not deny that it is important for participants in the assessment process to view the assessment as relevant and that efforts should be made to avoid face “invalidity” (in Wiggins, 1993). Other researchers (Hill, 1998; Nevo, 1985; Zeidner, 1990) advocate getting direct feedback from the participants in order to assess the face validity of an assessment method. However, it should be noted that while the importance of face validity cannot be denied, it cannot replace other types of validity. Direct performance-based assessments have been described as “appear[ing] to have the potential of enhancing validity,” (Linn, Baker, & Dunbar, 1991, p. 16) and having “seductive face validity,” (Aschbacher, 1991, p. 277). However, without knowing student attitudes towards getting marked on performance-based assessment, it is impossible to determine if it has this “seductive” quality.

#### Student Attitudes to Different Forms of Assessment

In one study of attitudes of students towards assessment in general, Alcorso and Kalantzis (1985), in a survey of 59 learners in the Adult Migrant Education Program (AMEP) in Australia, found that learners thought assessment to be an important component of their language learning experience. The same study revealed that 80% of the learners indicated a preference for more summative means of assessment such as

tests, rather than informal formative assessment methods that seemed to be commonly used in that program. Alcorso and Kalantzis also state that less-confident learners demonstrated a negative attitude towards assessment and viewed tests as unnecessary. A positive correlation between degree of preference for formal tests and levels of education was also discovered. In another study of 332 learners in the AMEP context, Kessler (1984) corroborates this finding between level of education and a desire for formal assessment. Brindley conducted a study in 1984 in which he surveyed 50 adult learners and drew a similar conclusion as the above two studies: students from backgrounds in which formal testing is emphasized in the educational system, expect and prefer formalized assessment procedures (in Brindley, 1989).

Another study seemingly contradicts those in which a correlation between level of education and assessment preferences was found. In a study involving 517 immigrant learners of English as a second language, Willing (1988) investigated whether there was a correlation between variables (such as language or cultural background, age, education, speaking proficiency) and learning style differences. He found that none of the biographical variables, including education, correlated with particular learning preferences and that it would be inaccurate to make sweeping statements about particular groups. Willing sampled a considerably larger pool of learners than the researchers in the previous studies so his claims would seem to be somewhat more reliable. His conclusion about biographical variables is in part corroborated by Zeidner (1988), who in looking at student preferences for types of tests, found that background biographic variables are “generally weak predictors of test attitudes,” (p. 83). Although both Willing and Zeidner reached similar conclusions regarding background variables, their dependent variables

differed with the former involving learning differences and the latter examining attitudes towards tests.

Scott and Madsen (1983) conducted a study with 73 adult ESL learners from Spanish and Japanese backgrounds and did find a correlation between cultural background and affective responses to different types of testing. Overall, Spanish-speaking students rated a battery of tests more positively than did Japanese-speaking students. For both language groups, an oral proficiency interview was the preferred test type over a grammar test, a reading test, and a listening exam. In the same study, students with lower levels of language proficiency did not rate the oral interview test as highly as did those students with higher levels of language proficiency. Scott and Madsen also found that all of the participants in the study gave more positive ratings on most of the affective measures with each successive administration of the test battery. However, due to the small sample size, such conclusions are only suggestive.

In the high school context, Grierson (1995) surveyed 33 students from different language and cultural backgrounds in an Intensive English Centre in Australia and noted a tension between the informal, spontaneous approach to assessment preferred by the teachers in that program to the formal testing procedures that students rated as their assessment of choice. However, in the same study, students did rank teacher-student discussions and corrections as their second and third preferences for assessment, indicating a desire for some interaction in the assessment process. Again, given the small sample size of this study, any conclusions are questionable at best.

In other research looking at student preferences for different types of assessment, Zeidner (1990) conducted a study with 100 participants in which he assessed college students' attitudes towards essay versus multiple-choice tests and found that students felt that essays reflected their ability or knowledge more than did the multiple-choice item format. Additionally, essay tests were perceived as more appropriate for addressing course objectives. On the other hand, students viewed multiple-choice tests more positively than they viewed essay tests in terms of the time needed for preparation and in terms of stress. In another study by Zeidner (1990), 279 undergraduate students indicated that they felt essays were fairer than multiple-choice tests. Zeidner (1990) also surveyed a group of 80 graduate students with similar results in that they indicated a preference for being evaluated on papers rather than on exams. These findings, although not on the types of assessment in the present study, do suggest that the learners preferred the more performance-type methods of assessment along the continuum between multiple-choice and performance-based assessment.

In another study looking at multiple-choice tests versus other types of testing, albeit not in the context of second language learning, 800 American high school students were surveyed to see if they perceived any differences between alternative assessment and traditional types of tests used to assess their ability and knowledge in mathematics (Herman, Klein, & Wakai, 1997). The general conclusion drawn in this study was that students found the alternative assessments to be more meaningful and motivating than traditional multiple-choice tasks. Background correlates such as gender and socioeconomic status revealed little difference in perceptions although the results

suggested that economically advantaged students were more receptive and comfortable with the alternative assessment.

Although no studies have been found on student attitudes to getting marked on presentations, there are several in which learner attitudes to oral tests have been surveyed. In one study in an English as Foreign Language (EFL) context, 170 learners in an advanced reading course were surveyed on their attitudes to written versus oral tests by means of a feedback inventory (Zeidner & Bensoussan, 1988) in which the participants filled out a semantic differential scale, an ethnic anxiety questionnaire, and a short personal data inventory. Results indicated that the learners showed a preference for written tests as the students felt that written tests were better and more appropriate for reflecting their knowledge than oral tests, and that chances of success were higher with them. Students' lack of enthusiasm for oral tests was attributed to the feeling that oral exams produced more anxiety, pressure and tension than did written tests. The researchers found that there was a slightly higher preference for tests from learners from some cultural backgrounds. As Zeidner and Benoussan caution, these findings are by no means generalizable as the sample population was not randomly selected, the learners had little experience with oral tests and the attitude survey may or may not have been a reliable instrument. Furthermore, previous studies by Savignon (1972) and Shohamy (1982) do not corroborate the findings of Zeidner and Benoussan. Both Savignon and Shohamy indicate that learners perceived oral tests and interviews as being low-anxiety test situations although in these studies learner preferences for oral interviews as opposed to traditional tests were not compared.



In the context of second language assessment, some of the literature seems to indicate that students show a preference for traditional tests as they are viewed as being formal in that they are well-defined, structured and assigned a mark. However, there is nothing inherently “informal” about performance-based assessment - criteria for success can make it just as formal as needed in the context in which it is used. What remains to be fully examined is how students react to getting assessed on performance-based tasks in a formalized way where marks are assigned. The literature on student attitudes to being marked on, let alone just doing, portfolios, presentations, and participation is virtually non-existent. In a presentation at a conference, Bane (1999) reported on a study in which he surveyed student reactions to Internet reading portfolios. Students read articles on-line and then wrote responses as a record of what they had read. Initial results from his study indicate that students reacted positively to the assignment and perceived that they improved in language development and reading and research skills. In a study on project work with ESL students, Eyring (1997) makes a call for more studies documenting student attitudes, presumably towards collaboration, so that implementation of such projects can be improved.

In an attempt to answer this call to survey student reactions to performance-based assessment, a study was conducted (Brooks, 1998) to examine adult student attitudes towards getting marked on portfolios, presentations, and participation. The results suggested that at the time, students in the non-credit Intensive English Program (N=87) had a positive attitude to these types of assessment. The results demonstrated a trend towards an increasingly positive attitude towards the three assessment types with increasing level of proficiency. The language background variable in this exploratory

study did show statistical significance in that different linguistic groups had varying attitudes to the three types of performance-based assessment. None of the other background variables (length of time in Canada, age, gender, previous experience with the three types of assessment) were found to be statistically significant with regard to attitude to performance-based assessment.

In the pilot study for this present study, 113 adult learners in an English as a Second Language program at a large university in Canada were surveyed for their attitudes towards getting marked on four assessment types: tests, portfolios, presentations, and participation. Results suggest that all four methods of assessment were rated positively with presentations significantly preferred over tests, portfolios and participation, respectively. In reporting these results it should be noted that portfolios were not used in every class that participated so the rating for that assessment type may not be truly indicative of the participants' attitudes to getting marked on that form of performance-based assessment. None of the background biographic variables such as age, gender, language background or education correlated significantly. However, for the proficiency variable, a significant difference in the overall means for type of assessment was found and in addition, an interaction effect of assessment type and level of proficiency was detected. Although beginner (N=47), intermediate (N=39) and advanced (N=27) students all indicated a significant preference for presentations, the attitude ratings for the assessment types seemed to depend on the level of language proficiency. Attitude ratings for the advanced level students showed the most marked differences between tests and the three forms of performance-based assessment, with students indicating a preference for portfolios, presentations, and participation over tests. It could

be that advanced level students, because of their language proficiency, felt more comfortable with performance-based assessment than did students with lower levels of proficiency. However, this notion of increasing preference for performance-based assessment with increasing levels of proficiency may not be true for participation in that intermediate students had a significantly lower mean for this type of assessment than did both beginner and advanced students. For the background variable of home country (N=83), only the attitudes of students from Japan (N=55), Korea (N=15) and Mexico (N=13), were analyzed as the other country groups had too few participants. An interaction effect between attitudes to the types of assessment and home country was significant. Students from all three countries rated presentations the most highly. For students from Japan and Korea, participation was the least preferred type of assessment whereas for students from Mexico, tests had the lowest mean score. As in most studies of this type, especially in examining correlation with background variables, the small sample sizes provide possibly suggestive results rather than conclusive findings.

As has already been mentioned, the literature on students' responses to performance-based assessment is rather thin. However, knowing student attitudes is a genre of research that seems to need some attention. Over a decade ago, Alderson (1986) predicted that with the growing concern that learners be involved in curricular decisions, test developers would also start to consult and survey learner views on the method and content of tests. However, his call for these innovations appears to have gone largely unheeded, at least from the perspective of the published literature. As Alderson (1986) states:

Students typically have considerable experience of their own language learning and so might be expected to have opinions on how they might best be assessed. Future innovations in testing should perhaps pay more attention to the students' own informed view on assessment and on the methods which will enable them to perform to the best of their ability. (p. 99)

It follows that in order to implement assessment methods that provide learners with optimal performance conditions, their attitudes towards the currently used performance-based assessment should be known. This premise is based on the simple notion that attitude to assessment can affect attitude to subject matter, which can in turn affect learning. Furthermore, attitude to assessment can affect performance in that students may not perform to the best of their abilities if they hold negative views towards the assessment type. There seems to be a need to find out what students think and this necessitates further research in this area of student attitudes.

## CHAPTER 3

### Design and Methodology

#### The Origin of the Study – the Pre-Pilot

As has already been stated, the idea for this study was born when a literature review on performance-based assessment revealed no shortage of opinions of researchers, administrators, and teachers but the attitudes and reactions of the students themselves appeared to be missing. Therefore, a study was conducted in an attempt to survey the general attitudes of adult ESL students in a non-credit Intensive English Program (IEP) based at a large Canadian university. In this original study, participants filled out a questionnaire to survey their attitudes towards three forms of performance-based assessment used in the program: portfolios, presentations, and participation. However, because of the small sample size in that initial study, no conclusions could be drawn at that time. Based on the results and limitations experienced in this pre-pilot study, refinements and additions were included in the pilot to the main study.

#### The Pilot Study – the Evolution of the Idea

##### The Context

The pilot study took place in about week nine of a twelve-week non-credit program for adult ESL students based at a large Canadian university, different from that of the original study and the main study. In this program, similar to the one in the main study, students are given grades and are marked using a variety of assessment methods including three of the assessment methods of interest: tests, presentations, and participation. Portfolios are used to some extent by several of the instructors but their use

is not program-wide. Apart from the use of portfolios, the nature of the program, the teaching methodologies, and the student population closely matched those of the main study.

### The Participants

The participants for the pilot study included 113 volunteers from all nine classes in the first time slot of the program. In addition, all 11 teachers in the program volunteered to participate. The student participants ranged in age from 18 to over 38 and came from 18 different countries. Of the 113 student participants 29 were male and 84 were female. The participants were classified as beginner, intermediate, or advanced based on their class levels at the time of the study.

### Instruments Used

In the development of the original instrument used in the pre-pilot study, first the construct was broadly defined as performance-based assessment, with three potential subconstructs of portfolios, presentations, and participation, recognizing, however, that in some classes, presentations may be part of the participation mark. Although the construct encompassed more than just these three forms of assessment, it was still felt that performance-based assessment defined the nature of the subconstructs. In the design of the instrument, several considerations limited the scope and length of the questionnaire. Foremost of these was the need to write items that would be relatively easy for students from elementary to advanced levels of English to understand. A second consideration was the effort to provide enough items so that reliability of the instrument could be checked but not so many as to be a burden for the participants to answer or be too much

of a disruption to the teachers. With these factors in mind, the original questionnaire was developed. Teachers whose classes participated in the pre-pilot study reported that the questionnaire took the students between 10 and 15 minutes to complete.

After the pre-pilot study, the questionnaire was modified based on feedback received from both students and teachers and on some re-thinking on the part of the researcher. The most significant change from the original instrument was the inclusion of tests under each item. In the pre-pilot study, tests were not included among the assessment types but rather were treated as items themselves in which students were asked to compare their opinions about performance-based assessment to their opinions about tests. However, in revising the questionnaire for the pilot study, it was decided that a more complete comparison between tests and performance-based assessment could be made if the items were the same for each assessment type. Therefore, the items in the questionnaire were refined and in some cases, rewritten completely to reflect this change. In the original questionnaire, students were asked to indicate their attitudes to performance-based assessment versus tests in only three of the items. However, in the revised questionnaire for this study, traditional tests were included under all 10 items so there were in effect 40 Likert-type scale items because there were 10 items answered for each of the four types of assessment.

At the end of the questionnaires in the pre-pilot study, participants were asked to circle whether they generally liked or disliked each form of performance-based assessment and then to indicate why. However, for the pilot study this was changed to a

more open-ended format in which participants were asked what they liked and/or disliked about each of the four types of assessment.

Another addition to the pilot study that was not present in the original pre-pilot, was the inclusion of teacher questionnaires asking the teachers their perceptions of student attitudes to the four types of assessment. The items in the teacher questionnaires were identical to those in the student questionnaires with the exception that the items were written so that teachers would rate what they thought the *students'* attitudes to the four types of assessment were. The purpose of the teacher questionnaires was three-fold, one being to check what types of assessment were actually used in the classroom as the teachers would be the most familiar with that, the second being to see what teachers' perceptions of students' attitudes were and how closely they matched those of the students' and the third being to provide a means of selecting teachers for interviews. For the main study, teachers with differing views on student attitudes to the four assessment types would be selected for interviews.

Whereas interviews had not been a part of the pre-pilot study, they were added to the pilot study so that a more in-depth picture of student attitudes to the four types of assessment could be conducted. In the pilot study, one student from each of the three levels of language proficiency was interviewed as well as two teachers, one who taught at the beginner level and one who taught at the intermediate level. The purpose of these interviews was to see if the questions were clear and elicited relevant responses that would provide insight for shedding light on the questions of the study.



### Data Collection and Analysis

As already stated, all nine classes in the program participated in the pilot study. Teachers were given envelopes with letters of informed consent and questionnaires, as well as instructions for administering the questionnaire. The class level was written on the outside of the envelope but not the class name or anything else that would identify the class or the teacher. After each class had completed the questionnaires, a volunteer student returned the sealed envelope to the receptionist at the institution. After the questionnaires had been turned in, the students in one of the classes in the program (a mixed intermediate/advanced class) were asked if they had had any difficulty in filling out the questionnaire. None of the students reported any difficulty although two students found a couple of the items quite similar. A couple of teachers reported that they had difficulty in filling out the questionnaire as they found the terms too vague and were reluctant to generalize. However, the other teachers did not report any such difficulty and several of them commented that they found it interesting to reflect on their perceptions of student reactions to the methods of assessment.

Before proceeding with the main study, the quantitative results of the pilot study were analyzed to check for problematic items. Item analysis was conducted using the computer program SPSS 8.0.0 (Statistical Product and Service Solutions, 1997) and an alpha scale reliability of .83 was calculated. At the time, none of the items stood out as being problematic so no changes were made to the scale itself. The qualitative data elicited from the questionnaires were analyzed and the types (and number!) of responses elicited seemed to indicate that the questions were clear to the participants. The interviews were listened to in order to ensure that the questions elicited responses

appropriate for answering the questions of the study. Also, the participants in the pilot study, both students and teachers, were asked if any of the questions were unclear, ambiguous, or confusing. None of the participants reported any concern or difficulty with the questions. Subsequently, but after conducting the main study, the results were fully analyzed so that a report of student attitudes could be provided to the institution where the pilot study had been conducted. A brief summary of the findings is included in the previous chapter.

### The Main Study

#### The Context

The study took place in the last three weeks of a 9-week non-credit Intensive English Program (IEP) in the Continuing Studies division of a large Canadian university. Students generally take three classes a day with each time slot of class having a different skill and content focus. Although the institution is non-credit granting, formal assessment methods were introduced into the program just over one year prior to the beginning of this study. Formal assessment was implemented as part of the newly adopted curriculum of the program. Another reason for more formalized assessment at the institution is that completion of the most advanced proficiency level will soon satisfy the university admission language requirements<sup>1</sup>. As a result, the criteria for successful completion of each level and the assessment procedures needed to be formally documented. Teachers are required to submit marks for all of the students at the end of each session. In terms of assessment, the motivation for students to do well is that they must pass the requirements for each level either to move to the next one or to receive a

---

<sup>1</sup> Senate approval has already been granted. This change will take effect in a year's time.

certificate if they do not plan on studying for another session. Successful completion of each of the first five levels entails completing the three core classes at any given level with a grade of 65% or more. Furthermore, those students intending to pursue degree programs at the university must achieve at least 80% in the sixth level of the program to satisfy the university language proficiency requirements. Additionally, those students who wish to obtain a certificate of advanced proficiency must get a grade of 80% in each of the three core courses in the top level of the program.

The possible assessment weightings for each class are broadly outlined in the still-evolving curriculum so that the teachers can select which of the assessment methods listed best meet the needs of their students. Because multiple assessment methods are used in each class, and given the requirement that the students take three classes, the participants in the study had been exposed to many different forms of assessment in their programs with different weightings for each form. For example, in some classes, participation may count for up to 25% of the final grade whereas in others, it may only be worth 10%. Similarly, the weighting for portfolios and presentations varies from class to class, depending on the skill focus of the class and depending on the individual teachers. Although the weighting of assessment methods is different for many of the classes, the weighting has been standardized for several of the classes with multiple sections.

### The Participants

#### Student participants.

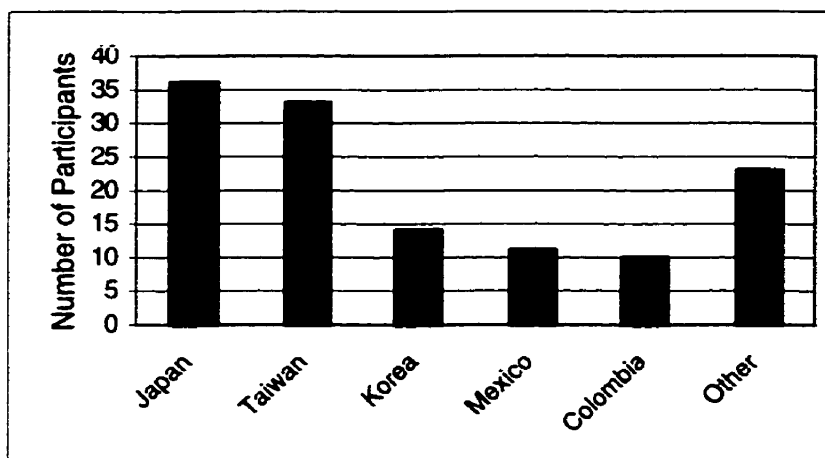
The 127 participants in this study were drawn from a population of adult ESL students in an Intensive English Program. They were at six different levels of proficiency

- lower elementary, upper elementary, lower intermediate, upper intermediate, lower advanced and upper advanced. Students are placed in each level based on the results of an in-house placement test for those new to the program or based on successful completion of the previous level class for those students who have been in the program for one or more sessions. The placement test consists of five components, an oral interview, a listening test, a grammar test, a reading test, and a writing sample. The raw scores from the tests appropriate to the skill area are used for class placement. For example, the listening and oral interview scores are blended and used to place the students in classes focusing on those two skill areas. Therefore, and this is relevant for the study, it is possible that students could be in a different level for each skill area. This will be discussed later as a possible source of error, particularly when looking at student attitudes with respect to level of language proficiency.

The potential student population for the study was 180. However, two classes did not participate (N=30) and some of the remaining potential students were either absent on the day the questionnaires were distributed or had left the program early as the study was conducted towards the end of the session (N=17). In total, 133 questionnaires were filled out but 6 were discarded because they were incomplete. That left the sample size for this study as 127 volunteer participants from the potential population of 180 enrolled in the second class time slot of the program. This time slot was chosen because it is the one from which the largest potential student sample could be drawn as all students were enrolled in classes at that time. A sample of convenience was used for the questionnaire part of this study since between class effects at any particular level were not a

consideration. The sample for student interviews was selected based on their willingness to be interviewed, their level of proficiency, and their responses to the questionnaire.

The six class levels were analyzed as three levels of proficiency with the lower and upper elementary classes comprising the beginner level (N=38); the lower and upper intermediate classes comprising the intermediate level (N=59) and the lower and upper advanced classes comprising the advanced level (N=30). Eighty-one of the participants were female and 46 of the participants were male. The participants came from 23 different countries with most coming from one of five countries: Colombia (N=10), Japan (N=36), Korea (N=14), Mexico (N=11) and Taiwan (N=33) as shown in Figure 1. The other eighteen countries had four or fewer participants, accounting for 23 of the total population of the study.



**Figure 3.1.** Countries represented by the participants. Included in the category of “other”: Argentina, Bangladesh, Brazil, Canada, China, Germany, Iran, Italy, Panama, Peru, Russia, Saudi Arabia, Spain, Switzerland, Thailand, Tunisia, and United Arab Emirates.

---

Most of the students fell into one of three age categories: 23 participants were in the 18-22 year old category, 63 were in the 23-27 year old category and 15 were in the 28-32 year old category. The 33-37 and the 38+ age categories each had three participants. One hundred and seven participants had post-secondary education. Ninety-three either had attended or were attending college or university; 10 were in or had completed graduate school and 4 were in or had finished technical school. The remaining 20 participants were high school graduates. Most of the participants had been in Canada for more than 2 months, with 65 indicating that they had been in the country 2-4 months, 26 for 5-8 months, 4 for 9-12 months, 3 had been in Canada more than a year, and 1 was from Quebec. The other 28 participants had been in the country for less than 2 months, indicating that at the time of the study, they were in their first English program since arriving. Reasons for studying English fell into one of four main categories, for college or university (N=33), for college and job (N=32), for job (N=42), and for travel (N=16). Participants' previous experience with the four types of assessment varied with 32% of the respondents having experience with tests only (N=41), 10% with tests and participation (N=13), 9% with tests and presentations (N=12), 17% with tests, presentations and participation (N=22) and 19% had experienced all four assessment types (N=24). The other 12% of the respondents had varying combinations and permutations of experience with the four methods of assessment.

A sample of six students at each of the three levels of language proficiency was interviewed. For the interviews, two students at each level who demonstrated negative attitudes towards performance-based assessment and two students at each level who gave responses on the questionnaire that indicated positive responses to the three forms of

performance-based assessment were selected for interviews. The other two interviews conducted at each level were with students who seemed to demonstrate a neutral or ambivalent attitude towards performance-based assessment. On the letter of informed consent, the students were told that the reason they were asked to write their names on the questionnaire was so that they could be identified for possible interviews based on their responses and their willingness to be interviewed. An additional consideration in selecting the participants for the interviews was a desire to get a range of ages, country backgrounds, and a mix of males and females representative of the percentages in the population of the study.

Of the students who participated in the interviews, five were from Japan, three were from Korea, three were from Taiwan, two were from Colombia, and four were from other countries including Iran, Germany, Bangladesh, Switzerland, and one student was from Quebec. A total of five men and thirteen women were interviewed and all of the age range categories were represented.

#### Teacher participants.

A total of thirteen teachers filled out the teacher questionnaires. All of the teachers had over nine years of teaching experience and all were currently using at least three of the assessment types under study in their classes.

A sample of six teachers was selected for interviews based on their willingness to participate and their responses to the teacher questionnaire. A sample of two teachers who taught at each of the three levels of proficiency was interviewed. The intention had been to choose teachers such that one teacher at each level would have given responses

on the teacher questionnaire indicating that they perceived students to be positive towards performance-based assessment and one teacher would have answered the items in a way suggesting that they perceived students to have negative or ambivalent attitudes towards the three performance assessment types under study. However, no such dichotomy was evident from the questionnaires so teachers were selected mainly based on their willingness and availability to be interviewed. In addition, the head teacher of assessment was interviewed in order to get a sense of the overall assessment process at the institution.

### Instruments Used

Student participants filled out a questionnaire of 10 Likert-type scale type items in which they indicated their attitudes towards four different methods of classroom-based assessment: tests, portfolios, presentations, and participation (see Appendix A). In effect, since the students were answering for each of the four types of assessment, they responded to 40 items in total. Students were also asked to fill in some background information about themselves (gender, home country, first language, the length of time spent in Canada, their educational background, and their previous experience getting marked on each of the four assessment types of interest in the study) which, according to previous research, may have influenced attitudes to the constructs defined in the instrument. Finally, participants were asked to write what they either liked or disliked about getting marked on each of the four types of assessment. A space for comments was also included at the end of the questionnaire.



Teacher participants filled out a questionnaire of 10 Likert-type scale type items in which they indicated their perceptions of student attitudes towards the four types of assessment: tests, portfolios, presentations and participation (see Appendix B).

Background information on the teachers was also collected including such factors as length of time teaching and experience using traditional tests, portfolios, presentations, and participation for summative assessment purposes.

The interviews of both the teachers and students followed a semi-structured interview format (see Appendices C and D). A visual was provided to aid those being interviewed, as well as the interviewer, in remembering the four types of assessment.

#### Procedures followed

Permission to conduct the study at the institution was granted. The Coordinator of the IEP first gave permission and the head teacher for assessment also gave consent. The head teacher had been interviewed in the pre-pilot study about the types of assessment that had been used since the implementation of formal assessment, which had taken place a year previously. After that initial interview, it was decided to limit the subject of the study to the four types of assessment. All four seemed to have been used in one of the time slots of the classes.

Each of the potential classes in the study received a package containing an instruction sheet for the teachers, consent letters for the students to sign, consent letters for the teachers to sign, the student questionnaires, and one teacher questionnaire. The instructors were asked to follow the procedures on the instruction sheet for the student questionnaires so that some consistency in administration of the questionnaire could be

achieved (see Appendix E). The students agreeing to participate in the study signed the consent letters (see Appendix F), which were put in an envelope and sealed. The teachers then handed out the student questionnaires which, when completed, were collected by a student, sealed in an envelope, and taken to the office by a volunteer in the class. This ensured the students' anonymity and that their teachers did not see their responses to the questionnaires. On the outside of each of the questionnaire envelopes, the class level (not the class name) was written.

After the questionnaires had been collected, they were put into groups according to the three levels of language proficiency and mixed so that the individual classes were not distinguishable. The questionnaires were then examined and participants were selected for interviewing based on their responses to the questionnaires. Students received letters asking if they would be willing to be interviewed and if so, with a possible time suggested (see Appendix G); the responses were sent back via the teachers. At the beginning of the interview, student participants were also asked to sign a letter of informed consent (see Appendix H).

Each teacher was given an envelope containing a teacher questionnaire, a letter of informed consent (see Appendix F) and an envelope in which to put the teacher questionnaire. After completing the questionnaire, each teacher was asked to put it in the envelope along with the letter of consent, seal the envelope, and return it to the office. For the interviews, teachers were approached and asked if they would be willing to participate in an interview (see Appendix G). Prior to the interviews, teachers also signed letters of informed consent (see Appendix H).

### Data Collection and Analysis

In this study, both quantitative and qualitative data were collected. The Likert-type scale items were scored and negative items were reverse scored. Therefore, the higher the mean for each item, the higher the positive attitude towards the construct of that assessment type. For the quantitative data, the computer program SPSS was used to conduct statistical analyses. An item analysis of the questions on the instrument was conducted, along with a calculation of coefficient alpha for internal consistency reliability. Descriptive statistics for each item and subconstruct were calculated. To test for significance between variables such as class level, home country, gender, age, length of time in Canada, educational background, and previous experience with the four assessment types, tests using GLM multivariate procedures as well as t-tests and simple effects tests were conducted. GLM repeated measures tests were also used to check for any interaction effects. In addition to correlating responses with the biographic variables, an exploratory factor analysis was conducted on the Likert-type scale items to see if the items on performance-based assessment emerged as one single factor or three separate factors as well as to explore whether the items about traditional tests appeared as one or possibly more than one factor. In addition, a secondary exploratory factor analysis was conducted on each of the subconstructs identified in the initial factor analysis. For the teacher questionnaires, descriptive statistics for each item were calculated so a comparison could be made with the item means from the student questionnaires. Due to the small sample size of the teachers (N=13) further analyses of the quantitative data were not conducted.

As a check to see how many participants were in different proficiency levels for different skill areas, the class lists for the other two time slots (not the one in which the study was conducted) were examined and a tally was made.

The qualitative data from the questionnaires were first coded as to whether they demonstrated a positive, negative, or neutral attitude towards the assessment type. Each comment could have received more than one code in that a participant may have mentioned both positive and negative attributes of an assessment type within the same comment. To get a sense of the number and affective nature of these comments, a tally was made. The qualitative data were then analyzed for common themes and patterns as to the content of the comment using the categories in Table 3.1.

General comments such as “I like it” or “Dislike” were not coded for content, as the affective nature of such comments would have been captured in the initial coding into the positive and negative categories. However, specific affective responses to the types of assessment were included in the coding. Once common themes had been extracted from the data, the scheme to code the data according to content was revised to facilitate analysis.

Some of the categories were used with positive or negative superscripts so that the exact nature of the content could be described. When a tally was made of the number of comments for each category, some of the categories had to be subdivided into positive and negative categories. For example, the category “Demonstration of Ability” had to be divided to account for the negative comments so that a new category “Not Reflective of Ability” was added. Table 3.1 lists the categories used for the refined coding scheme.

Table 3.1

**Categories Used to Code the Qualitative Data from the Questionnaires**

---

1. Affective Responses
    - Positive
    - Negative
  2. Assessment Format or Type
  3. Assessment Involving a Process
  4. Assessment Resulting in a Product
  5. Concerns about Marking
  6. Concerns about Time Involvement
  7. Demonstration of Ability
    - Demonstrates Ability
    - Not Reflective of Ability
  8. Difficulty of Assessment Method
    - Easy
    - Difficult
  9. Equity Issues
    - Fair
    - Not Fair
  10. Familiarity
    - Familiar
    - Not familiar
  11. Feedback to Students
  12. Frequency
    - Too frequent
    - Not frequent enough
  13. Motivation
  14. Necessity
  15. Opportunity to Learn
    - Opportunity to Learn
    - Not an Opportunity to Learn
- 

When the comments were coded as to their content, each comment could have received multiple codes but no more than one code from each category. To check the reliability of the coding scheme, a second coder rated a random sample of 10 comments on each of the four types of assessment. When the total number of matching codes was divided by the total number of codes, a reliability of 85% was achieved.

Both the student and teacher interviews were taped and the transcriptions were analyzed for common themes and patterns. In addition, some of the student responses in the interviews were checked against what they had written on the questionnaires as a check on the reliability of their responses. The interview question regarding what participants like and/or dislike about getting marked on each type of assessment was coded using the categories in Table 3.1.

In both the quantitative and qualitative results sections, the comments made by the participants remain in their original form. That is to say, no corrections in terms of grammar or spelling were made so as to preserve the exact quality of the comments.

## CHAPTER 4

## Quantitative Results

Results of the Factor Analysis of the Scale

In an attempt to reduce the data, principle axis factor analysis (Varimax rotation) was conducted on the 40 Likert-type scale items. Preliminary analysis of the scree plots indicated that there were 5 factors with eigenvalues greater than 2. However, when one item was deleted from the factor analysis, and four factors were extracted, the four types of assessment loaded onto factors corresponding to each assessment type. That is, all of the items about tests loaded onto factor 1; similarly all the items regarding portfolios, participation, and presentations loaded onto factors 2, 3, and 4, respectively. These results suggest that the questionnaire consisted of four separate subconstructs. The results of the factor analysis are shown in Table 4.1.

Table 4.1

Factor Loadings

	Factor 1	Factor 2	Factor 3	Factor 4
Test1	.728	Port1 .702	Part1 .477	Pres1 .590
Test2	.656	Port2 .669	Part2 .661	Pres2 .717
Test3	.802	Port3 .742	Part3 .711	Pres3 .667
Test4	.730	Port4 .769	Part4 .763	Pres4 .749
Test5	.445	Port5 .594	Part5 .540	Pres5 .347
Test6	.827	Port6 .670	Part6 .774	Pres6 .532
Test8	.685	Port8 .469	Part8 .684	Pres8 .695
Test9	.675	Port9 .730	Part9 .699	Pres9 .673
Test10	.610	Port10 .556	Part10 .587	Pres10 .395
% variance	19.3%	13.1%	8.9%	7.6%

Note: Test1 refers to the test subitem of item 1. Similarly, Port1, Part1 and Pres1 refer to the portfolio, participation, and presentation subitems of item 1, respectively.

The one item that did not fit into the factor analysis (item 7 – “Doing \_\_\_\_\_ takes too much time” – see Appendix A) was therefore not included in the subsequent statistical analyses of the four subconstructs but rather was analyzed separately.

### Subscale Reliability

To check the reliability of each subscale corresponding to the four assessment types under study, the Cronbach alpha coefficient was calculated. With the deletion of item 7, the coefficient for the test subscale was .89; the coefficient for the portfolio subscale was .88; and the coefficients for the presentation and participation subscales were .85 and .88, respectively. The coefficient alphas obtained with item 7 remaining were all lower with values of .85, .84, .80, and .85 for the test, portfolio, presentation, and participation subscales, respectively. All of the coefficient alphas show that the scale had acceptably high internal consistency, given that the value of alpha should be at least .70 for a scale to demonstrate internal consistency (Spector, 1992, p.32).

### General Response to the Four Assessment Types

#### Student attitudes

As a check on the overall attitudes to the four subconstructs, the mean score for each of the four types of assessment was calculated. The most positive attitude, “Strongly Agree” was scored as a 5 and the most negative attitude, “Strongly Disagree” was scored as a 1. “Agree” and “Disagree” were scored as 4 and 2, respectively. “Don’t know” was scored with a value of 3. Therefore, a mean score above 3 indicates a positive attitude toward the assessment type. The means of the items for presentations, portfolios, tests and participation were 4.12, 3.99, 3.92, and 3.83, respectively (see Table



4.2). All are on the positive end of the scoring scale, suggesting a favourable response to all four types of assessment. Paired t-tests were conducted to check for differences between these means. The significance level was adjusted using the Bonferroni procedure of dividing alpha (.05) by the number of comparisons being made (Norusis, 1998). With this correction for multiple comparisons, the mean for presentations was found to be significantly higher ( $p < .01$ ) than the means for tests and participation.

Table 4.2

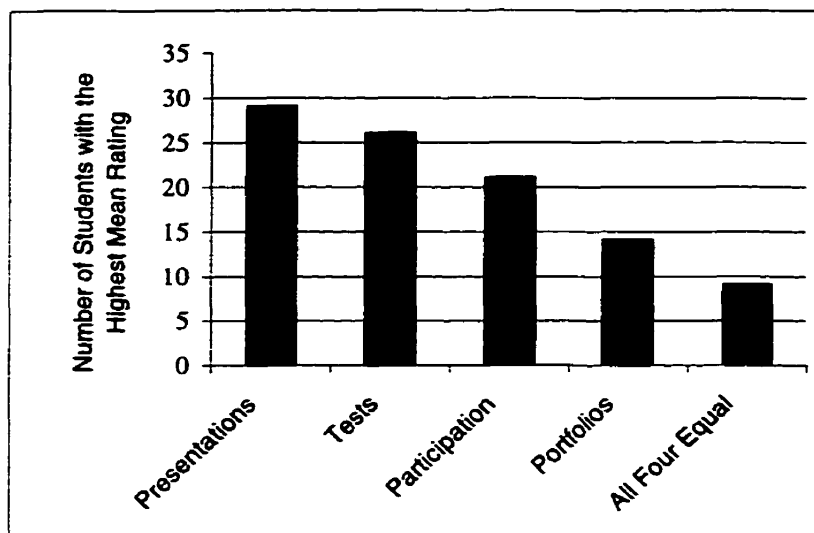
Descriptive Statistics for the Four Assessment Types (N=127)

Assessment Type	Mean	Standard Deviation	Median
Presentations	4.12*	.56	4.2
Portfolios	3.99	.56	4.0
Tests	3.92	.69	4.0
Participation	3.83	.76	4.0

\* at  $p < .01$  significantly higher than the means for tests and participation

Because looking at the means in this way could mask individual participant ratings for each type of assessment, a tally was made of which form of assessment had the highest mean score for each participant. Results of this tally (see Figure 4.1) indicate that presentations were rated most highly by the largest number of participants (N=29), with tests being preferred by 26 of the participants and participation and portfolios rated most highly by 21 and 14 participants, respectively. The remaining participants either rated all four assessment types equally (N=9) or rated some other combination of assessment types equally (N=28). This, in part, could account for the discrepancy between the results in Table 4.2 and those in Figure 4.1. In addition, the slight change in order of preferences could be due to the fact that the latter results were obtained by

ranking the participants' mean scores for each assessment type and therefore, a slight difference in mean scores would be reflected in the tally.



**Figure 4.1.** The results of a tally of frequencies of each participant's highest mean rating of the four assessment types.

Overall, 3% of the participants' responses were "Don't know".<sup>2</sup> The range was from 0% (for test item 2) to 7.1% (for participation item 7) or 6.3% (for participation item 6) if item 7 is not included in the analysis. The items about tests resulted in the lowest frequency of this response (1.4%) followed by the items about presentations (2.9%), portfolios (3%), and participation (4.5%). Given that all of the participants had had prior experience with tests, it is not surprising that the items about tests resulted in a low frequency of "Don't know" as a response.

<sup>2</sup> It should be noted that the "Don't know" response could have been chosen by those with a neutral attitude. In the scale that was used, this distinction was not made.

### Teachers' Perceptions of Student Attitudes

The results from the questionnaires of teachers' perceptions of student attitudes to the four assessment types were analyzed by scoring the items and getting mean scores for tests, portfolios, presentations, and participation. Through conducting paired t-tests with a Bonferroni correction for multiple comparisons, the mean of the teachers' ratings of student attitudes towards tests was found to be significantly higher than the means for portfolios and participation. All of the descriptive statistics for teachers' perceptions of student attitudes to the four assessment types are shown in Table 4.3. Even though due to the small sample size of 13 no factor analysis was carried out, item 7 was removed from calculating the descriptive statistics just as it was for the student data.

Table 4.3

#### Descriptive Statistics for the Teachers' Perceptions (N=13) of Student Attitudes to the Four Assessment Types

<u>Assessment Type</u>	<u>Mean</u>	<u>Standard Deviation</u>	<u>Median</u>
Tests	4.32*	.44	4.33
Presentations	3.93	.60	3.93
Portfolios	3.86	.55	3.86
Participation	3.46	.69	3.46

\* at  $p < .01$  significantly higher than the means for portfolios and participation

For presentations and portfolios, the means of the teacher ratings of student attitudes matched those of their students quite closely but for tests and participation, the means differed considerably. The mean score for portfolios for students was 3.98 while the teachers' mean score was 3.86. For presentations, the students had a mean score of 4.12 whereas the teachers had a mean score of 3.93. Tests received mean scores of 3.92

and 4.32 for students and teachers, respectively. That is, teachers felt that students would rate tests more positively than they actually did. Conversely, teachers thought students would rate participation more negatively than the findings indicate. For participation the students' mean score was 3.83 while the teachers' mean score was only 3.46.

When the rank ordering of the four assessment types was compared, the teachers' perceptions of student attitudes also differed from the students' attitudes (see Table 4.4). Although the teacher sample size of 13 is considerably smaller than that of the student sample size of 127, it is interesting to note the differences, with the exception of that for participation, in the ranking of the four assessment types. The teachers possibly assumed that since students typically have more experience with tests than with the other three assessment types, the students would prefer tests. However, the data suggest that familiarity with the assessment type was not necessarily related to attitude.

Table 4.4

Comparison of the Students' and the Teachers' Ranking of the Four Assessment Types based on the Mean Scores

Assessment Type	Student Ranking	Teacher Ranking
Presentations	1	2
Portfolios	2	3
Tests	3	1
Participation	4	4

## Background Variable Results

### Testing the Assumptions of Normality and Sphericity

In SPSS, General Linear Model (GLM) multivariate and repeated measures procedures on the means of the subconstructs of each assessment type were used to look for any background correlates. However, first the assumption of a normal distribution of the data had to be verified. The data indicate that the participants had a favourable attitude to all four types of assessment; therefore, the data are negatively skewed with greater medians than means. The data, however, do just fit the test of normality (dividing the kurtosis by the standard error) so multivariate tests, which are based on an assumption that the data are normally distributed, were conducted. The kurtosis values and their standard errors are found in Table 4.5. The ratio of kurtosis to its standard error should fall within the range of  $-2$  to  $2$  for the assumption of normality to be satisfied (SPSS, 1997).

Table 4.5

#### Tests of Normality for each of the Four Subconstructs

Subconstruct	Kurtosis	Standard Error	Ratio
Tests	.618	.427	1.45
Portfolios	.610	.427	1.43
Presentations	.630	.427	1.48
Participation	.820	.427	1.92

The dependent variables in each analysis were the means of the students' attitudes towards tests, portfolios, presentations, and participation. For repeated measures analysis, in addition to the assumption of normality, the assumption of the sphericity of the variance-covariance matrix of the dependent variables needed to be satisfied in order

to use univariate results of the within-subjects factors and the interaction of the within-subjects factors with the between-subjects factors. If the variance-covariance matrix is circular in form, the F statistic is considered to be valid for the univariate analysis (SPSS, 1997). Therefore, to check this assumption, Mauchly's Test of Sphericity was conducted. Because the assumption of sphericity could not be satisfied for these data, the Greenhouse-Geisser epsilon adjustment, which is a conservative correction for the numerator and denominator degrees of freedom, was used to assure the validity of the F statistic. When required, this adjustment is reflected in all of the reported F statistics in this chapter.

For all of the GLM multivariate and repeated measures statistical analyses in examining background variables and attitudes to the four types of assessment, alpha was set at .05. To protect against Type I errors in multiple comparisons, Bonferroni corrections were made. The background variables of age, gender, time in Canada, reason for studying, and prior experience with the four assessment types revealed no significant differences. The results of the other background variables will be discussed below.

#### Background Variable of Level of Language Proficiency

For the independent variable of level of language proficiency, significant differences ( $F=6.24$ ;  $df= 2.5, 304$ ;  $p=.001$ ) were obtained in the overall means for the four types of assessment. Pairwise comparisons with a Bonferroni adjustment for multiple comparisons suggest that the mean of students' attitudes towards presentations was significantly higher than the attitudes held towards tests and participation (as was the

finding from the results in Table 4.2). The descriptive statistics for the variable of level of language proficiency can be found in Table 4.6.

Table 4.6

Descriptive Statistics for the Variable Level of Language Proficiency

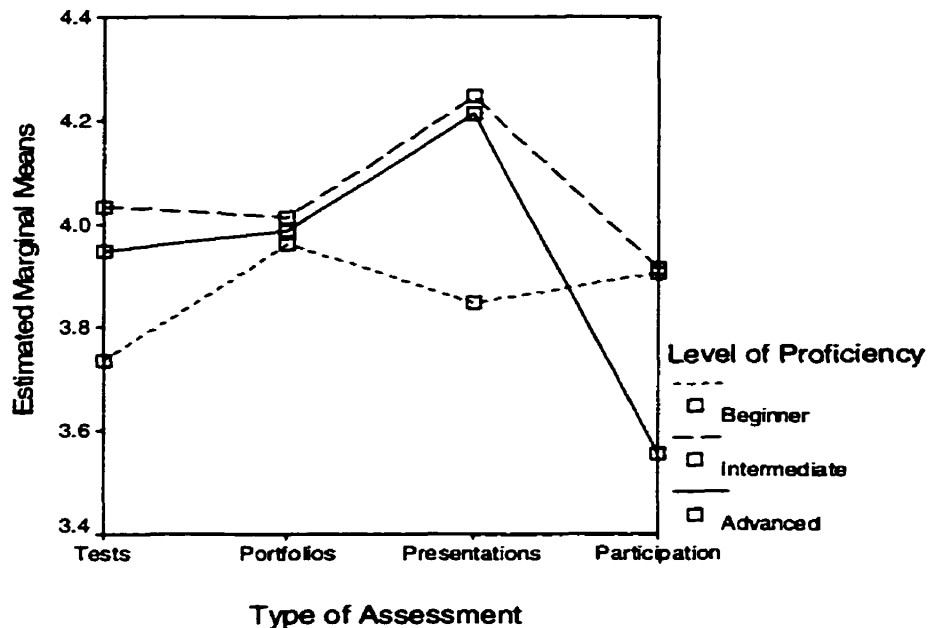
Type of Assessment	Level of Language Proficiency							
	Beginner (N=38)		Intermediate (N=59)		Advanced (N=30)		Overall (N=127)	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Tests	3.74	.72	4.03	.67	3.95	.64	3.92	.69
Portfolios	3.96	.48	4.01	.58	3.99	.62	3.99	.56
Presentations	3.85	.63	4.25 <sup>b</sup>	.45	4.21 <sup>b</sup>	.55	4.12 <sup>a</sup>	.56
Participation	3.91	.62	3.91	.72	3.56	.94	3.83	.76
Overall	3.86	.40	4.05	.38	3.93	.43	3.97	.41

<sup>a</sup>sig. ( $p < .05$ ) higher than the means of the three other assessment types

<sup>b</sup>sig. ( $p < .05$ ) higher than the mean for beginner students

Analysis of the overall means for level of language proficiency did not yield significant results ( $F=2.73$ ;  $df=2, 124$ ;  $p=.069$ ). To check for any interaction effects, a univariate GLM repeated measures analysis was conducted with the result that an interaction ( $F=2.83$ ;  $df=4.9, 304$ ;  $p=.02$ ) was found between assessment type and level of language proficiency in these data (see Figure 4.2).

A cursory glance of the interaction plot suggests that beginner students appear to have held less positive attitudes to tests, portfolios, and presentations than did intermediate and advanced students. The results of the interaction plot also suggest that attitude to an assessment type depends on the level of language proficiency of the students in that the plot lines cross between presentations and participation. To determine the source of the interaction, the between-subjects simple main effects and multiple comparisons were examined.



**Figure 4.2.** Interaction plot of the estimated marginal means for level of language proficiency and the four types of assessment.

Repeated contrasts of the interaction indicate a significant difference between presentations and participation ( $F=6.98$ ;  $df=2, 124$ ;  $p=.001$ ). A simple effects test indicates that for intermediate and advanced students, there was a significant difference in their attitudes towards presentations and participation. The findings suggest that intermediate students held significantly more positive attitudes towards presentations than they did towards participation ( $F=10.4$ ;  $df=1, 124$ ;  $p=.002$ ). Similarly, advanced students rated presentations more positively than they did participation ( $F=20.6$ ;  $df=1, 124$ ;  $p<.0001$ ). Post hoc Bonferroni tests indicate that both intermediate and advanced students had significantly more positive attitudes towards presentations than did beginner students (see Table 4.6). For beginner students, no significant differences in attitudes between the four types of assessment were found.



### The Language Background Variable

When the background variable of language was explored, only four language groups were included since the other language groups had only three or fewer participants. Examining language groups rather than home country had the effect of combining those participants from China and Taiwan into one language group (Chinese) and combining participants from Colombia, Mexico, Panama, and Peru into one language group (Spanish). In the GLM repeated measures procedure, a significance difference was found in the overall means for type of assessment ( $F=3.74$ ;  $df=2.4, 257$ ;  $p=.019$ ) but not in the overall means for language background ( $F=1.12$ ;  $df=3, 108$ ;  $p=.343$ ), or in the interaction between language background and type of assessment ( $F=1.28$ ;  $df=7.1, 257$ ;  $p=.260$ ). The univariate results of the GLM multivariate procedure, showed a significant difference ( $p<.05$ ) for mean scores on attitudes towards tests with respect to the language background variable (see Table 4.7).

Table 4.7

#### Results of the Tests of Between-Subjects Effects for the Variable Language Background (N=112)

<u>Assessment Type</u>	<u>df</u>	<u>F</u>	<u>Sig.</u>
Tests	3, 108	2.78	.045*
Portfolios	3, 108	1.13	.340
Presentations	3, 108	.251	.860
Participation	3, 108	.522	.668

sig. at  $p = .05$

However, when multiple comparisons were made between language groups, no significant differences in means were detected. In this sample, those participants from a Korean ( $N=14$ ) language background had a mean score of 4.19 for tests, Chinese ( $N=37$ )

speaking students had a mean score of 4.10, Japanese speaking participants (N=36) had a mean score of 3.84, and Spanish speakers (N = 24) had a mean score of 3.73. The descriptive statistics for this variable are found in Table 4.8.

Table 4.8

Descriptive Statistics for the Variable Language Background

Assessment Type	Language Background	Mean	Standard Deviation	N
Tests	Chinese	4.10	.46	37
	Japanese	3.84	.58	36
	Korean	4.19	.43	14
	Spanish	3.73	.99	25
Portfolios	Chinese	4.10	.46	37
	Japanese	3.88	.57	36
	Korean	4.05	.56	14
	Spanish	4.00	.60	25
Presentations	Chinese	4.11	.50	37
	Japanese	4.14	.53	36
	Korean	4.02	.61	14
	Spanish	4.17	.57	25
Participation	Chinese	3.84	.69	37
	Japanese	3.69	.76	36
	Korean	3.94	.47	14
	Spanish	3.89	.97	25

The Background Variable of Home Country

Although examining the variable home country decreases the sample sizes, tests were run to see if any significant differences in means could be detected for the countries Colombia, Japan, Korea, Mexico, and Taiwan. All of the other countries represented in these data had four or fewer participants so they were not included in the statistical analysis for the variable of home country. Significant differences were found in the overall means for type of assessment but not in the overall means for home country;

however, an interaction effect was significant. The results of these GLM repeated measures tests are found in Table 4.9.

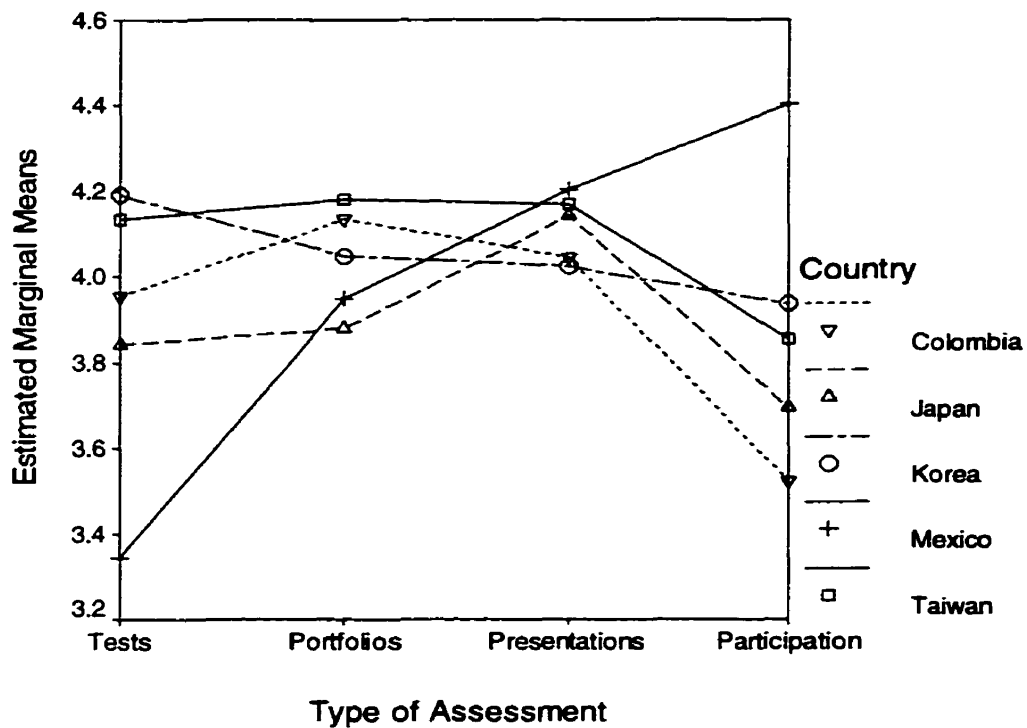
Table 4.9

**Results of a Repeated Measures ANOVA on Student Preference Ratings for Assessment Type and the Independent Variable of Home Country (N=104)**

	F value	df	Sig.
Overall for Assessment Type	3.47	2.4, 239	.025
Overall for Home Country	1.19	4, 99	.318
Assessment Type * Home Country	3.10	9.7, 239	.001

The interaction plot suggests that considerable differences in attitude towards tests and participation existed among the participants from the countries represented with students from Mexico appearing to hold considerably different views about tests compared with students from Korea. Similarly, from the interaction plot, it would seem that students from Mexico also had quite divergent views about participation when compared with those held by students from Colombia (see Figure 4.3).

Simple effects tests show that when the attitudes towards portfolios of students from Colombia were compared with their attitudes towards presentations and participation, a significant difference was found ( $F=4.25$ ;  $df=1, 99$ ;  $p=.042$ ) and similarly, a difference in the mean attitude scores for Colombian students was found with regard to their attitudes towards presentations and participation ( $F=3.93$ ;  $df=1, 99$ ;  $p=.05$ ). Colombian students ( $N=10$ ) seemed to rate portfolios the most highly, with a mean score of 4.13, while the mean scores for presentations and participation were 4.04 and 3.52, respectively.



**Figure 4.3.** Interaction plot of the estimated marginal means for attitude towards assessment type for home country.

For Japanese students ( $N=36$ ), a difference was also found in their attitudes towards presentations and participation ( $F=10.4$ ;  $df=1, 99$ ;  $p=.002$ ) with mean scores of 4.14 and 3.69, respectively. Mexican students had a significant difference in attitudes when tests were compared to the other three types of assessment ( $F=17.5$ ;  $df=1, 99$ ;  $p<.0001$ ) and when portfolios were compared to presentations and participation ( $F=4.77$ ;  $df=1, 99$ ;  $p=.031$ ). From the interaction plot, it is clear that for Mexican students ( $N=11$ ), participation was the most preferred form of assessment, followed by presentations, portfolios and tests with means of 4.40, 4.20, 3.95, and 3.34, respectively. A simple effects test indicates that for students from Taiwan ( $N=33$ ), there was a significant difference in their attitudes towards presentations and participation ( $F=4.66$ ;  $df=1, 99$ ;

$p=.033$ ), with the former preferred to the latter. Korean students ( $N=14$ ) showed no significant differences in their attitudes towards the four types of assessment. The descriptive statistics for the variable home country are found in Table 4.10.

Table 4.10

Descriptive Statistics for the Variable Home Country (N=104)

	Colombia N=10		Japan N=36		Korea N=14		Mexico N=11		Taiwan N=33	
	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.	mean	s.d.
Tests	3.96	.29	3.84	.58	4.19	.43	3.34	1.16	4.13	.48
Portfolios	4.13	.62	3.88	.57	4.05	.56	3.95	.63	4.18	.41
Presentations	4.04	.71	4.14	.53	4.02	.61	4.20	.45	4.17	.47
Participation	3.52	1.23	3.69	.76	3.94	.47	4.40	.23	3.86	.75

Tamhane post hoc multiple comparisons of the between-subjects effects were not significant for tests, portfolios, or presentations. However, for participation, significant differences between students from Mexico and students from Japan ( $p<.0001$ ), Korea ( $p<.04$ ) and Taiwan ( $p<.003$ ) were observed. As has already been mentioned, a glance at the interaction plot suggests that the Mexican participants held considerably different attitudes towards participation than did participants from Colombia but in the post hoc tests, the differences were found to be not significant (perhaps due to the small sample size and large standard error of .320 for that comparison).

The Level of Education Variable

Although no interaction effects were found in the repeated measures analysis, tests of between-subjects effects in the GLM multivariate procedure revealed a significant difference ( $p < .05$ ) between the mean scores for attitudes towards tests held by university students and high school graduates versus the attitudes of those students

who were either in or had finished graduate school. University (N=92) and high school students (N=20) had mean scores of 3.99 and 3.91, respectively, for tests whereas graduate students (N=10) had a mean score of 3.39 for tests. Because of the unequal sample sizes for these three groups of participants, any conclusions suggested might be spurious but it is still interesting to note the differences (see Table 4.11).

Table 4.11

Results of the Tests of Between-Subjects Effects for the Variable Level of Education (N=127)

Assessment Type	df	F	Sig.
Tests	2, 119	3.4	.036*
Portfolios	2, 119	.34	.709
Presentations	2, 119	.31	.736
Participation	2, 119	.58	.561

sig. at p = .05

Results of the Time Item

Because item 7 (Doing \_\_\_\_\_ takes too much time – see Appendix A) did not load neatly onto the factor analysis, it was analyzed separately. However, because this is a one-item response and therefore, not highly reliable, the analysis was conducted for exploratory purposes rather than for drawing any conclusions. The item may have been problematic because it is the only one that is negatively worded. When reversed scored, the mean values for this item for tests, portfolios, presentations, and participation were 3.24, 2.78, 2.85, and 3.55, respectively. This could be interpreted as the participants' finding participation and tests the least time consuming and portfolio and presentations requiring more time. This is the type of response that one might intuit but perhaps the item did not load onto the factor analysis because it was misunderstood by some of the

participants. Therefore, the means for each of assessment types for each level of language proficiency were calculated. The mean values for tests and participation were highest for advanced students and lowest for beginner students, indicating that perhaps the beginner students may not have fully understood the item. However, the standard deviations for the three groups are not markedly different so level of language proficiency may not have had any effect on how students responded. The means and standard deviations of item 7 for the three language proficiency groups, as well as the overall means for all the students together, are shown in Table 4.12.

Table 4.12

Descriptive Statistics for Item 7 for the Three Levels of Language Proficiency

Assessment Type	Beginner		Intermediate		Advanced		Overall Total	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Tests	3.03	1.31	3.25	1.24	3.50	1.14	3.24	1.23
Portfolios	2.89	1.11	2.80	1.26	2.60	1.22	2.78	1.20
Presentations	2.82	1.23	2.90	1.30	2.80	1.27	2.85	1.26
Participation	3.00	1.38	3.78	1.05	3.80	1.03	3.55	1.20

Results of the Secondary Factor Analysis

To explore whether any identifiable factors were present within each of the four subconstructs, including item 7, a secondary principle axis factor analysis (Varimax rotation) was conducted. Initially, extracting eigenvalues of greater than 1 resulted in two factors for each of the four subconstructs. Removing item 7 resulted in the extraction of 1 factor for tests, portfolios, and presentations and 2 factors for participation. Therefore, in the secondary factor analysis, eigenvalues greater than .5 were extracted so that a greater number of factors could be explored.

Although the factor analysis did not result in clean factor loadings, some interesting patterns did emerge (see Table 4.13). It should be noted that in Table 4.13 the order of the factors has been changed so that the factors are not presented in the order of variance explained by each factor. The factors and the items under each factor are organized to demonstrate loading patterns of the factor analyses conducted for the four assessment types.

Table 4.13

Rotated Factor Loadings in the Secondary Factor Analysis Conducted on each

Subconstruct

Subconstruct	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
Tests	1 (.467)		5 (.319)	10 (.659)	9 (.675)	7 (.565)
	2 (.741)		3 (.538)			
	4 (.684)		6 (.498)			
			8 (.696)			
Portfolios	2 (.785)	1 (.582)	6 (.723)	5 (.598)	3 (.587)	7 (.384)
		4 (.544)	8 (.607)	10 (.543)		
		9 (.882)				
Presentations	1 (.770)	4 (.550)	6 (.594)	5 (.521)	3 (.645)	7 (.401)
	2 (.537)	9 (.688)	8 (.493)			
			10 (.550)			
Participation	1 (.646)	3 (.583)	5 (.655)			7 (.528)
	2 (.698)	4 (.709)	6 (.750)			
		9 (.542)	8 (.543)			
			10 (.600)			

Note: Each value on the left refers to the item number (see Appendix A) and each value on the right is the loading. The factors are not presented in the order of variance explained by each factor.



For all four of the subconstructs, tests, portfolios, presentations, and participation, item 6 and item 8 loaded onto the same factor, which could be designated as “Opportunity to Improve or Show Ability”. However, for presentations and participation, item 10, which taps into the in-class motivational response to the assessment, also loaded onto the same factor. Opportunity to improve, opportunity to show ability, and motivation are three themes that are prevalent in the qualitative results. Items 1 and 2 loaded onto the same factor for tests, presentations, and portfolios. This factor could be designated as “Affective Response” since both items tap into preference and general attitude to the assessment type. For all three forms of performance-based assessment, items 4 and 9 loaded onto the same factor. Since the former item taps into fairness and the latter has to do with whether or not the assessment type should be used for marks, this factor could be “Equity and Marking”. Item 7, the one that did not load cleanly onto the initial factor analysis of the scale, emerged as a single factor for each of the four subconstructs. Only this item loaded onto the factor so it could be designated as “Time Involvement”. The factor loadings of the other items did not result in any clear patterns. Although this was just an exploratory analysis, many of the themes reflect those in the qualitative data in the following chapter.

### Summary of the Quantitative Results

The findings of this part of the data analysis seem to indicate that the participants generally held positive attitudes towards all four types of assessment: tests, portfolios, presentations, and participation. Based on the results of the factor analysis and the calculation of the Cronbach alpha coefficient for each of the subscales, the participants seemed to be rating four different assessment subconstructs. Presentations and portfolios

had the highest mean ratings followed by tests and participation, respectively. A tally of individual student preferences for assessment type also indicated that presentations ranked the most highly. However, the order of preference for the three other forms of assessment was different, with tests rated second, participation third, and portfolios last. The means of the teacher ratings of student attitudes for portfolios and presentations matched those of their students quite closely but for tests and participation, there was a mismatch between teachers' perceptions and actual student attitudes. The rank ordering also differed considerably in that teachers indicated that students' attitudes towards tests would be the most favourable, followed by presentations, portfolios, and participation, respectively.

The background variables of age, gender, time in Canada, reason for studying and prior experience with the four assessment types revealed no significant differences. For the background biographic variable of level of language proficiency, an interaction between level of proficiency and attitudes towards the four types of assessment was found. Both intermediate and advanced students indicated a preference for presentations over participation and the between-subjects simple main effects were significant in that for presentations, advanced and intermediate students had higher mean scores than did beginner students. The independent variable of language background was significant for tests but when multiple comparisons between language groups were made, no significant differences in means were detected. Multiple comparisons of the within-subjects simple effects showed no significant differences in attitude between assessment types. For the home country variable, a significant interaction effect was present although the between-subjects main effects of country and the within-subjects effects of assessment type were

not significant. Results suggest that Colombian, Japanese, and Taiwanese participants showed a more positive attitude towards presentations than participation. Additionally, students from Colombia and Taiwan preferred portfolios to participation. Japanese students seemed to prefer presentations to participation while Mexican students rated participation more highly than they rated tests. The between-subjects effects for the variable level of education showed a significant difference in attitudes for the dependent variable tests. Results suggest that those with a graduate level of education did not rate tests as highly as did students with university and high school levels of education.

Because it did not load cleanly in the factor analysis, item 7 regarding the time requirement for the assessment type, was analyzed separately. Results indicate that participants found presentations and portfolios the most time-consuming. It may be that because the item was negatively worded, participants with lower levels of language proficiency may have misunderstood it although the results are what one might have predicted.

The secondary factor analysis of each of the four subconstructs of tests, portfolios, presentations and participation did not result in clean factor loadings but some interesting patterns were observed. A potential factor that emerged was “Opportunity to Improve or Show Ability or Provide Motivation”. Two other factors that seemed to be present were “Affective Responses” and “Equity and Marking”. Finally, the time item (item 7) emerged as a single Time Involvement factor in all four secondary factor analyses. While not in any way conclusive, the tentative designation of the factors from the

exploratory factor analysis are suggestive of the themes in the qualitative data that will be presented in the next chapter.

## CHAPTER 5

### Qualitative Results

#### The Qualitative Data from the Questionnaires

The qualitative data from the questionnaires were compiled and analyzed for common themes and patterns. To get a sense of the attitudinal nature of the comments, they were classified as being neutral, positive, or negative. Some participants had comments that had both positive and negative components so these were counted separately and classified accordingly. The 263 positive comments far outnumbered the 143 negative comments (see Table 5.1). Tests and presentations both elicited the greatest number of positive comments, which is in keeping with the finding from the tally of student preferences from the quantitative data as indicated in Figure 4.1. Tests also resulted in more negative comments than the three performance-based assessments with a total of 45 being negative as opposed to the totals of 29, 32, and 37 for portfolios, presentations, and participation, respectively.

Table 5.1

#### Tally of the Type and Number of Comments from the Qualitative Data

Type of Assessment	Negative Comments		Positive Comments		Neutral Comments		Total	
	N	%	N	%	N	%	N	%
Tests	45	31.5	72	27.4	6	17.1	123	27.9
Portfolios	29	20.3	60	22.8	13	37.1	102	23.1
Presentations	32	22.3	70	26.6	9	25.7	111	25.2
Participation	37	25.9	61	23.2	7	20.0	105	23.8
Total	143	100	263	100	35	100	441	100

In addition to classifying the comments as to whether they were positive or negative, the comments were also coded for content or nature of the comment to see if any patterns emerged. The coding scheme in Table 3.1 was used.

Coding the qualitative data in this way and tallying each of the codes resulted in some patterns emerging from the data. The three performance-based assessment methods, portfolios, presentations, and participation each had the greatest number of comments related to the assessment being an opportunity to learn with 18, 33, and 22 comments, respectively. Typical comments of this nature are included in Table 5.2.

Table 5.2

**Comments Reflecting the Idea of the Assessment Being an Opportunity to Learn**

---

I like getting marked on portfolios because I can more learn my English.

I like doing presentation because I have to understand many vocabulary and sentences. We can learn a lot in the presentations.

Presentation makes our speaking skills develop. Also, we can get additional knowledge through preparing presentations.

For my this is the best way for learn, you have to participate all time for practice your English and achieve more knowledge.

I believe [participation] is a way to help me learn more.

---

For tests, the most frequent comment (N=18), which accounted for 14% of the comments, had to do with this form of assessment acting as a motivational factor in terms of studying. Typical comments are included in Table 5.3.

Table 5.3

Comments Reflecting the Idea of the Assessment being Motivational


---

I like getting marked on the tests. Because there is a test, I will push myself to study hard. Besides, I care about my grade.

If I will take a test someday, I will study hard more and more, so I often want to take a test.

Tests can push us to study English. Students usually need some pressure to study.

Although the test let me nervous, I still like test. Then I will study more harder.

I like to know how well I did on the test. Getting a good mark encourages me to keep on studying.

If I have a test, I learn for it, I am pushed!

---

For portfolios and presentations, the next most frequent comment was that these assessment methods allowed students to demonstrate their language ability or improvement. For participation the second most frequent category of comments had to do with concerns regarding equity. For example, some participants felt that perhaps not all students had an equal chance to contribute to the class or students might have gotten penalized for being absent from class. The second most frequent type of comment regarding tests was that students liked the feedback that tests provided them. They liked being able to see perceived tangible evidence of their mastery (or non-mastery) of the class material and to compare their achievement to that of their classmates. Because of the number of categories, only a tally of the most frequent categories of comments for the coded results was made using 5% as the cut-off point. The tallies were converted into percentages of total comments and the results are shown in Table 5.4. Samples of typical

comments for all of the codes in Table 5.4 for each of the four types of assessment are included in Appendix I.

Table 5.4

Results of Coding the Questionnaire Comments for the Content of the Response

Tests	Comments	Portfolios	Comments
1. Motivation	14%	1. Opportunity to Learn	16.5%
2. Feedback to Students	13%	2. Demonstration of Ability	12%
3. Demonstration of Ability	10%	3. Feedback to Students	9%
4. Not Reflective of Ability	10%	4. Concerns about Time	9%
5. Assessment Format or Type	10%	5. Assessment Format or Type	7%
6. Opportunity to Learn	9%	6. Concerns about Marking	7%
7. Negative Affective Responses	8%	7. Assessment Resulting in a Product	5%
		8. Assessment Involving a Process	5%
<b>Total Percentage</b>	<b>74%</b>	<b>Total Percentage</b>	<b>70.5%</b>
Presentations	Comments	Participation	Comments
1. Opportunity to Learn	21%	1. Opportunity to Learn	26%
2. Demonstration of Ability	12%	2. Concerns about Marking	18%
3. Negative Affective Response	10%	3. Concerns about Equity	13%
4. Provides Practice	9%	4. Provides Practice	10%
5. Concerns about Marking	7%	5. Demonstration of Ability	8%
6. Concerns about Time Involvement	5%	6. Not Reflective of Ability	6.5%
		7. Motivation	6.5%
<b>Total Percentage</b>	<b>64%</b>	<b>Total Percentage</b>	<b>88%</b>

Another theme present in the qualitative data from the questionnaires, though by no means prevalent in the data, is the notion suggested by some of the participants that



they did not really care about the marks themselves. In the responses on the questionnaire, 11 of the participants independently made comments to the effect that marks were of little consequence to them. This is not a significant number in terms of the 441 total responses but it is worth noting in that the idea resurfaced, unsolicited, in the interviews. A sample of such comments from the questionnaires is shown in Table 5.5.

Table 5.5

Comments Suggesting that Marks are of Little Consequence to the Participants

---

I like [tests] because show me how I have improved my English but I think the most important thing is learn, not your mark on a test.

I don't like getting marked on portfolios. Because portfolios show what you have learned, I prefer it for my own reference, not for marking.

I like [presentations]. It shows me that the way I explain or present is good or not, but marks is not a very important thing.

I don't like to be marked in participation. One should not participate in a class because of a note, a number, but because the class is real good.

In general, I don't like to be marked in any way. It makes me stressed.

In my opinion, marking in ESL studies is not necessary at all. ESL students usually take interest in improving their language anyway.

---

Note: The last two comments in the Table 5.5 were written in the General Comments section of the questionnaire and were therefore not included in the tally provided in Table 5.1.

Another theme that emerged from the qualitative data from the questionnaires has do to with concerns about criteria for marking, particularly for two of the types of performance-based assessment, participation and presentations. Concerns about marking and equity were the second and third most prevalent comments about participation. These three categories can be viewed as inter-related in that they all have to do with

criteria for marking and ensuring fairness. It would seem from the comments that some of the participants were not clear on what criteria they were being judged for participation. Concerns about marking was also the fifth most common category of comment about presentations. Some of the typical comments relating to these issues are found in Table 5.6.

Table 5.6

Comments about Marking Criteria and Equity

---

I think presentation is the most helpful among these four choices to improve student's communication skills. However, when teachers evaluate students, they must have some criteria for fairness.

It's very difficult to mark for the teacher, all the presentations are so different.

I sometimes feel not fair. It's depend on teachers, I guess.

I feel [participation is] unfair because even though I do my best and others don't, we get the same score.

Dislike [participation]: sometimes teachers mark this part unfairly.

I don't know exactly what the teacher is waiting for when he/she wants to evaluate us on participation.

---

Several of the teachers explicitly commented on their questionnaires that students may not like the subjectivity of presentations and participation and that students may not be sure what to do to get a good mark. One teacher summed up the problem by stating (in relation to item number 5, "Students know what they must do to get a good mark on \_\_\_\_\_") that "Students are not always told in advance what they need to do in order to get a good mark on presentations, portfolios, tests, participation," and then went

on to say, "I give them details on how they will be marked before I collect their work." From the teachers' questionnaire responses (and from the teacher interviews), it is clear that some of the teachers do give out explicit criteria for what students need to do to get high marks on the types of assessment. However, based on the students' comments, not all of the teachers make the marking criteria transparent.

### The Qualitative Data from the Interviews

#### Experience with the Four Types of Assessment

##### The students' responses.

In response to the first interview question as to how much experience the 18 participants had with the four types of assessment, not surprisingly, all had had extensive previous experience with tests of all kinds including short answer, long answer, fill-in-the-blanks, and multiple-choice. Of the four types of assessment, portfolios were the least known. Only five of the 18 participants had some experience and/or familiarity with portfolios although three of those five qualified their responses to say that they had had little experience with portfolios. Twelve of the 18 participants had previous experience with presentations and 15 had experienced being assessed on participation. No response patterns related to the home country (or province, in the case of the participant from Quebec) were observed. The amount of experience with the four assessment types seemed to be unique to the individual's experience and not necessarily indicative of the entire educational system of their country. With the exception of one student, all of the answers regarding previous experience with the assessment types matched those answers that had been given by each of the participants on the questionnaires. One of the participants had indicated previous experience with all four

types of assessment prior to beginning her studies at the institution where the study took place but in the interview indicated that she had only experienced tests and participation in her home country. However, she had studied at the institution for more than one session so she had experienced each of the types of assessment prior to filling in the questionnaire.

#### The teachers' responses.

The range of responses provided by the teachers was similar to that provided by the students. All of the teachers felt that their students had extensive experience with tests, some but limited exposure to presentations, some familiarity with the concept of participation although perhaps not the expectations, and little or no experience with portfolios. Several of the teachers commented that of the four types of assessment, students seemed to have the most difficulty understanding how to do portfolios, even after handouts and explanations had been provided.

#### Definitions of the Different Types of Assessment

##### Tests.

All of the 18 student participants were asked to define each of the four types of assessment so that it would be possible to determine whether there was a common understanding of each of the terms. After the initial surprise of the question, for all the participants, defining tests seemed to be the least problematic or difficult, which is not surprising given that this was the most familiar type of assessment for them. A range of representative responses is included in Table 5.7. Tests were described mostly by the participants' giving examples of the kinds of tests that they had experienced, such as multiple-choice or fill in the blanks.

Table 5.7

Student Definitions of a Test


---

I hope it's a way to see if we understand something in the class...a good test, I think, should include everything for the comprehension of the course, we have to write something, we don't have a choice.

To remind about our knowledge. About what we have learned; that's a test.

For me, test is like a tools that the teacher have to know if his or her students are working in the process, are learning in accordance with the parameters...[a test] is like a parameter or measure that the student know I have to work more about any topic.

It is a way for the teacher to know if you know or don't something but is write always. So is the way that he knows that you learned something; it is always writing.

Test? It's a silly things that teachers would tell you when you're going to have the test and you cannot reach and then you have a mark on it. They check all the answer that you did and the important is they give you the mark.

If somebody says test to Korean, they usually get much stress about that...if somebody says tests, it's stress, believe me.

It is like, how could I say, it's like front line in a war. It's really like this. It's the last point. You'll be killed or you'll be rescued because I mean, due to my experiences, it's like this.

---

Other commonly cited characteristics of tests included that tests tended to be written, tended to be scored, and test-takers tended to have no freedom or choice. What is also interesting to note is that the words "mark" or "score" were commonly used in the definition of tests but were rarely mentioned for the three forms of performance-based assessment. Again, this could relate back to the familiarity factor. If tests figured prominently in the educational background of the participants, it is hardly surprising that tests would equate with "marks". Five of the participants described tests as being a

chance to demonstrate ability or mastery of a subject whereas others seemed to provide more of an emotional or in some cases, dramatic definition of what a test meant to them. A table of all of the definitions of tests given in the interviews can be found in Appendix J.

Although the last comment in Table 5.7 is somewhat more dramatic than most of the responses, it does show the tremendous impact personal experience has on attitudes. The significance of this participant's comment became apparent during the course of the interview in which he stated that due to illness, he had missed a university entrance exam for private universities that was offered only once a year. Because he missed that exam, he had tremendous pressure when writing the admissions exam for public universities. As he said himself in the interview, "I had to die for that exam" because had he failed it, he would not have been able to attend university at all. Interestingly, but by all means not surprisingly, the negative reactions were all from participants coming from countries in which testing is frequent and the consequences of exams, particularly high stakes ones such as university admissions, are severe.

### Portfolios.

In the program where the study took place, there is no set definition of portfolio so each teacher does it somewhat differently. Similarly, each student had a slightly different perception of what it was exactly. Most of the participants had experience with a writing portfolio although the mechanics and instructions may have differed somewhat. Two of the participants had done a portfolio in their reading classes. In all cases, their definitions of portfolios involved some process of producing pieces of writing, revising them, putting them together, and then submitting the collection to the teacher for

marking. Rather than describing the process, several of the participants reflected on the purpose of doing a portfolio (see Table 5.8). Another common theme in the definitions was that portfolios involved some kind of revisiting of writing in that they provided an opportunity to look back at previous work and reflect on it. A complete list of definitions can be found in Appendix K.

Table 5.8

Student Definitions of a Portfolio

---

Portfolio, this is a good and new assignment, new job, new work in Canada. Portfolio is a way because the student can show their creativity...not only creativity, you show or talk about their feelings, their likes.

Portfolio is a lot of thing, a lot of what can I say? Items.... Little by little you class your knowledge for to see your degree or progress. ... But the important, why we do the portfolio is for to see where we are, where we was and where we are now.

It is clear. For me, life is like a portfolio in which we have it after. I mean, I have 19 years of life which can be a portfolio, what I've done in my life.

It is a writing again or which we finish in our class before. We write that, review, review, review, we just review again. ... Give to the teacher again and teacher just look how we improve our writing.

Well, you choose some of your works and put together, put them in a nice form and hand them in.

---

Presentations.

All of the participants seemed to have a clear idea about presentations. According to the participants, typically presentations involved choosing a topic, researching or preparing something, and then presenting it to the class or to groups of students. In giving definitions, several participants noted that it was a chance to express themselves in

front of others. A common word that was used in the responses was “show”, one of the characteristics of performance-based assessment. Typical participant definitions of presentations are given in Table 5.9. All of the definitions given can be found in Appendix L.

Table 5.9

Student Definitions of a Presentation

---

Well, my definition of presentation is the teacher give us a topic or we choose a topic and then gives a limited time like 10 minutes or 20 minutes and focus on your topic and then you express your ideas and according to the resources that you find.

Choose the subject and prepare, and prepare for that about topic and I don't know and then give others information about the topic that I choose.

Presentation is can show your opinion but you have to prepare before the presentation maybe for a long time. And then you have to collecting many datas.

I think presentation it's a, you have to show, to show your idea, but not by writing but your work, your speaking and your body language and the topic, I mean the topic of your talking, maybe the paper, the paper maybe come from yours, maybe come from the other. That's why I think the presentation it's that kind of the show how you professional or your opinion or about your thinking about a topic.

A presentation. I think that's the ability to say what you know but is more, you don't have to repeat it in the same way that, that was in the book or in the class. You change it with what you think, what you know in, it's like your words. Say the things but in your words so people can understand.

---

Participation.

Of the four types of assessment, the students found articulating a definition of participation the most problematic. During the interviews, two of the participants asked for clarification of the term and one of them stated that before studying at the institution, she had never heard of the term even though the concept was familiar to her in her own



educational background. Some of the participants equated the term with attendance while others had more complex definitions of the term that included attentiveness, completion of homework assignments, collaborating with fellow classmates, or simply being alert and active in the class. Some typical responses are listed in Table 5.10. The complete set of responses can be found in Appendix M.

Table 5.10

Student Definitions of Participation

---

Participation is that, I think it's basic, it's basic thing to evaluate. It shows students' motivation or how they, how they, how much they want to study. Maybe it's only just sit in the class and just hear, it's not participation, just try to catch something of the teacher or maybe student need to ready to participate, maybe doing homework or at least read what they are going to do, to study.

Participation is like working together, working, team-working. And sometimes the leader is the teacher.

Participation is attendance. Attendance to participate in the class. Then to help each student in the group. Then to make some group ... atmosphere in the group or group activity. So how much.

I didn't have ... a stronger feeling about this word, participation. You have to work together. I mean it's a teamwork. ... I mean you have to pay attention, listen carefully to the teacher what you did, what did the teacher say today and about the teacher's opinion, ..what does the teacher want us to do. You have to do what the teacher ask you to do.

I don't know exactly because in Japan, you know, participation is regarded as kind of obligation so almost every student attended the class so maybe teachers couldn't give any mark according to attendance I think. ... Of course, attitude is included, included in the participation I think. ... 60 % attendance and 40% attitude.

---

Summary of teachers' definitions.

The definitions that the six teachers gave were similar, albeit much more elaborate and detailed, than those that the students offered. The whole gamut of tests was

described. Most teachers commented that portfolios were some kind of collection of work involving a process of going back and reflecting on what had been previously written. Presentations typically were described as involving preparation and then a performance. For the teachers, like for the students, participation was the most difficult to define precisely and the teachers held a wide range of beliefs of what constitutes participation.

#### Likes and Dislikes about the Four Types of Assessment - The Student Responses

Although all of the participants had already been asked this question on the questionnaire, they were asked again during the interview in an attempt to get at more detail and as a check on the accuracy of the responses to the questionnaire items. For all eighteen of the participants, the qualitative written responses on the questionnaires closely mirrored the responses in the interviews in terms of attitude and the nature of the comments. The only differences noted were that several of the participants tended to give both positive and negative characteristics for each assessment type in the interview whereas on the questionnaire may have provided one or the other.

#### Tests.

In answering what they like or disliked about getting marked by tests, the participants' reactions were predictably mixed. The positive attributes of tests closely matched the categories that were evident in the qualitative data from the questionnaires. For that reason, the responses were coded and tallied using the same coding scheme as shown in Table 3.1. The two coding categories with the highest number of comments are shown in Table 5.11.

Among the reasons for having a favourable attitude towards tests, seven of the participants said that they liked getting feedback on their level or on whether they had mastered the material that was being tested. Six of the participants mentioned that tests encouraged them to review their work and motivated them to study. Four of the participants said they liked tests because they thought they were fair while three mentioned familiarity as one of the positive aspects of tests as an assessment method. Similarly, four answered that they generally liked tests but that it depended on the test type.

Table 5.11

Results of Coding the Interview Responses about the four Assessment Types

Tests	Comments	Portfolios	Comments
1. Motivation 2. Negative Affective Response	20% 17.5%	1. Demonstration of Ability 2. Opportunity to Learn	21% 13%
Presentations	Comments	Participation	Comments
1. Opportunity to Learn 2. Negative Affective Response	19% 16%	1. Opportunity to Learn 2. Concerns about Equity	19% 17%

The most common negative comment about getting marked on tests was that tests resulted in stress and pressure, as noted by seven of the participants. Four participants made comments related to the time limitation of tests while three said that they did not like the preparation time that was required for tests. Two mentioned that they disliked

having to memorize information for tests and one said that anything that had been studied was soon forgotten after the test.

### Portfolios.

The most common theme in the positive responses about portfolios was that this form of assessment allowed the participants to improve their writing skill and thus learn (see Table 5.11). Seven participants made comments to this effect. Five noted that they liked having a product at the end to keep as a record of their experience and/or improvement. The fact that it was possible to be creative in portfolios was mentioned by four of the participants.

In terms of negative comments, the most common had to do with the time involved in the preparation of portfolios but only four commented on this aspect, which reflects the range of opinions on this form of assessment. Two participants objected to aspects related to the scoring of portfolios. One felt it was not fair to be evaluated on how much time was spent on the aesthetics of the product while another participant commented on the potentially subjective nature of marking such an assessment type. One participant did not like the fact that other pieces of writing throughout the term were corrected but that she did not receive any marks per se. She felt she would have liked to have feedback in the form of marks throughout the term rather than just at the end. Three of the participants, although generally positive about portfolios, felt that this type of assessment did not show their true ability either because it was possible to receive outside help from others or from grammar checks in word processing programs.

### Presentations.

Twelve of the participants had positive comments or mixed comments about presentations whereas six of the participants had only negative comments about this assessment method. This result appears to be somewhat different from the quantitative findings in which presentations rated the most highly in terms of the overall mean (see Table 4.2) and in terms of receiving the highest mean of the four assessment types for the greatest number of participants (see Figure 4.1). However, as the results of the coding of common themes in the data suggest, students may consider a type of assessment an opportunity to learn, for example, but may still have negative feelings towards having to perform in the assessment. As noted by eight of the participants, the most commonly expressed reason for liking this form of assessment was that it gave students a chance to speak and express themselves in front of their classmates and teachers. Four mentioned that they liked the process of preparing for the presentations because it included research, organization of ideas and in some cases, collaboration with others. For two of the participants, presentations were a good chance to gain more confidence and in effect, change their personality. In particular, one participant explained this process by giving a personal example. In the interview, to demonstrate her feeling the first time she made a presentation, she lowered her voice, and spoke quietly and hesitantly; the second time she spoke a little more forcefully, and the third time she spoke in a loud confident voice. In the interview she said, "But it is very good because modern people, today's people is presentation is important....If I do many presentation, my personality change."

Not surprisingly, the most common theme in the negative comments about presentations had to do with affective responses, nervousness and shyness in particular.

Nine of the participants mentioned some kind of negative affective reaction to having to do presentations. However, six of those felt at the same time that presentations were useful or beneficial for their language learning (see Table 5.11). The next most common theme had to do with marking or scoring the presentation itself. Four participants mentioned that the marking might be unfair. Two suggested that this might be because those who had never done a presentation might not know what to do to succeed and were therefore, at a disadvantage. Another two participants said that the marking might be unfair because it was not clear on what was being evaluated and that it was inappropriate to mark content. As one participant stated:

What I strongly dislike is the fact – well, I do not like getting marked on presentations because everybody who gives a presentation, prepares it and puts lots of effort in it. So it's unfair. It does not – if you really appreciate the work he or she has done, you cannot say your presentation was better than another person's ones. Especially if they are on different topics... There's not a common ground or something for marking.

The issue of transparency of marking criteria will be discussed later in this chapter as well as in the following chapter.

#### Participation.

Most of the participants had either positive or mixed feelings about getting marked on participation. Of the eighteen participants, nine had only praises for participation, while six had both positive and negative comments. Typically, the positive comments had to do with participation being an integral or necessary part of studying

English with seven of the participants commenting on this aspect (see Table 5.11). Three mentioned that participation allowed the teacher to get a feel for how the class was going and whether or not students were enjoying the class. For two of the participants, not participating in the class would be akin to throwing money away. As one so succinctly stated, “If we are here in this building, it’s for to be present. I can take another class, another course by e-mail, I don’t know.”

As for the negative comments, concerns about marking, particularly the problem of the subjectivity of being assessed on participation, were the most commonly expressed, with five of the participants raising these issues as potential problems. Several participants noted that they felt that the teacher just evaluated those who spoke a lot and that not all the students had an equal opportunity to participate. Another two participants commented that they felt it was unclear what was meant by participation and one suggested that some guidelines needed to be available to students (see Table 5.12).

Table 5.12

Comments on Concerns about Marking Participation

---

But again the main point is to define what participation is, to put rule, to make rules...

Participation is just every student has his or her own style of expressing him or herself and his or her own likes and dislikes so it’s hard to find general standards to mark students, to know what is good participation or bad participation.

That’s often just written, “participation”, so no, I don’t know how the teachers see the word.

The first time I understanding will be attendance but I don’t know at the end what the teacher really mark for attendance or how you act in the class.

---

Apart from the comments in Table 5.12, only three other negative comments were made: that being marked on participation was childish, that group marks for participation were unfair, and that students should not be penalized for being absent.

#### Student Likes and Dislikes about the Four Types of Assessment – the Teacher Responses

When the teacher participants were asked what students would either like or dislike about the four types of assessment, their responses closely matched those of the students. For tests, the only slight difference was that teachers tended to cite familiarity more frequently than did the students. Another common theme in the teacher responses was that students might like the objectivity of tests whereas that comment was not so common among the students. With regard to portfolios, the teacher and student responses very closely mirrored each other. Teachers commented that students might have concerns due to the lack of familiarity, the subjectivity of scoring, and the amount of work involved. Similar to what the students said about presentations in their responses, the teachers also felt that students would feel nervous about having to speak in front of others but also commented that the students tended to like presentations after the performance was over. One teacher comment that highlights this duality of emotions is:

They're nerve-wracked about it. Also, what we're asking them to do – I mean, to stand up in front of a group of people and speak in a language that isn't your own – aaagh! And you're going to be assessed! What are the circumstances in which you would be asked to do that? I think for the most part students like them. They like it after the fact.



For participation, the teacher responses also reflected the range of attitudes of the students. The teachers felt that students would think that getting marked on participation would be easy but at the same time might dislike the subjectivity. One teacher joked, “The whole reason I’m participating in this study is I want you to tell me because I don’t know. I can base it on my own experience. I find the idea horrifying that I’m being judged. It makes me self-conscious.” This same feeling was echoed by several of the student participants, too.

### Ranking of the Assessment Methods

#### Students’ responses.

Towards the end of the interview, the participants were asked to rank the four assessment types based on their own preference (see Figure 5.1). Five participants indicated that they liked participation the best although one qualified that statement to say that she only liked individual participation, not group participation.

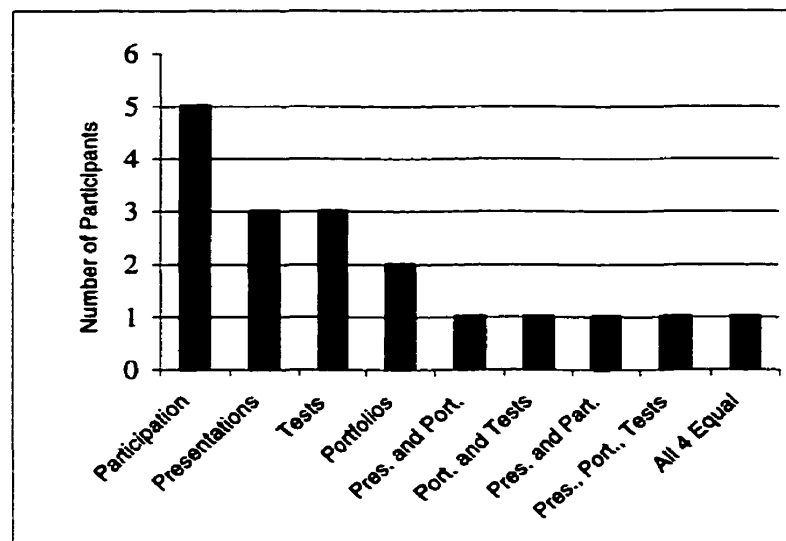


Figure 5.1. Ranking of assessment preferences in the interviews.

Tests and presentations were each preferred by three participants. Being marked on portfolios was the method of assessment of choice for two of the participants while the remaining five participants liked various combinations of assessment methods equally, as shown in Figure 5.1.

As a check on the reliability of participant responses, their assessment preferences as stated in the interviews were compared to those that they had selected on the questionnaires for item 1 (“I would prefer to be marked on \_\_\_\_\_” - See Appendix A). Of the eighteen participants, all had given their preferred method of assessment as per the interviews, the highest rating, or tied for highest rating on the questionnaires. As an additional check, the mean scores for each participant were examined to see if the participant’s preferred method had the highest mean score on all the items on the questionnaire. For thirteen of the eighteen participants, their highest mean score matched or tied their stated assessment preference. For example, if participants ranked presentations as their favourite method of assessment, they also had the highest mean score for presentations as determined by their answers on the questionnaire. One participant of the five, whose preferences did not match her highest mean scores, indicated in the interview a preference for tests but on the questionnaire had a mean score of 3.5 for tests as opposed to 4.3 for presentations. However, in the interview she indicated a difference between preference for a type of assessment and her thoughts on which assessment type was best for her English. Although she said she preferred tests to presentations, she thought the latter was better for improving her language ability. Similarly, another participant ranked participation over presentations in the interview but on the questionnaire these items had mean scores of 4.3 and 4.6,

respectively. However, in the interview the same participant indicated that presentations were the most exciting form of assessment but ranked it second to participation because she felt nervous performing in front of her classmates and teachers. One participant stated that getting marked on participation was her preference but only if it was individual participation. That is possibly why on the questionnaire the mean score of 2.7 for participation was lower than her highest mean score of 3.7 for tests.

### Teachers' perceptions.

Five of the teachers felt that students would prefer tests but two of these said they could not really answer because the preference would depend on the nature of the class and that the method of assessment had to reflect the skill area of the class. One teacher felt that participation would rank the most highly. However, the comments and range of responses for the most part were highly reflective of the comments made by the students.

### Other Ways of Being Assessed

When asked if there was another way of assessment that should be used in an English program, eleven of the eighteen student participants answered that they could not think of any. One suggested that an exit test, similar to the placement test, would be useful so that she could compare her level at the beginning and end of the program. Another participant suggested that an oral interview, such as the one she was experiencing for this study, would be a good method of assessment. Yet another participant suggested that perhaps being marked on speaking only English while in the building might be useful but impossible to implement. Finally, one of the participants

joked that whoever could develop an alternative to tests for assessing language proficiency would be a very wealthy person.

Most of the teachers felt that most types of assessment they used fell into one of the four categories of tests, portfolios, presentations, or participation. One thought that some type of self-assessment might be useful whereas another commented on the effectiveness of assessing students in a writing class on a research paper.

### Issues Arising in the Interviews

#### Marks in a relatively low stakes environment.

As in the qualitative data from the questionnaires, the issue arose in the student interviews of whether an English program such as the one in the study should have marks at all. However, this was by no means the prevailing attitude among the participants. It was the exception rather than the rule but it raises some points that bear mentioning. Typical comments about a lack of concern or need for marks are presented in Table 5.13.

Table 5.13

#### Comments Indicating an Ambivalence Towards Marks

---

I don't care very much about the marks. I think it's the same for everybody. When I talk to everyone, we always think the same. We just want to improve our English so it's for ourselves more than for a diploma, for example. It's not the Same [as other disciplines].

I don't know if it's necessary to be marked at all because everybody here at [name of institution] wants to learn English, so we are all motivated.

Then it's different why do you study. So tests or no test, I must study.

I do participation, never mind mark.

---

For the participants in this study, the stakes were not high because at that time, completion of the highest level of the program meant little more than the personal satisfaction of a certificate in proficiency. Therefore, it is worth noting these comments from the interviews, as was done in the questionnaire data, with the speculation that possible ambivalence towards marks might change in the future for some of the student population as the stakes become higher.

To provide a balance of comments, it is also worth noting that four of the participants in the interviews mentioned that marks were generally motivating. According to these participants, without marks, they would not necessarily study outside of class. However, the prevailing attitude seemed to be that developing their language skills was a priority. One student summed up this sentiment by commenting, “Maybe get more marks is better...but I think improving English is the most important.”

#### Marking criteria

Another theme that emerged from the interviews had to do with marking criteria and subjectivity of marking. Five of the eighteen participants commented on one of these aspects of marking. Comments of this nature were made about all three forms of performance-based assessment but none of the participants made any such remarks about tests. From the comments it seems like not all of the students are being made aware of the marking schemes used for the assessment. Some of the comments reflective of this can be found in Table 5.14. What is telling from the last four comments in Table 5.14 is the repetition of the phrase “I don’t know.” Either some of the teachers did not make the

marking criteria clear to the students, or the information was somehow getting lost in the communication process.

Table 5.14

Comments about Marking and Marking Criteria

---

But portfolio and presentation and participation, there's no points. I mean teacher, only teacher can give points so it depends on the teacher so ... there is some difference between teacher's thinking and student's thinking...if I share the same idea with teacher, I might get a good score. If I write opposite side with teacher, I might not have a good score... If I write the same article ten years ago and now, I might get different point so maybe difficult.

Some people join many presentations, some people not join, don't join but estimate just the teacher estimate that, I don't know.

I don't know if teachers mark the presentators voice or for example, the way of speaking, sometimes unfair for those people, for those who are shy and not get used to do presentations.

[The teachers] just say participation and the first time I understanding will be attendance but I don't know at the end what the teacher really mark for attendance or how you act in the class.

That's just often written participation so I don't know how the teachers see the word. ... Really I don't know how the teacher can see, can think "Oh, this student is very active, participated well in the class and another one, no." I don't know how they can make the difference.

---

Summary of the Qualitative Results

The qualitative results seem to corroborate the quantitative findings in that the participants in this study generally had a positive response to all four assessment types. On the written responses on the questionnaires, positive statements (N=263) far outnumbered those of a negative nature (N=143). For all three types of performance-based assessment, opportunity to learn was the most common theme reflected in the

comments. The results from the coding of the data indicate that the most common comment about tests had to do with aspects of motivation.

The comments in the interviews also suggest that in general, the participants were pleased with the balance of types of assessment used in the language program. All of the participants had previous experience with tests, and some had varying degrees of experience with the three type of performance-based assessment. Asking the participants to define the types of assessment provided a check as to whether the concepts were familiar to them and therefore, as a further check on the validity of the questionnaire. Only the definition of participation proved to be problematic, with two of the participants asking for clarification during the interview. In the interviews, the comments on the positive and negative aspects of each assessment type generally matched those made on the questionnaires. The most common comment about presentations and participation had to do with these assessment types providing an opportunity to learn. This finding was the same as that of the qualitative data from the questionnaires. For portfolios, the most common theme in the interviews was that in this type of assessment it was possible to demonstrate one's ability. The second most frequent comment was that portfolios were an opportunity to learn. These two categories were reversed in the questionnaire data. Similar to the rest of the participants in the study (N=127), the 18 participants in the interview commented on the motivational effect of tests. The highest number of participants ranked participation as their assessment of choice and this finding is different from that in the quantitative data.

As for other ways of being assessed, the prevailing attitude of most of the participants was that they liked the balance of assessments used in the program although some students made comments on the questionnaires and in the interviews that perhaps marks were not necessary at all. Five of the interview participants also commented on some aspect of marking, in particular the subjectivity of assessing portfolios, presentations, and participation and the lack of transparency of the criteria used in marking. These concerns about marking were reflected in the coding of the questionnaire data for participation in particular. The teachers' perceptions of student responses to all of the interview questions echoed those of the students quite closely although as with the quantitative data, teachers seemed to feel that the themes of familiarity and objectivity of tests would be more prevalent in the students' responses than they actually were.

The implications and interpretations of the qualitative and the quantitative results follow in the next chapter. In addition, the research questions of this study will be addressed.



## CHAPTER 6

### Discussion

#### Attitudes to Performance-based Assessment

In answering the first research question of this study, the results seem to indicate that the participants had positive attitudes towards all three forms of performance-based assessment that were under investigation: portfolios, presentations, and participation. The mean item scores for each of the three subconstructs were all on the positive end of the scoring scale, with presentations rated the most favourably (see Table 4.2). Additionally, in the qualitative data from the questionnaires, positive comments outnumbered negative or ambivalent comments. Of the three types of performance-based assessment, presentations received the highest number of positive responses (see Figure 5.1) and the tally of frequencies of highest mean rating for each type of assessment (see Figure 4.1) also corroborate these findings from the qualitative questionnaire data.

The results of the coded qualitative questionnaire data indicate that participants viewed the three performance-based methods of assessment as an opportunity to learn. This is in keeping with the notion that performance assessments involve both products and processes (Aschbacher, 1991; Fitzpatrick & Morrison, 1971). It also is suggestive of the positive washback that is possible with performance-based assessment (Miller & Legg, 1993; Shohamy, 1995; Norris et al., 1998). If students feel that they are learning as a result of engaging in the assessment process, then the type of assessment would seem to have positive pedagogical effects. The second most frequent comment for two of the performance-based assessments, portfolios and presentations, was that they both gave

students the opportunity to demonstrate their ability; one of the defining features of performance-based assessment is that it involves demonstrated behaviour. Shohamy (1983b), in characterizing aspects of performance-based oral assessment, makes the distinction between “doing” and “knowing” and Jones (1985) states that demonstration is integral to performance-based assessment. The performance aspect is one that deserves further investigation since it raises many questions regarding output, feedback and L2 learning. Although students may hold positive views towards performance-based assessment, the effectiveness of such assessment methods for evaluative purposes and for learning requires further research.

#### Resistance to Performance-based Assessment?

To answer the second research question of this study, little evidence was found to support the claims in the literature that performance-based assessment may be unfamiliar or culturally inappropriate for second language learners (Burt & Keenan, 1995; Wrigley & Guth, 1992). Even though the students in the present study were generally unfamiliar with portfolios, this form of performance-based assessment was still viewed positively and moreover, regarded as a vehicle for learning. The other two forms of performance-based assessment, presentations and participation, were quite familiar to the students, as most had had experience with them in their educational systems. Basically, most of the students in this study, with a few exceptions, were very comfortable with the methods of assessment used and viewed all the forms of assessment positively, including tests.

### Attitudes to Traditional Tests versus Performance-based Assessment

In this study, the participants had favourable attitudes towards not only performance-based assessment but also tests, the subject of the third research question. Tests scored on the positive end of the rating scale and in terms of overall mean, was ranked third of the four types of assessment. What differs when comparing attitudes towards performance-based assessment versus attitudes towards tests, however, was the nature of the comments. Participants most frequently commented on the motivation that tests provide. It is interesting, although not entirely surprising, that this type of assessment resulted in a different category of comment being the most common. Motivation can be a positive effect of tests but the participants did not seem to get the same sense of learning when they wrote tests as when they were involved in a performance-based assessment. Although some researchers have demonstrated that motivation encourages learning and affects achievement (e.g. Bradshaw, 1990; Gardner & Lambert, 1972; Shohamy, 1982), the types of comments the participants made seemed to differentiate learning and studying. The fact that “feedback” was the second most frequent comment about tests indicates that the students did like to get some sort of tangible score. Another interesting observation regarding the results of the coding of the data is that there were equal numbers of comments stating that tests demonstrated ability, as there were comments about tests not demonstrating ability. This was also reflected in the interviews. Because students typically have no input into the content of tests and the test questions may or may not tap into what the students have studied or learned individually, sometimes tests do not provide students with a chance to show what they know. However, performance-based assessment methods, because they tend to be more interactive in that they generally require input and output, students have multiple

opportunities to perform and thus might feel like these forms of assessment are more indicative of their ability. Zeidner (1990) had a similar finding in that university students perceived essays, which are performance-based, to be reflective of their knowledge whereas multiple-choice tests were characterized as not indicative of their knowledge.

### Background Variables

#### Language Proficiency Variable

For this background variable, a significant interaction effect was observed in which attitude towards assessment depended on level of language proficiency. Attitude towards presentations was found to be statistically significant with regard to level of language proficiency with intermediate and advanced students rating this form of assessment more highly than did beginner students. Although this is what one might intuit, given that beginner students are less proficient and therefore, likely to be less enthusiastic about having to get up in front of their classmates, several explanations can be offered for the difference between the beginner class and upper level classes. Foremost of these is the possibility that the beginner class did not fully understand the items so that the data were confounded by language ability. Having said that, the beginner participants did not seem to have any problem understanding or answering the questions in the interview but one cannot rule out the possibility that on the questionnaires, some of the beginner students may not have understood all of the items. Another potential reason for the apparent difference in attitude toward presentations is that proposed by Alcorso and Kalantzis (1985) who suggest that less confident learners do not like to be assessed as much as learners with more confidence. Scott and Madsen (1983) also found that students with lower levels of language proficiency rated oral

interview tests less positively than did those students with higher levels of language proficiency. It could be that beginner students, by nature of their level of English, are less confident and thus less willing to perform as part of the assessment process. As noted by one of the participants in the interviews, her confidence increased with each presentation. It seems likely that increases in overall language proficiency could also, therefore, have a positive effect on confidence and affective responses to oral presentations.

However, having said that about level of language proficiency, a few additional points bear mentioning. Because the placement testing at the institution uses different aggregate test scores for placement in the different skill areas of classes, it is possible that some students may be in different levels in each of their three classes. That is to say, based on the placement test scores, a student may be placed in a beginner writing class but in an intermediate speaking class. Because this is a potentially confounding factor in looking at any correlation between level of language proficiency and attitude towards the different forms of assessment, and because through oversight this was not asked on the questionnaire, a tally was made to identify any students for whom this would be the case. The class lists of the two other time slots of the program (not the one in which the data collection took place) were examined but the number of students in different levels for each skill area was not very consequential in that only 18 of the 127 participants had two of their three classes at a different level than that in which the study took place. At the beginner level, 13 of the 38 participants were at different levels in different skill areas with eight students being lower in writing than in speaking and four being lower in speaking than in reading or writing. In other words, in the time slot in which the study

took place, 9 of the beginner students were in a higher speaking class in a different time slot and 4 were in a higher reading or writing class in a different time slot. At the intermediate level, only 5 of the 59 participants were in different level classes in another time slot. All of the advanced students were at the advanced level in all three of their classes.

Another issue to consider when looking for possible correlations between level of proficiency and attitudes to assessment is that the construct of language proficiency itself is problematic as there is still debate among those in the field (e.g. McNamara, 1996) as to what constitutes knowing a language and assessing performance in its use. Looking at language proficiency as a potential variable begs the question “Proficient in what?” However, given the above observation that only 18 of the 127 participants were in different levels in different skill areas, and given that the students were given a battery of tests for placement purposes, it is fairly safe to say that the students were in three distinct groups based on proficiency. The theoretical issue of defining language proficiency is still unresolved, however, but as Shohamy (1998) suggests, “recent trend toward performance testing as a means of alternative assessment, may eventually provide better and more expanded definitions of language ability,” (p. 158).

### Language and Country Variables

Although Willing’s (1988) study differed from the present study in that he was looking at learning styles and not attitudes or preferences per se, his finding that cultural or language background has no effect on attitudes to learning preferences is partially supported by this data set as is Zeidner’s (1988) conclusion that sociocultural background

variables generally do not correlate well with attitudes to types of tests. In the present study, the language background variable was found to be significant for tests but the statistical significance was lost in the post hoc tests in which multiple comparisons were made. However, for the country variable a significant interaction effect was detected. The result suggest that those students coming from Colombia, Japan, Korea, and Taiwan were more inclined to rate tests more highly than those students from Mexico, who tended to rate participation more highly. Since the students from Mexico had the largest standard deviation of all the country groups and seeing as there were only 11 participants, it must be said that these observations are speculative given the relatively small sample sizes and the statistical results. Further studies with larger sample sizes would have to be conducted to attempt to look at any possible correlates between language or culture and assessment preferences.

#### Level of Education Variable

A statistically significant difference between means was found with regard to tests among participants with different levels of education. Although post hoc tests did not reveal any significance, a cursory glance at the means of the three groups for level of education would suggest that university and high school students rated tests more positively than did graduate students. This finding differs from those in the Alcorso and Kalantzis (1985), Kessler (1984) and Brindley (1984 [in Brindley, 1989]) studies in which a high level of education correlated with a preference for formal tests. It must be noted though, that in the Brindley study, the average level of education was just over nine years. However, in the present study all of the participants had a high level of education as over 84% had advanced study beyond the high school level and all had completed high

school. From these studies, it is not clear, nor should it be assumed, that level of education acts as a continuum with preference for tests getting higher as level of education increases. This was certainly not the case in the present study. In fact, there were no significant differences in attitudes to either tests or the three types of performance-based assessment with regard to level of education

### Other Background Variables

Despite the finding that the variables of age, gender, time in Canada, reason for studying and prior experience using each type of assessment revealed no significant differences in the quantitative data, some interesting comments worth noting arose in the interviews. Of these background variables, one might expect experience with the assessment type as having a significant effect on attitude. As Scott and Madsen (1983) found, students' affective responses to a battery of tests got increasingly positive with each successive administration although this may have perhaps been the result of a practice effect more than the result of true familiarity. In the present study, the participants' familiarity with tests was often mentioned but it was by no means the most prevalent comment. Similarly, the lack of familiarity with performance-based assessment, portfolios in particular, did not seem to be problematic for these participants. It seemed like life experience, rather than experience with the assessment, had more of an influence. One participant commented that she thought her answers might have been different had she been interviewed immediately after finishing her university studies. Because she had been in the work force for several years and was routinely required to make presentations, she stated a preference for this form of assessment but noted that several years earlier she probably would have preferred tests. With regard to age, several



of the older (over the age of 28) participants commented that they felt their attitudes might be different from those of younger students. They felt that since they were paying for the program out of their own pockets, and perhaps the younger students were not, the older students might be more motivated in terms of pushing themselves to learn.

#### Teachers' Perceptions of Student Attitudes to Types of Assessment

The last research question of this study was to compare student attitudes with teachers' perceptions of student attitudes since a significant mismatch could possibly impede learning (Gardner & Lambert, 1972). Generally, the teachers' perceptions agreed with those of their students in that the teachers predicted that students would respond positively to all four types of assessment, which was the case. However, in the overall means and the ranking of the four assessment types, teachers' perceptions and students' attitudes differed. Due to the small sample size of the teachers relative to the sample size of the students, no conclusive inferences can be drawn from the data but a few points bear mentioning. The teachers' responses differed from those of the students in that the teachers felt students would rank tests more highly than they did. Teachers also tended to cite that student familiarity with tests might result in a preference for that type of assessment but that was not a common theme in the students' responses.

One clearly common ground among the teachers and students was the ambiguity of participation as evidenced by their ratings of participation as the students' least preferred method of assessment. In the interviews both groups tended to view participation as a grey area of assessment. According to several of the teachers, participation is sometimes intentionally kept vague with a lack of transparency of

assessment criteria so that marks can be shifted in making decisions either to promote students to the next level of proficiency or conversely, to prevent students from progressing to the next level. If participation is to be used as an assessment method as the stakes in the English language program increase for students who want to enter the university, this issue of a lack of criteria for participation will have to be addressed as it raises some of the general concerns about using performance-based methods for assessment rather than pedagogical purposes.<sup>3</sup>

### Issues in Performance-based Assessment

#### Fairness

Although it did not figure prominently in either the quantitative or qualitative data, issues of equity and fairness were mentioned by some of the participants. Tests were perceived as being fair in that everyone in the class had an equal opportunity, or not fair as the tests tended to provide only a snapshot or sampling of ability. Concerns with equity for the three performance-based assessments had mostly to do with transparency of scoring criteria. Marking criteria should be explicit so that students know what they are expected to do (Alderson & Clapham, 1995). As noted above, participation seemed to be particularly problematic in this respect in that students tended not to be clear on what exactly was being assessed. Many of the students equated participation with attendance while several of the teachers explained that they did not consider attendance to be included in participation. This confusion is also relevant to this study as a potential confounding factor. If attendance is weighted heavily in the participation mark, then that

---

<sup>3</sup> It should be noted that at least two of the teachers who were interviewed did have clear guidelines that they had developed for assessing participation.

component of the mark should not be considered a performance-based assessment as it does not include rater judgment, one of the defining criteria of these assessment types. Perhaps the term should have been qualified to indicate in-class participation only.

### Validity

Although the psychometric concerns about performance-based assessment should not be discounted, at this time, they do not pose a real concern in the context of a language program such as the one in this study. However, they should be considered, particularly if in the future the institution will be granting certificates of proficiency that meet the university language requirements. Two aspects of validity are particularly pertinent, although not sufficient by themselves: face validity and consequential validity.

It is important for an English Language Program's assessment procedures to have face validity for both the students and the university at large. If students perceive that they are learning, as the results suggest they do, then that provides some evidence of face validity. The students in this study indicated that they generally liked all four types of assessment and that the three performance-based assessments provided them with an opportunity to learn. According to Nevo (1985) in order to determine to what extent an assessment has face validity, student attitudes should be ascertained. This attention to student attitudes can also guide development and implementation of tests. As outlined in Zeidner and Bensoussan (1988) after Nevo (1985):

In the course of assessing a test's face validity or in determining testing policy and procedure, measurement specialists would appear to agree that examinee's attitudes, perceptions and motivational dispositions concerning various types of tests administered at the university level should be given due consideration and weight by instructors and researchers. (p. 100)

Another aspect of validity that should be considered is consequential validity. With the recent approval of the upper level of the program satisfying the proficiency requirement for the university, the stakes are raised in terms of consequential validity. If the program starts to act as more of a feeder to the university, students may become more concerned with the ways that they are being assessed. In this study none of the participants mentioned the consequences of the assessment. Much of the discussion in the literature has been on the consequential validity of implementing performance-based assessment such as portfolios (Linn, Baker, & Dunbar, 1991; Hamp-Lyons, 1996) and the importance of including consequential validity as an important factor in development of assessments (Shepard, 1997). However, apart from the comments about the assessment of their learning, the participants did not directly address any concerns about the consequences of the marks they received in the program. In general, there was a feeling among the participants that any marks in the program were for themselves and not necessarily important in the long term. It would be interesting to survey the upper level students intending to continue in academic programs at the university to see if they would prefer that their grades be based on performance-based assessment or on more traditional tests. As it stands now, the advanced students, in order to receive a certificate in proficiency, have to pass an interview assessment conducted by two different

raters, in addition to completing the core skill courses at the highest level of the program.

The adult ESL program in this study is in a period of transition between being a formerly low to no stakes environment, as many such non-credit courses are, to being a higher stakes environment in which judgments about students' ability to perform in an academic setting at the university will be determined. This period of transition can be viewed as being along a continuum with an instructionally-oriented, low stakes approach to assessment at one end and more of a summative, high stakes approach to assessment at the other end of the continuum. These ends of the continuum represent the distinction between pedagogy and assessment.

#### Pedagogy vs. Assessment

From a pedagogical perspective, the results of this study are encouraging in that students indicated that they view the three types of performance-based assessment as an opportunity to learn. Based on the students' responses on both the questionnaires and in the interviews, the inclusion of portfolios, presentations, and participation for marks in the English program was considered positive. As has already been mentioned, performance-based assessment had high face validity in the program since students perceived the assessment process as a tool for learning. Because at the time the stakes were low for the students, it is not surprising that anything considered to enhance their learning was viewed as being valid since presumably the only motivation they would have for entering the program at all would be to improve their English proficiency. Although the present study found no significant correlation between reason for studying and attitude to assessment, in a few years' time, it would be interesting to survey the

student population in the Intensive English Program to see if a shift in attitudes towards assessment has occurred. Results of such a study might be confounded by the fact that in the future, the program might attract students with different motivations for studying English. If the stakes are raised, some students might object to having their language admission requirement for the university hinge on being assessed on performance-based assessment, particularly if it is viewed as being subjective or unfair in any way (or conversely, such students might view such a program as less of a hurdle than passing the TOEFL test).

While it is difficult to dispute the pedagogical benefits of performance-based assessment as it is being used in the Intensive English Program in this study, there are some areas that are being addressed and have yet to be addressed as the transition from low stakes to higher stakes progresses. One of the areas is the establishment of common weighting of assessment criteria for each skill area class at each level, which had already been started at the time the study took place. This is particularly important for classes at the top level of proficiency so that all students receive the same distribution of marks across multiple sections of the same skill area. Another area of concern is the use of participation as a form of assessment. Without clear criteria, participation should not be included in the forms of assessment. The teachers and administration of the program had started grappling with this issue but at the time of the study, no definitive definition, weighting or criteria by which to judge it had been agreed upon. Transparency of marking criteria is a defining feature of performance-based assessment so validated scoring rubrics for presentations, portfolios, and even participation need to be developed so that students are judged on the same criteria within each skill area. At present,

individual teachers devise their own scoring schemes, which they may or may not show the students in advance of the performance. Finally, although subjectivity can never be eliminated from performance-based assessment, it is possible to reduce the rater variability through training, particularly if common rating schemes are adopted or through the use of multiple raters (Bachman & Palmer, 1996; Houston, Raymond, & Svec, 1991; McNamara, 1996).

When using performance-based assessment as a pedagogical tool, classroom teachers in adult ESL programs do not need to be overly concerned with issues of reliability, transparency of marking criteria and validity. However, if the assessment has high stakes consequences, then the distinction between pedagogical and assessment purposes becomes more pronounced.

### Limitations of the Study

#### Scale Reliability and Validity

Although the instrument scale was found to have a relatively high internal consistency in both the pilot study and the present study, this alone is not adequate evidence that the items in the scale reflect one construct of attitudes towards assessment with four subconstructs of attitudes to tests, portfolios, presentations and participation. However, the scale reliability calculated, together with the results of the factor analysis, does provide some evidence for the reliability and validity of the instrument. The interviews also acted as a kind of reliability check as some of the participants' answers could be compared to those they gave on the questionnaires. The similarity in responses in the interviews and on the questionnaires, lend support to the validity and reliability of

the scale but more evidence, in the form of similar studies with larger sample sizes, is needed. Finally, the use of a negative item (item 7) in the scale proved to be problematic, as researchers in the literature caution (e.g. Melnick & Gable, 1990), so in future studies that one negatively worded item would need to be reworded.

### Preference vs. Attitude

Through conducting the interviews and comparing the quantitative and qualitative data, a distinction between attitude and preference became apparent. That is, some participants may not have liked any given type of assessment but at the same time recognized that the assessment either aided their language learning or motivated them. With the exception of the first item, the items on the questionnaire mostly have to do with ability, opportunity to learn, equity, and motivation. A high mean score on each item shows a favourable attitude but not necessarily a preference. It would be interesting to conduct a study to test whether indeed, attitude and preference are interchangeable or distinct terms. This study tapped into both notions in that the quantitative data appeared to measure attitude or response to the statements whereas the qualitative data elicited both.

### Individuality of Preferences

Every learner has his or her own style, motivation and preferences for learning. As Bachman (1990) states, “individuals with different backgrounds and personalities may perform differentially in different types of language tests,” (p. 113). The key notion in this statement is “individuals”. An investigation such as this in which an attempt was made to survey general attitudes can be suggestive but not conclusive. However, the data



do suggest some common patterns so that once more of a body of literature is built up around the topic of student attitudes towards assessment procedures, it may be possible to see if these patterns are in any way generalizable.

### The Connection between Attitude, Marks and Performance

Because it would have been too intrusive, the participants were not asked what their marks were. It is possible that attitude toward each assessment type could be affected by the marks received although some researchers have found otherwise (e.g. Zeidner, 1988). However, the general feeling among the participants seemed to be that the actual mark itself was of little import. What mattered to the students was the feedback about their learning. This apparent lack of concern about marks might have influenced responses considerably. A semantic differential or a few items about the importance of marks in the context of the Intensive English Program might have proved insightful. With regard to performance, the literature suggests that types of assessment can affect scores on tests (Bachman & Palmer, 1990; Shohamy, 1982; Shohamy, 1997b). It follows that attitudes towards types of assessment probably also affect performance.

### Implications of the Findings

The findings of this study suggest that the participants in the Intensive English Program at the university studied view the balance of assessment practices, which includes tests as well as performance-based assessment, in a positive light. This has several implications for the program itself. Since formalized assessment had just been implemented slightly over a year before the study took place, the findings of this study suggest that in general, the students were positive towards being assessed as all four types

of assessment were seen as a means to create motivation or provide opportunities to learn. The majority of the assessment methods used in the English language program in this study are performance-based and if one subscribes to the view that these types of assessment have positive washback effects (e.g. Miller & Legg, 1993; Shohamy, 1995; Norris et al., 1998), then the finding that the students react positively to them, is significant in terms of pedagogical implications.

Assessment methods have an impact on what is learned and taught (Crooks, 1988; Gardner & MacIntyre, 1993; Moss et al., 1992) so it is important to gauge how they are implemented and how students respond to them. The most common theme in the student comments about performance-based assessment was that it provided an opportunity to learn. In an Intensive English Program where students may or may not want to continue their studies in degree programs, this is important to note. If the assessment methods are viewed as a natural extension to the typical classroom activities in a communicative language classroom, then regardless of whether the stakes of the assessment are low or high for each individual student, the students are likely to find the performance-based methods relevant and engaging.

Knowing students' attitudes to what goes on in the classroom can possibly help to provide a better learning environment. Affective variables may influence performance, which may in turn lead to inaccurate assessment. If as teachers, researchers, and test developers, we know what student perceptions are, we are better able to provide the kinds of instructions, task types, and feedback that the students need in order to perform at their best.

## References

- Alcorso, C., & Kalantzis, M. (1985). *The learning process and being a learner in the AMEP*. Paper prepared for the Committee of Review of the Adult Migrant Education Program. Wollongong, Australia: Centre for Multicultural Studies.
- Alderson, J.C. (1986). Innovations in language testing? In M. Portal, (Ed.), *Innovations in language testing* (pp. 93-108). Windsor, England: NFER-Nelson.
- Alderson, J.C., & Clapham, C. (1995). Assessing student performance in the ESL classroom. *TESOL Quarterly*, 29 (1), 184-187.
- Aschbacher, P.R. (1991). Performance assessment: State activity, interest, and concerns. *Applied Measurement in Education*, 4 (4), 275-288.
- Bachman, L.F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L.F., & Palmer, A.S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Baker, N.W. (1993). The effect of portfolio-based instruction on composition students' final examination scores, course grades, and attitudes. *Research in the Teaching of English*, 27 (2), pp. 155-174.
- Baker, E. L., & O'Neil, H.F. (1996). Performance Assessment and Equity. In M.B. Kane & R. Mitchell (eds.) *Implementing Performance Assessment - Promises, Problems and Challenges*. (pp. 183-199). Mahway, New Jersey: Lawrence Erlbaum Associates.
- Bane, J. (1999). Student reactions to Internet reading portfolios [Abstract]. *Language Testing Update*, 25, 65.
- Barnwell, D.P. (1996). *A History of foreign language testing in the United States from its beginning to the present*. Tempe, AZ: Bilingual Press, Arizona State University.
- Bradshaw, J. (1990). Test-takers' reactions to a placement test. *Language Testing*, 7 (1), 13-30.
- Brindley, G. (1989). *Assessing Achievement in the learner-centered curriculum*. Sydney, Australia: National Centre for English Language Teaching and Research.

- Brindley, G. (1998). Describing language development? Rating scales and SLA. In L.F. Bachman & A.D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 112-140). Cambridge: Cambridge University Press.
- Brooks, L.A. (1998). *Student attitudes to performance-based assessment: A Pilot study*. Unpublished course paper, Ontario Institute for Studies in Education at the University of Toronto.
- Brown, J.D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32 (4), 653-675.
- Burt, M., & Keenan, F. (1995). *Adult ESL Learner Assessment: Purposes and Tools*. ERIC Digest. (ERIC Document Reproduction Service No. ED 386 962)
- Canale, M. (1983). On some dimensions of language proficiency. In J.W. Oller (Ed.), *Issues in language testing research* (pp. 333-342). Rowley, MA: Newbury House Publishers.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1 (1), 1-47.
- Carey, P.A. (1996). *A Review of psychometric and consequential issues related to performance assessment*. (TOEFL Monograph Series No. 3). Princeton, NJ: Educational Testing Service.
- Carroll, J.B. (1961). Fundamental considerations in testing for English language proficiency in foreign students. In H.B. Allen & R.N. Campbell (Eds.), 1972, *Teaching English as a second language: a book of readings*. (2nd ed., pp. 313-321). New York: McGraw Hill.
- Cray, E., & Currie, P. (1996). Linking adult learners with the education of L2 teachers. *TESOL Quarterly*, 30 (1), 113-130.
- Crooks, T.J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58 (4), 438-481.
- Cummins, J. (1984). *Bilingualism and special education: Issues in assessment and pedagogy*. Clevedon, England: Multilingual Matters.
- Damico, J.S. (1991). Performance assessment of language minority students. *In Focus on Evaluation and Measurement Volumes 1 and 2: Proceedings of the National Research Symposium of LEP Student Issues* (pp. 137-171). Washington, D.C. Sept. 4-6, 1991.

- Darling-Hammond, L. (1994). Performance-based assessment and educational equity. *Harvard Educational Review*, 64 (1), 5-30.
- Dunbar, S.B., Koretz, D.M., & Hoover, H.D. (1991). Quality control in the development and use of performance assessments. *Applied Measurement in Education*, 4 (4), 289-303.
- Eyring, J. (1997). *Is project work worth it?* ERIC Research Report. (ERIC Document Reproduction Service No. ED 407838).
- Fitzpatrick, R., & Morrison, E.J. (1971). Performance and product evaluation. In R.L. Thorndike (Ed.), *Educational measurement*. (2nd ed., pp. 237-270). American Council on Education, Washington, D.C.
- Frechtling, J.A. (1991). Performance assessment: Moonstruck or the real thing? *Educational Measurement: Issues and Practice*, 10 (4), 23-25.
- Gardner, R.C., & Lambert, W.E. (1972). *Attitudes and motivation in second language learning*. Rowley, MA: Newbury House Publishers.
- Gardner, R.C., & MacIntyre, P.D. (1993). On the measurement of affective variables in second language learning. *Language Learning*, 43 (2), 157-194.
- Gipps, C.V. (1994). *Beyond testing: Towards a theory of educational assessment*. London: The Falmer Press.
- Grierson, J. (1995). Classroom-based assessment in Intensive English Centres. In G. Brindley (Ed.) *Language Assessment in Action*. (pp. 195-237). Sydney, Australia: National Centre for English Language Teaching and Research.
- Haertel, E. (1992). Performance measurement. In *The Encyclopedia of educational research*. (Vol. 3, pp. 984-989). New York: Macmillan Publishing Company.
- Hambleton, R.K., & Murphy, E. (1992). A Psychometric perspective on authentic measurement. *Applied Measurement in Education*, 5 (1), 1-16.
- Hamp-Lyons, L. (1996). Applying ethical standards to portfolio assessment of writing in English as a second language. In M. Milanovic & N. Saviile (Eds.), *Studies in language testing 3: Performance testing, cognition and assessment* (pp. 151-164). Cambridge, England: Cambridge University Press.
- Hamp-Lyons, L. (1997). Ethics in language testing. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language and education: Vol. 7. Language testing and assessment* (pp. 323-333). Dordrecht, The Netherlands: Kluwer Academic Publishers.

- Herman, J.L., Aschbacher, P.R., & Winters, L. (1992). *A Practical Guide to Alternative Assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Herman, J.L., Klein, D.C.D., & Wakai, S.T. (1997). *American students' perspectives on alternative assessment: Do they know it's different?* (CSE Technical Report 439). Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Hill, K. (1998). The effect of test-taker characteristics on reactions to and performance on an oral English proficiency test. In A.J. Kunnan (Ed.), *Validation in language assessment*, (pp. 209-229). Mahwah, NJ: Lawrence Erlbaum Associates.
- Houston, W.M., Raymond, M.R., & Svec, J.C. (1991). Adjustments for rater effects in performance assessment. *Applied Psychological Measurement*, 15 (4), 409-421.
- Hymes, D.H. (1967). Models of the interaction of language and social setting. *Journal of Social Issues*, 23(2), 8-38.
- Jones, R.L. (1985). Second language performance testing: An Overview. In P.C. Hauptman, R. Leblanc & M. Bingham Wesche (Eds.), *Second language performance testing* (pp. 15-24). Ottawa, Canada: University of Ottawa Press.
- Kessler, S. (1984). *AMEP Wastage Survey*. Sydney, Australia: AMES.
- Lado, R. (1960). English language testing: Problems of validity and administration. *English language teaching*, 14, 153-161.
- Lado, R. (1964). *Language testing*. New York: McGraw-Hill.
- Lam, T. (1995). *Fairness in performance assessment*. ERIC Digest. (ERIC Document Reproduction Service No. ED 391 982)
- Linn, R.L., Baker, E.L., & Dunbar, S.B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, 20 (8), 15-21.
- McNamara, T. (1996). *Measuring second language performance*. Harlow, England: Addison Wesley Longman.
- McNamara, T. (1997). Performance testing. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language and education: Vol. 7. Language testing and assessment* (pp. 131-139). Dordrecht, The Netherlands: Kluwer Academic Publishers.

- Mehrens, W.A. (1992). Using performance assessment for accountability purposes. *Educational Measurement: Issues and Practice*, 11 (1), 3-9, 20.
- Melnick, S.A., & Gable, R.K. (1990). The use of negative item stems: a cautionary note. *Educational Research Quarterly*, 14 (3), 31-36.
- Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement*. (3rd ed., pp. 13-103). New York: American Council on Education and Macmillan.
- Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23 (2), 13-23.
- Meyer, C.A. (1992). What's the difference between authentic and performance assessment? *Educational Leadership*, 49 (8), 39-40.
- Miller, M.D., & Legg, S.M. (1993). Alternative assessment in a high-stakes environment. *Educational Measurement: Issues and Practice*, 12, 9-15.
- Moss, P.A. (1994). Can there be validity without reliability? *Educational Researcher*, 23 (2), 5-12.
- Moss, P.A., Beck, J.S., Ebbs, C., Matson, B., Muchmore, J., Steele, D., Taylor, C., & Herter, R. (1992). Portfolios, accountability, and an interpretive approach to validity. *Educational Measurement: Issues and Practice*, 11 (3), 12-21.
- Murray, D. (1994). Using portfolios to assess writing. *Prospect*, 9 (2), 56-65.
- Nevo, B. (1985). Face validity revisited. *Journal of Educational Measurement*, 22 (4), 287-293.
- Norris, J.M., Brown, J.D., Hudson, T., & Yoshioka, J. (1998). *Designing second language performance assessments*. Honolulu, HI: University of Hawai'i Press.
- Norusis, M.J. (1998). *SPSS guide to data analysis*. Upper Saddle River, N.J.: Prentice-Hall, Inc.
- Nunan, D. (1995). Closing the gap between learning and instruction. *TESOL Quarterly*, 29 (1), 133-158.
- O'Malley, J. M., & Pierce, L.V. (1996). *Authentic assessment for English language learners: Practical approaches for teachers*. New York: Addison-Wesley Publishing.
- Pierce, L.V., & O'Malley, J.M. (1992). *Performance and portfolio assessment for language minority students*. Wheaton, MD: National Clearinghouse for Bilingual Education.

- Savignon, S. (1972). *Communicative competence: an experiment in foreign language teaching*. Philadelphia, PA: The Center for Curriculum Development.
- Scott, M.L., & Madsen, H.S. (1983). The influence of retesting on test affect. In J.W. Oller (Ed.), *Issues in language testing research* (pp. 270-279). Rowley, MA: Newbury House Publishers.
- Shepard, L.A. (1997). The centrality of test use and consequences for test validity. *Educational Measurement: Issues and Practice*, 16 (2), 5-8, 13, 24.
- Shohamy, E. (1982). Affective considerations in language testing. *Modern Language Journal*, 66, 13-17.
- Shohamy, E. (1983a). Interrater and intrarater reliability of the oral interview and concurrent validity with cloze procedure in Hebrew. In J.W. Oller (Ed.), *Issues in language testing research* (pp. 229-236). Rowley, MA: Newbury House Publishers.
- Shohamy, E. (1983b). The stability of oral proficiency assessment on the oral interview testing procedures. *Language Learning*, 33 (4), 527-540.
- Shohamy, E. (1995). Performance assessment in language testing. *Annual Review of Applied Linguistics*, 15, 188-211.
- Shohamy, E. (1997a). Second language assessment. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language and education: Vol. 4. Second language education* (pp. 141-149). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Shohamy, E. (1997b). Testing methods, testing consequences: Are they ethical? Are they fair? *Language Testing*, 14 (3), 340-349.
- Shohamy, E. (1998). How can language testing and SLA benefit from each other? The case of discourse. In L.F. Bachman & A.D. Cohen (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 156-176). Cambridge: Cambridge University Press.
- Short, D. (1993). Assessing integrated language and content instruction. *TESOL Quarterly*, 27 (4), 627-656.
- Spector, P.E. (1992). *Summated rating scale construction: An introduction*. Newbury Park, CA: Sage Publications.
- Spolsky, B. (1968). Language testing – the problem of validation. *TESOL Quarterly*, 2 (2), 88-94.



- Spolsky, B. (1995). *Measured words*. Oxford: Oxford University Press.
- SPSS Advanced statistics 7.5. (1997). Chicago, IL: SPSS, Inc.
- SPSS for Windows, Release 8.0.0 [Computer software]. (1997). Statistical Product and Service Solutions.
- Stiggins, R.J. (1987). Design and development of performance assessments. *Educational Measurement: Issues and Practices*, 6 (1), 33-42.
- Supovitz, J.A., & Brennan, R.T. (1997). Mirror, mirror on the wall, which is the fairest test of all? An examination of the equitability of portfolio assessment relative to standardized tests. *Harvard Educational Review*, 67 (3), 472-506.
- Swain, M. (1984). Large-scale communicative language testing: A Case study. In S.J. Savignon & M.S. Berns (Eds.), *Initiatives in communicative language teaching* (pp. 185-201). Reading, MA: Addison-Wesley Publishing Company.
- Swain, M. (1993). Second language testing and second language acquisition: is there a conflict with traditional psychometrics? *Language Testing*, 10 (2), 193-207.
- Tannenbaum, J. (1996). *Practical Ideas on alternative assessment for ESL students*. ERIC Digest. (ERIC Document Reproduction Service No. ED 395 500)
- Weir, C.J. (1990). *Communicative language testing*. Hemel Hempstead, England: Prentice Hall International.
- Willing, K. (1988). *Learning Styles in Adult Migrant Education*. Adelaide, Australia: National Curriculum Resource Centre.
- Wiggins, G.P. (1993). Assessment: Authenticity, context and validity. *Phi Delta Kappan*, 74 (6), 200-214.
- Wolf, D.P., Bixby, J., Glenn, J., & Gardner, H. (1991). To use their minds well: Investigating new forms of student assessment. *Review of Research in Education*, 17, 31-74.
- Wrigley, H.S., & Guth, G.J.A. (1992). *Adult ESL literacy: Programs and practice*. Technical Report. San Mateo, CA: Aguirre International. (ERIC Document Reproduction Service No. ED 348 895)
- Yap, K.O. (1993). *Integrating assessment with instruction in ABE/ESL programs*. Paper presented at the annual meeting of the American Educational Research Association. (ERIC Document Reproduction Service No. ED 359 210)

- Yawkey, T.D., Gonzalez, V., & Juan, Y. (1994). Literacy and biliteracy strategies and approaches for young culturally and linguistically diverse children: Academic excellence P.I.A.G.E.T. comes alive. *Journal of Reading Improvement, 31* (3), 130-141.
- Yen, W.M. (1997). The technical quality of performance assessments: Standard errors of percents of pupils reaching standards. *Educational Measurement: Issues and Practice, 16* (3), 5-15.
- Zeidner, M. (1988). Sociocultural differences in examinees' attitudes toward scholastic ability exams. *Journal of Educational Measurement, 25* (1), 67-76.
- Zeidner, M. (1990). College students' reactions towards key facets of classroom testing. *Assessment and Evaluation in Higher Education, 15* (2), 151-169.
- Zeidner, M., & Bensoussan, M. (1988). College students' attitudes towards written versus oral tests of English as a foreign language. *Language Testing, 5* (1), 100-114.

## Appendix A

**Student Questionnaire - Attitudes about Different Kinds of Assessment**

For my research, I would like to know how you feel about getting marked on tests, portfolios, presentations and participation in your classes. Here are some terms you will need to understand before you fill out the questionnaire.

Tests - for example, multiple choice tests, fill in the blank tests

Portfolios - collections of your writing which are put in a folder and then evaluated

Presentations - you prepare a talk on a topic and present it to your class

Participation - you are marked on things like attendance, attitude, cooperation in group work, completing class assignments

**Please fill out this information.**

Male  Female

Home Country \_\_\_\_\_ First Language \_\_\_\_\_

Age  18 - 22  23 - 27  28 - 32  33 - 37  38 or over

Length of time in Canada (from when you arrived until now)

less than 2 months  2-4 months  5-8 months  9-12 months  more than 1 year

Educational Background (level you are in or have finished)

high school  college or university  technical school  graduate school

Reason(s) for studying English You can put a check (✓) for more than one.

for college or university  for traveling  for my job  other \_\_\_\_\_

**Before** coming to the (name of institution), which kinds of assessment had you experienced? You can put a check (✓) for more than one.

Tests  Portfolios  Presentations  Participation

How much of your marks at the (name of institution) are determined by these kinds of assessment? (If you mark "none", it means that you are not marked on that type of assessment in any of your classes. "A lot" means that it counts towards a large percentage (%) of your marks).

	(0%)	(5-10%)	(11-20%)	(more than 20%)
<b><u>Tests</u></b>	<input type="checkbox"/> None	<input type="checkbox"/> A little	<input type="checkbox"/> Some	<input type="checkbox"/> A lot
<b><u>Portfolios</u></b>	<input type="checkbox"/> None	<input type="checkbox"/> A little	<input type="checkbox"/> Some	<input type="checkbox"/> A lot
<b><u>Presentations</u></b>	<input type="checkbox"/> None	<input type="checkbox"/> A little	<input type="checkbox"/> Some	<input type="checkbox"/> A lot
<b><u>Participation</u></b>	<input type="checkbox"/> None	<input type="checkbox"/> A little	<input type="checkbox"/> Some	<input type="checkbox"/> A lot

Please circle one of the five opinions for each statement. Circle your opinion (Strongly Agree, Agree, Disagree, Strongly Disagree or Don't know) for tests, portfolios, presentations AND participation.

1. I would prefer to be marked on \_\_\_\_\_.

<b><u>tests</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>portfolios</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>presentations</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>participation</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know

2. Getting marked on \_\_\_\_\_ is good because I have opportunity to show my ability.

<b><u>tests</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>portfolios</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>presentations</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>participation</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know

3. Getting marked on \_\_\_\_\_ helps me learn.

<b><u>tests</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>portfolios</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>presentations</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>participation</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know

4. I think getting marked on \_\_\_\_\_ is fair to me.

<b><u>tests</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>portfolios</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>presentations</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>participation</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know

5. I know what I must do to get a good mark on \_\_\_\_\_.

<b><u>tests</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>portfolios</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>presentations</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>participation</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know

6. Getting marked on \_\_\_\_\_ helps me improve my English.

<b><u>tests</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>portfolios</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>presentations</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>participation</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know

7. \_\_\_\_\_ take(s) too much time.

<b><u>tests</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>portfolios</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>presentations</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>participation</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know

8. \_\_\_\_\_ are (is) good for showing my ability in English.

<b><u>tests</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>portfolios</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>presentations</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>participation</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know

9. \_\_\_\_\_ should be used for marks.

<b><u>tests</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>portfolios</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>presentations</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>participation</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know

10. Getting marked on \_\_\_\_\_ makes me put in more effort in class.

<b><u>tests</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>portfolios</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>presentations</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>participation</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know

**Please answer the following questions.**

What do you like or dislike about getting marked on **tests**? \_\_\_\_\_

---



---



---

What do you like or dislike about getting marked on **portfolios**? \_\_\_\_\_

---



---



---

What do you like or dislike about getting marked on **presentations**? \_\_\_\_\_

---



---



---

What do you like or dislike about getting marked on **participation**? \_\_\_\_\_

---



---



---

Comments \_\_\_\_\_

---

Thank you!

## Appendix B

### Teacher Questionnaire - Student Attitudes about Different Kinds of Assessment

For my research, I would like to know what you think **students'** attitudes are towards getting marked on tests, portfolios, presentations and participation in their classes. Here are some definitions of the terms used in the questionnaire.

- Tests - for example, multiple choice tests, fill in the blank tests
- Portfolios - collections of student writing which are put in a folder and then evaluated
- Presentations - a student prepares a talk on a topic and presents it to the class
- Participation - students are marked on things like attendance, attitude, cooperation in group work, completing class assignments

#### Please fill out this information.

##### Teaching Experience

- 1-4 years       5-8 years       9-12 years       13-16 years       more than 16 years

##### Level(s) Currently Teaching

- Level 100 or 200       Level 300 or 400       Level 500 or 600

Put a check (✓) beside those forms of assessment you are currently using in either of your classes

- Portfolios       Presentations       Participation       Tests

What weight is given to the following forms of assessment in your class(es)? Please fill in a percentage for the weighting of each of the four types of assessment.

##### **Class 1**

- Level 100 or 200       Level 300 or 400       Level 500 or 600

\_\_\_\_\_ portfolios      \_\_\_\_\_ presentations      \_\_\_\_\_ participation      \_\_\_\_\_ tests

Other types of assessment used \_\_\_\_\_

##### **Class 2 (if applicable)**

- Level 100 or 200       Level 300 or 400       Level 500 or 600

\_\_\_\_\_ portfolios      \_\_\_\_\_ presentations      \_\_\_\_\_ participation      \_\_\_\_\_ tests

Other types of assessment used \_\_\_\_\_

**Please circle one of the five opinions for each statement. Circle your opinion (Strongly Agree, Agree, Disagree, Strongly Disagree or Don't know) for tests, portfolios, presentations AND participation.**

1. I think students would prefer to be marked on \_\_\_\_\_.

<u>tests</u>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<u>portfolios</u>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<u>presentations</u>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<u>participation</u>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know

2. Students think getting marked on \_\_\_\_\_ is good because they have opportunity to show their ability.

<u>Tests</u>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<u>portfolios</u>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<u>presentations</u>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<u>participation</u>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know

3. Students think getting marked on \_\_\_\_\_ helps them learn.

<u>tests</u>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<u>portfolios</u>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<u>presentations</u>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<u>participation</u>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know

4. Students think getting marked on \_\_\_\_\_ is fair to them.

<u>tests</u>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<u>portfolios</u>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<u>presentations</u>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<u>participation</u>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know



5. Students know what they must do to get a good mark on \_\_\_\_\_.

<b><u>tests</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>portfolios</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>presentations</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>participation</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know

6. Students think getting marked on \_\_\_\_\_ helps them improve their English.

<b><u>tests</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>portfolios</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>presentations</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>participation</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know

7. Students think doing \_\_\_\_\_ take(s) too much time.

<b><u>tests</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>portfolios</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>presentations</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>participation</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know

8. Students think \_\_\_\_\_ are (is) good for showing their ability in English.

<b><u>tests</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>portfolios</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>presentations</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>participation</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know

9. Students think \_\_\_\_\_ should be used for marks.

<b><u>tests</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>portfolios</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>presentations</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know
<b><u>participation</u></b>	Strongly Agree	Agree	Disagree	Strongly Disagree	Don't know

10. Students think getting marked on \_\_\_\_\_ makes them put in more effort in class.

**tests**                      Strongly Agree      Agree      Disagree      Strongly Disagree      Don't know

**portfolios**                      Strongly Agree      Agree      Disagree      Strongly Disagree      Don't know

**presentations**                      Strongly Agree      Agree      Disagree      Strongly Disagree      Don't know

**participation**                      Strongly Agree      Agree      Disagree      Strongly Disagree      Don't know

**Please answer the following questions.**

What do students like or dislike about getting marked on **tests**? \_\_\_\_\_

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

What do students like or dislike about getting marked on **portfolios**? \_\_\_\_\_

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

What do students like or dislike about getting marked on **presentations**? \_\_\_\_\_

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

What do students like or dislike about getting marked on **participation**? \_\_\_\_\_

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

Comments \_\_\_\_\_

\_\_\_\_\_

Thank you!

## Appendix C

### Interview Questions for Student Interviews

These questions will be preceded by a warm-up in which the interviewee is put at ease.

1. Before your current studies, how much experience did you have with these four types of assessment?
2. How would you define the four types of assessment? That is, what do you mean by presentations, portfolios, participation and tests?
3.
  - a. What do you like or dislike about getting marked on tests?
  - b. What do you like or dislike about getting marked on portfolios?
  - c. What do you like or dislike about getting marked on presentations?
  - d. What do you like or dislike about getting marked on participation?
4. Which do you prefer tests, portfolios, presentations or participation and why?
5. Is there another way of getting marked that you think you would prefer?

## Appendix D

### Interview Questions for Teacher Interviews

1. Before their current studies, how much experience do you think your students had with these four types of assessment? Did your students seem to be familiar with these four types of assessment at the beginning of the program?
2. How would you define the four types of assessment? That is, what do you mean by presentations, portfolios, participation and tests?
3.
  - a. What do students seem to like or dislike about getting marked on tests?
  - b. What do students seem to like or dislike about getting marked on portfolios?
  - c. What do students seem to like or dislike about getting marked on presentations?
  - d. What do students seem to like or dislike about getting marked on participation?
4. Which do you think students prefer: tests, portfolios, presentations or participation and why?
5. Is there another way of getting marked that you think the students would prefer?

**Appendix E****Procedures for Administering the Questionnaires**

Dear 10:30 teachers,

Here are the questionnaires on student attitudes to different forms of classroom-based assessment that I e-mailed you about. Please take a look at the questionnaires and if you have any specific questions, e-mail me before giving them out to your class (e-mail address).

If you agree to give your class the questionnaires, please do so between (dates given). Please follow these procedures so that the administration of the questionnaire is relatively consistent across classes.

1. Explain to your students that you are helping out a colleague with some research about different ways of assessing students.
2. Give each student a letter from me and please go over the key points of the letter:
  - they do not have to participate
  - their answers will remain anonymous
  - I will not show their individual answers to you (although I will provide you with a summary of the compiled results, of course)
  - they are not answering about your class only - they could have experienced these forms of assessment in any of their classes this session or in previous sessions
3. Have each willing student sign the letter.
4. Collect the letters and put them in the letter envelope. Give each student who has signed the letter a copy of the questionnaire.
5. Please make sure students understand these words:  
portfolios, presentations, participation, to get marked, to get a mark, fair, opportunity
6. The questionnaire should be fairly self-explanatory. Perhaps in lower level classes go over the first one together. Please explain that they are to answer each question for each of the four types of assessment. They should not fill in the blanks.
7. For the questions at the end, I would like the students to write what they like and/or dislike about the particular type of assessment.
8. Please have one of your students collect the completed questionnaires, put them in the questionnaire envelope, seal it and take it (together with the consent letter envelope) to the office to give to (teacher's name) after class. I have written the class level (not the class name) on the envelopes so I can see if the language level has an effect on how students answer.

If you have any questions, please ask (teacher's name). She has kindly agreed to be my contact person for this.

Thank you VERY much for helping me with this. It should be very interesting to see what the students think about these types of assessment.

## Appendix F

## Letters of Informed Consent

(date)

Dear (name of institution) student,

I am a (name of institution) teacher who has taken some time off to do my Master's degree in Second Language Education at the University of Toronto. I am writing a paper on **Adult ESL Student Attitudes to Different Kinds of Assessment** that teachers use in giving you a mark. For my research, I would like to know how you feel about getting marked on tests, portfolios, presentations and participation in your classes.

- Tests-               for example, multiple choice tests, fill in the blank tests
- Portfolios -       collections of your writing which are put in a folder and then evaluated
- Presentations -   you prepare a talk on a topic and present it to your class
- Participation -    you are marked on things like attendance, attitude, cooperation in group work, completing class assignments

You do not have to participate in this study. If you agree to help me by filling out my questionnaire, please sign below and then fill out a questionnaire. Please put your name on the questionnaire. **I am only asking you to write your name in case I would like to ask you for an interview. I will not use your name in my research or show your answers to your teacher. Your answers will NOT affect your grades.** You can withdraw from the study at any time without consequence. The questionnaires will be put in an envelope and then a student in your class will take them to the office. Thank you for your time.

Sincerely,

Lindsay Brooks

I have read this letter and agree to fill out the questionnaire.

\_\_\_\_\_  
Your signature

(date)

Dear Teachers,

For my thesis, I am writing a paper on "Adult ESL Student Attitudes to Different Kinds of Assessment". For my research, I would like to know how you think **students** feel about getting marked on tests, portfolios, presentations and participation in their classes.

- Tests - for example, multiple choice tests, fill in the blank tests
- Portfolios - collections of student writing which are put in a folder and then evaluated
- Presentations - a student prepares a talk on a topic and presents it to the class
- Participation - students are marked on things like attendance, attitude, cooperation in group work, completing class assignments

You do not have to participate in this study. If you agree to help me by filling out my questionnaire, please sign below and then fill out a questionnaire. Please put your name on the questionnaire. **I am only asking you to write your name in case I would like to ask you for an interview. I will not use your name in my research or show your answers to anyone else.** You can withdraw from the study at any time without consequence. You have been provided with an envelope in which to put your questionnaire before handing it in at the office. Thank you very much for your time.

Sincerely,

Lindsay Brooks

I have read this letter and agree to fill out the questionnaire.

\_\_\_\_\_  
Your signature

**Appendix G**  
**Letters Requesting an Interview**

(date)

Dear (Student Name),

Thank you very much for completing the questionnaire on “Adult ESL Student Attitudes to Performance-based Assessment”. I have looked over the questionnaires and would like to do some interviews with some students. The answers on your questionnaire are very interesting and I would like to talk to you about your ideas. Would you be available to meet with me for about half an hour on **(date) at (time) in (room number)?** **You do not have to be interviewed if you do not want to.** Please put a check (✓) beside your answer.

- I would like to be interviewed. I will see you (proposed time)
- I would like to be interviewed but I'm busy at the time you suggested. How about \_\_\_\_\_  
(suggest another day and time)
- Sorry, I do not want to be interviewed.

Please put this letter in the envelope and give the envelope to your teacher.

Thank you,

Lindsay Brooks



(date)

Dear (Teacher's Name),

Thank you very much for completing the questionnaire on "Adult ESL Student Attitudes to Performance-based Assessment". I have looked over the questionnaires and would like to do some interviews with some teachers. I would like to explore in more detail your thoughts on performance-based assessment and tests. Would you be available to meet with me for about half an hour on \_\_\_\_\_  
date at time in room? **You do not have to be interviewed if you do not want to.** Please put a check (✓) beside your answer.

- Sorry, I do not want to be interviewed.
- I would like to be interviewed. I will see you [on date, at time, in room]
- I would like to be interviewed but I'm busy at the time you suggested.

Please put this letter in the envelope and return the envelope to my temporary mailbox.

Thank you,

Lindsay Brooks

## Appendix H

## Letters of Informed Consent for the Interviews

(date)

Dear (Student name),

Thank you for agreeing to be interviewed about your perceptions of student attitudes about tests, portfolios, presentations and participation. The interview will take no more than half an hour and you can leave the interview at any time or refuse to answer any of the questions. Your answers will be kept confidential. Only I will know what you said. I will not talk about your answers with anyone else at (name of institution). Your name will not appear anywhere in my written report. You can withdraw from the study at any time without consequence.

If you agree, I would like to tape record our interview so I can analyze what you said later on when I am writing my paper. Nobody except me will ever hear the tapes.

Thank you very much for participating!

Sincerely,

Lindsay Brooks

---

 (date)

I \_\_\_\_\_ agree to be interviewed by Lindsay Brooks about my  
 (please print your name)  
 perceptions of student attitudes to different types of assessment. I understand that  
 the purpose of the interview is purely for research purposes and that:

- my name will be kept confidential
- my answers will remain confidential
- nobody else will know what I said
- no one else besides the researcher will listen to the tape recorded interview
- I can withdraw from the study at any time without consequence

Signed,

---

 (signature)

(date)

Dear (Teacher name),

Thank you for agreeing to be interviewed about your perceptions of student attitudes about tests, portfolios, presentations and participation. The interview will take no more than half an hour and you can leave the interview at any time or refuse to answer any of the questions. Your answers will be kept confidential. Only I will know what you said. I will not talk about your answers with anyone else at the ELI. Your name will not appear anywhere in my written report. You can withdraw from the study at any time without consequence.

If you agree, I would like to tape record our interview so I can analyze what you said later on when I am writing my paper. Nobody except me will ever hear the tapes.

Thank you very much for participating!

Sincerely,

Lindsay Brooks

---

December 11, 1998

I \_\_\_\_\_ agree to be interviewed by Lindsay Brooks about my  
 (please print your name)  
 perceptions of student attitudes to different types of assessment. I understand that

the purpose of the interview is purely for research purposes and that:

- my name will be kept confidential
- my answers will remain confidential
- nobody else will know what I said
- no one else besides the researcher will listen to the tape recorded interview
- I can withdraw from the study at any time without consequence

Signed,

---

(signature)

## Appendix I

## Typical Comments for Each of the Codes for the Four Types of Assessment

Table I1

Tests - Typical Comments Reflective of the Codes**Motivation**

I like about getting marked on tests because it can push me to study hard.

Tests can push us to study English. Students usually need some pressure to study.

I think test is a way to push me to learn hard but the result may not mean anything.

**Feedback to Students**

I like about getting marked on tests because it help me and my teacher which I did progress or improve.

Is good because I can to be feedback.

The marks of tests show me how much understand what I learned and what I need to review.

**Demonstration of Ability**

I liked because prove what I really know but it doesn't really teach me.

I like tests because really you can show what you know.

I like that because it show to me my ability.

**Not Reflective of Ability**

I think that do the test don't show our total skills.

Most of the time a test doesn't reflect what you really know because during a test there are a lot of factors that can influence your grade.

Tests are quite important but in my opinion a test doesn't show at all what you know.

---

table continues

Table II continued

---

**Assessment Format or Type**

I like tests of vocabulary, which is multiple choice, but I don't like reading test.

I like tests when they have different kinds of questions and before to take the exam the teacher explain how to solve it.

I like getting marked on grammar, sentence structure and vocabulary. But at the same time, I don't think that vocabulary can show my ability in English. I can learn by heart new words without making too many efforts.

**Opportunity to Learn**

I like it because obligated me to study more and in this way I can learn more.

Tests are necessary to control and check the student's understanding. If I have a test, I learn for it, I am pushed.

I like getting marked on tests toward how much I have learned, because in the process or preparing for the test, I can repeat and make sure what I learned.

**Negative Affective Response**

I don't like getting marked on tests, because low scores depress me.

I dislike about getting marked on test because when I have test, I always feel nervous, nervous make me got low marked.

Sometimes I feel a lot of pressure during the test because we have to answer it in 1 hour, for example.

---

Table I2

Portfolios - Typical Comments Reflective of the Codes**Opportunity to Learn**

I like getting marked on portfolios because I can more learn my English.

I like the portfolios, because help me to improve my English, sometimes it's difficult for me, but I like it.

I like get marked on portfolios because it help me to develop my skills in essays or paragraphs. It's very important to improve my vocabulary.

**Demonstration of Ability**

That's a good way to know how students show their thinking or writing skills.

I think on portfolio, my real ability will emerge.

I like portfolios. It shows what I have learned and how much I have improved.

**Feedback to Students**

I like about getting marked on portfolios because that can help me find out where is my wrong.

I like portfolios because if I have some wrongs on writing, my teachers can correct me.

I like about getting marked on portfolios because I have more opportunities to revise or organize my tasks and can get more comments from teachers.

**Concerns about Time Involvement**

I don't like portfolios because it needs too much time.

I don't like portfolios because it takes much time.

Portfolios are a good way to mark on. But we usually need to spend too much time to finish my portfolio.

---

table continues

Table I2 continued

---

**Assessment Format or Type**

When the topic is interesting is very good and enjoyable to write about it.

It is very practical if I want to improve in writing. However, it must be proper level. If teacher choose too difficult topic for writing, it's nothing but stressing.

Like: have chance to write different kinds of things.

**Assessment Resulting in a Product**

I like to review my writings which I wrote in class, so the collection of them will be a memorial object.

I like getting marked on portfolios, because I can feel the accomplishment. It's nice to read some portfolios which I wrote in some former sessions.

After the course, there is something who rest.

**Assessment Involving a Process**

I like very much the idea to do portfolios because you can see at the end a picture about your learning.

I like it because make portfolio is a process.

I like that the work I have done for the session can be improved when presenting the portfolio. It reflects the process I've been passing through along the course.

---

Table I3

Presentations - Typical Comments Reflective of the Codes

---

**Opportunity to Learn**

I think the presentations are very good tool for learn.

I like doing presentation because I have to understand many vocabulary and sentences. We can learn a lot in the presentations.

Presentations makes our speaking skills develop. Also, we can get additional knowledge through preparing presentations.

**Demonstration of Ability**

I like the presentations because ...I have the opportunity to show my little ability.

In my opinion, presentations are the best thing to show our English ability, because I think that presentation is Western way that how to express ourselves.

I think on presentations, my real ability will emerge, too, so I don't like it. I can't cheat like on tests!

**Negative Affective Response**

I don't like presentations because I'm very nervous while I'm talking in front of public.

Dislike. That makes me feel stress and not so comfortable when I'm in presentations.

I don't like it so much because I feel nervous when I have to talk in front of the class.

**Provides Practice**

I think it's good to practice English in front of people.

I like the presentation because I can practice my English.

Presentations are good to let us practice and organize our ideas on a precise topic.

---

table continues



Table I3 continued

---

**Concerns About Marking**

When teachers evaluate students, they must have some criteria for fairness.

Sometimes it's very difficult to mark for the teacher, all the presentations are so different.

Teachers have to be very careful when they are marking a presentation.

**Concerns About Time Involvement**

Presentations take too much time but they are the best way to improve my English.

It takes too much time to prepare.

A negative aspect is that the preparation requires too much time.

---

Table I4

Participation - Typical Comments Reflective of the Codes**Opportunity to Learn**

I like getting marked on participation because I can improve my English.

For me this is the best way for to learn, you have to participate all time for practice your English and achieve more knowledge.

I believe it is a way to help me learn more. If I participate in every classes, I think at least I can learn something from teacher and it is also a way to improve my listening comprehension.

**Concerns About Marking**

It should not to use for mark. Because he/she should decide by himself to go classes or not. We are not children.

We cannot mark a student is good or not good depend in his participation.

It doesn't really show how is my participation. What is the difference between 8.5 or 8.6?

**Provides Practice**

I think this is a point more important because all the time, you practice your knowledge about language English.

I like it because that helps you to practice your conversation skills.

Good. I can practice my speaking.

**Demonstration of Ability**

I like participation because it shows my abilities for speak and use vocabulary.

Participation is the best way to mark because the people show their knowledge and nothing is memorize at that moment.

Like because is showing how are you doing all days.

---

table continues

Table I4 continued

---

**Not Reflective of Ability**

I don't like this way for getting marked on because it is not really relation with students' English skills and participation.

Participation never shows any ability.

I disagree that participation defined someone English ability is marked for students.

**Motivation**

Participations who motivation for learning English. I think it is the base when I learn something.

I think marking from participation encourage people to attend the class and it's really (how can I say?) legitimate.

Getting marked on it makes students study, I think.

---

## Appendix J

### Student Definitions of a Test

---

1. I hope it's a way to see if we understand something in the class.... A good test, I think, should include everything for the comprehension of the course; we have to write something. We don't have a choice and to see if we understood the correction we had to make a test. Everything should be on the test so it's a little bit long.
2. It means reproducing learned facts or memorized facts. ... Yeah, well, if you say tests, it makes me think of like vocabulary tests or word tests. Uh, it's more if you talk of exams. It means transferring knowledge. Transferring previously acquired knowledge and methods.
3. Test? It's a silly things that teachers would tell you when you're going to have the test and you cannot reach and then you have a mark on it. They check all the answer that you did and the important is they give you the mark.
4. It is like, how could I say, it's like front line in a war. It's really like this. It's the last point. You'll be killed or you'll be rescued because I mean, due to my experiences, it's like this.
5. I had many tests, tests was just fill in the blanks and make a multiple choice.
6. To remind about our knowledge. About what we have learned; that's a test.
7. I can do portfolio and presentation the topic what I'm interested in or what I like to do but in the test, I can't do what I like to do. ... Because teacher gave us tests, we have to do. I don't have any choice or freedom.
8. When we take test, we have to write the answers on the paper. That's it.
9. Yes, it is very popular because for every student, for every student, every student like this. It is take a small time and they don't need any writing, just tick, tick.
10. Test always make me nervous. Make me be nervous. ... I want to get good marks but always study just a little so worry about before test day, I'm so nervous and worried. I have to do, I have to do study, study but just thinking.
11. Test. Paper. Paper test and yeah, it's very hard.

---

table continues

- 
12. Test means. I think tests – there is two kinds of tests. One is choose answer and other one is like describe and write so it's very different between these two...Multiple choice tests is not so important because I can guess and choose but other hand, writing test or something like that, at least can evaluate their ability, real ability, so I think that's more important...I think it's a good chance to see my ability.
  13. Terrible ... and it also make a lot of point during your class and in your program. Almost just the teacher give you a paper, lots of question and you need check, you write your opinion or you have to check, write anything, something like that, paper test.
  14. Exam. Teacher want, teacher wants to know student, student knowledge, then test.
  15. If somebody says test to Korean, they usually get much stress about that because we had a lot of tests. Even though they just say, it's just a quiz, don't give it too much stress, but we still get stressed. ... If somebody says tests, it's stress, believe me.
  16. For me, test is like a tools that the teacher have to know if his or her students are working in the process, are learning in accordance with the parameters...[a test] is like a parameter or measure that the student know I have to work more about any topic.
  17. Test mean scores, all scores, decide by test.
  18. It is a way for the teacher to know if you know or don't something but is write always. So is the way that he knows that you learned something; it is always writing.
-

## Appendix K

### Student Definitions of a Portfolio

---

1. As [the teacher] described, we have to choose five our best writing and to improve them, to correct them and to make a letter to present why we chose those items.
2. Well, you choose some of your works and put together, put them in a nice form and hand them in.
3. Collection of writings. ... [The teacher] told us to choose five of them, the one you like and then rewrite it and type it and put it in the portfolio.
4. It is clear. For me, life is like a portfolio in which we have it after. I mean, I have 19 years of life which can be a portfolio, what I've done in my life. It's like this and it's, that's more practical than tests.
5. What is a portfolio? Just giving my writing in a file and sometimes teacher check it and correct my writing and teacher put back it for us and then we make a new writings according to teacher's corrections.
6. Portfolio is a lot of thing, a lot of what can I say? Items. ... Little by little you class your knowledge for to see your degree or progress. ... But the important, why we do the portfolio is for to see where we are, where we was and where we are now.
7. I don't know if it's general meaning of portfolio or not but what I did, was first I choose the topic and that I was interested in and I read it. ... First we have to read and later we have to summarize it and we have to check new vocabulary and list it and we have to write some reflection about that article. That's what I did. ... The teacher marked some detail, every detail about portfolio and finally marked the whole.
8. I have to rewrite my writing before I wrote. I have to add some more details.
9. It is a writing again or which we finish in our class before. We write that, review, review, we just review again. ... Give to the teacher again and teacher just look how we improve our writing.
10. Biography. Biography and hometown and a formal letter, informal letter and describe of a traditional Korean food. Yeah, and then describe my room, my house, many many things. ... Then I keep.

---

table continues

- 
11. Portfolio. This is the process. I don't think this is the test. Just only process and we have, we show the teacher our process for what we did in the class and show our knowledge, our effort. ... Each time we handed the essay to teacher, then teacher commented, had a comment on my essay or correct my essays. Then so – in portfolio, I corrected my mistakes or errors, then I made more information, I added more information in my essays, then in portfolio, I handed it to good essays.
  12. Portfolio. Actually I don't know the real meaning of portfolio. This time I handed to my teacher, for example, writing teacher and then I can my all writing papers, my all journals, and just set it and give it to them. But before I hand it to them, I need to check again and correct, correct more good portfolio.
  13. The teacher give you a topic and you write down your journal or something. Sometimes it's an interview. Maybe interview with somebody. Maybe in school, or I mean, outside, on the street and you write down your opinion, make, make, write down all of your answers, all of your opinions, and then you give it to the teacher maybe at the end of the program. ... Maybe one time for the portfolio so you have a lots of time to prepare if you think back my opinion I done before very stupid, now I change, something like that. You will have chance to write down my new idea so we don't feel lots of pressure about the portfolio.
  14. Portfolio mean is collect. When I, when I write, write everything I collect, I collect and find, means portfolio.
  15. Before I came here I thought portfolio is a kind of like a for the model, so when they were get a job, they prepare a lot of their own picture and then hand it out to the - ... That is what I know is a portfolio. [Teacher's name] she explained that we should choose the article and then we should talk on what we learned during the class time, and then also write about new word I learned from the article.
  16. Portfolio, this is a good and new assignment, new job, new work in Canada. Portfolio is a way because the student can show their creativity...not only creativity, you show or talk about their feelings, their likes. ... But it's more writing. Portfolio is writing class.
  17. Maybe in one semester, you have to write many report paper or article and at the end of semester you collect all the article and put it together to make a book.
  18. ... I don't know if I got it because [the portfolio] was like a it's kind of presentation but you write it, it's a lot of questions, open questions that you have to write it and make a big package of papers with a lot of information.
-

## Appendix L

### Student Definitions of a Presentation

---

1. It's just to express ourselves the best as we can 'cause we are learning a second language so it's not the same as for example, in microbiology, we had to be precise, lots of terms and here it's just to see if we are able to speak fluently in English, I think. I hope so anyway. And the topic is maybe is less important than the way we speak.
2. Presentation is an interesting situation because you stand in front of the class and you have a sort of auditory and that's the one side. The other side is the preparation for presentation which takes quite a while but since it's about something, since the presentation should be on a topic you are interested in, it's okay. And the preparation.
3. Well, my definition of presentation is the teacher give us a topic or we choose a topic and then gives a limited time like 10 minutes or 20 minutes and focus on your topic and then you express your ideas and according to the resources that you find.
4. Presentation, it's like a speech... Most of students in high school didn't like that because they had to go to libraries, they had to find some, to do some researches, find some information and do presentation they sometimes feel shy.
5. Presentation, uh, roughly make a some, research and understand it and maybe I need some analyze the data which I get. I, firstly, I have to make a summary for, to, for audience to understand and usually I was in Electronic Engineering department so uh, make a transparency.
6. Presentation. For example, about one items, I don't know social security in Geneva for example. ... We must explain what, how the social system work, why exists, what we need. ... Explain to a group or to teacher.
7. To show one's thoughts or one's study to another person, to another people or – presentation, to show what the person did. ... I have to collect some article or some source and then organize it, so that I can give speech what I want, and then to presentation.
8. I talk about the study, in front of classmates.

---

table continues



- 
9. **Presentation.** Our teacher select every group, every group one, for every group one subject. Our group, example, our group have community centre project, another group have medical doctor project, another group have RCMP project, and another have [university name] radio, radio project another group have [university name] newspaper project. ... Yes, many, many project and we have community centre project and it is firstly, or in, we, our work, our first work is going to the community centre and our teacher select name from the community centre, person, a person name, give us a person name. We went there and took his interview and we write his interview and take many information, take many guide from him and know about many things from the community centre and take some picture from, picture from the community centre and we took picture from the interviewee and we came to our home and take our information, lot of information. Now we, we our teacher gave us big board for making, for make big poster with our own, our own thinking and we, and we cut many information, not writing only. We cut many picture with point information exactly and we write down many information in the bottom of the picture. And after finishing our poster, we, our teacher give us, every student give 15 minutes, and we talk about that. We showed our poster to another student and we talk, talk about 15 minute about this poster and this community centre and we give them, we'll give them many information about the community centre. About the poster with picture this how this picture, what is the, who is,...we cannot explain this. In this picture, which subject we use this picture or this subject, we use this picture.
  10. I'm just first class and last class I never did, never do that but second classes sometimes our teacher gives us some assignment for writing and then that day after day ... the teacher appointed the students, she said please come up ... come up and then okay, she want to read, or she want to talk to my thinking of work, thinking of work and then talked to other students.
  11. Presentation is to speak in front of everyone. Then yeah, both [portfolios and presentations] are same process to survey to study harder before these. ... presentation needs us knowledge, and presentation skills.
  12. Presentation is the chance, the chance to explain what I'm thinking or what I am, I was studying. And the chance to convince the people. ... I have to research in the Internet or some books and actually I worked with my partner. ... We have to work together and especially my partner was not Japanese so I have to speak English with her and I have to learn, learn how she think about that.
  13. I think presentation it's a, you have to show, to show your idea, but not by writing but your work, your speaking and your body language and the topic, I mean the topic of your talking, maybe the paper, the paper maybe come from yours, maybe come from the other. That's why I think the presentation it's that kind of the show how you professional or your opinion or about your thinking about a topic.

- 
14. Presentation in class or seminar, I express my prepare one subject. ... But here ... I should prepare, I mean the focus in on the speaking not how I well prepare so it's different [from my previous experience].
  15. Choose the subject and prepare, and prepare for that about topic and I don't know and then give others information about the topic that I choose.
  16. Presentation is a, like a, about any topics using tools, for example blackboard or computer or television or anything but before I have to prepare that topic, summary. Talk about that and after that discuss about that ... For example in presentation, you have to do a presentation about free topic. You can find in the newspaper one topic and prepare it, you prepare that topic and you're going to do, you can do presentation. It's very good because that things improve our skills.
  17. Presentation is can show your opinion but you have to prepare before the presentation maybe for a long time. And then you have to collecting many datas.
  18. A presentation. I think that's the ability to say what you know but is more, you don't have to repeat it in the same way that, that was in the book or in the class. You change it with what you think, what you know in, it's like your words. Say the things but in your words so people can understand so it's not the same as in the book whatever you have to say.
-

## Appendix M

### Student Definitions of Participation

---

1. ... Usually I thought it was to see if we, of course, if we are active listening but active speaker, to answer the questions the teacher asked.
2. It's the quantity and the quality. How often you say something. How often you contribute to the, the learning process in class counts as well as what you say.... It's active participation. It's not attendance.
3. Attendance. ... Some discussion they required the participation mark but very few as I said. ... I think here teacher require us to answer the question, to have the more response and to know if you can understand the question or not. ... Attendance and the times that you answer the questions or you drop out your ideas.
4. Some people think that it's just doing homework and taking part in classes but personally, I don't agree. ... Participation is not necessarily doing homeworks or just taking part in class. ... Everyone can talk about tests, but what is exactly participation? ... I believe participation is a combination of everything. Physically you have to be present in the class. It's a part of participation. You have to have chance to discuss with other students about the topic or about the textbook. You have to do your homeworks. ... You have to, do researches maybe. You have to show them that, I mean, teachers have to encourage students to do, to act in a way that teachers can understand that people like the class and like what they're doing. That's participation for me.
5. Just presence in the class or some teacher require the, requires the how do I join the class. How much do I talk to the teacher or make a questions.
6. First to be present. After to be wake up in class, to respond, to help maybe teacher or the other student...I think the class no depend only the teacher. If we want a nice class, we must to do something too.
7. I don't know exactly because in Japan, you know, participation is regarded as kind of obligation so almost every student attended the class so maybe teachers couldn't give any mark according to attendance I think. ... Of course, attitude is included, included in the participation I think. ... 60 % attendance and 40% attitude.
8. Participation is joining the class, do homeworks, and I don't know any more.
9. I cannot understand. Just it is talking, participate any sports or talking about, or introduce another student, talking to student or more student and how can they talk, what is their talking, is perfect or non perfect. I think this is participation.

---

table continues

- 
10. I don't know exactly but I think part activity or class activity. Yes, that's it. I trying to activity, I will join.
  11. Participation is attendance. Attendance to participate in the class. Then to help each student in the group. Then to make some group ... group activity. So how much.
  12. Participation is that, I think it's basic, it's basic thing to evaluate. ... It shows students' motivation or how they, how they, how much they want to study. I think it's very important. ... Maybe it's only just sit in the class and just hear, it's not participation, just try to catch something of the teacher or maybe student need to ready to participate, maybe doing homework or at least read what they are going to do, to study.
  13. I didn't have ... a stronger feeling about this word, participation. That was my first time to know, to learn about this word at the [name of institution]. ... I think because in a class, you have to work together. I mean it's a teamwork. ... So if you want to help, do yourself a favor, or help yourself, you have to participation in the class. I mean you have to pay attention, listen carefully to the teacher what you did, what did the teacher say today and about the teacher's opinion, ... what does the teacher want us to do. You have to do what the teacher ask you to do.
  14. In class teacher question, some student is talk but ah, yes.
  15. Most of the meaning is attendance. And also another is assignment. How they well prepare for their assignment.
  16. Participation is like working together, working, team-working. And sometimes the leader is the teacher.
  17. Participation don't be absent, don't be skipped class. Participation of how do you write your homework.
  18. That he has to be in class, not sleep in class. You have to ask questions to see if he understand, ask teacher, say things, I think that is that; that you're active in class.
-