

The Canadian Academic English Language (CAEL) Assessment

Test Score and Users' Guide

The CAEL Assessment Testing Office
Carleton University
228 Paterson Hall
1125 Colonel By Drive
Ottawa, Ontario Canada K1S 5B6

www.cael.ca

The Canadian Academic English Language (CAEL) Assessment Test Score and Users' Guide has been prepared for deans, registrars, graduate supervisors, administrators, program coordinators, student advisors, and teachers who are using CAEL test scores for selection or placement purposes. The *Guide* is designed to provide information about the CAEL Assessment which will be useful in score interpretation. Additional information about the CAEL Assessment is available from our website or from the CAEL Assessment Office, Carleton University.

The CAEL Assessment Testing Office
Carleton University
228 Paterson Hall
1125 Colonel By Drive
Ottawa, Ontario Canada K1S 5B6

www.cael.ca

The Canadian Academic English Language (CAEL) Assessment *Test Score and Users' Guide* is owned exclusively by Carleton University. The CAEL Assessment *Test Score and Users' Guide* and all related materials are fully protected by copyright from any and all unauthorized use or duplication, and they remain the exclusive property of Carleton University. These materials may not be photocopied or otherwise reproduced by any means or adapted without the express permission of Carleton University.

TABLE OF CONTENTS

PAGE

Acknowledgments	6
List of Tables	7
The CAEL Philosophy of Language Testing	8
• Chapter One:	9-23
OVERVIEW	9
Test Audience and Purpose	9
How does the CAEL Assessment compare with other standardized tests of English?	10
Background Regarding the CAEL Assessment:	11
Test Development	
The CAEL Assessment: Past	11
Present	12
Future	12
Administration	13
Procedures	13
Stage One: Registration	13
Stage Two: The Oral Language Test (OLT)	13
Stage Three: The CAEL Written Test	14-17
Availability of the CAEL Assessment	17
Where the Test is Offered	17
Identification Requirements	17
The Cost of Taking a CAEL Assessment	17
Measures to Protect Security	18
CAEL Assessment Test Administrators	18
Testing Irregularities	18
CAEL Assessment Score Reports	19
Information on the CAEL Score Report	19
Release of Test Results	19
Additional Score Reports	19
Official Test Reviews	20
Re-writing the CAEL Assessment	20
Test takers with Special Requirements	20
Preparing to Take the CAEL Assessment	20

Frequently Asked Questions About the CAEL Assessment	21-22
Strengths of the CAEL Assessment	22-23
• Chapter Two:	23-50
CAEL ASSESSMENT SAMPLE TESTS	24
The CAEL Assessment Sample Written Test	25-40
Appendix 1 – Answer Key	41-45
Appendix 2 – Transcript of Lecture	46
Appendix 3 – Writing Sample	47
The CAEL Assessment Sample Oral Language Test	48-50
• Chapter Three:	51-56
CHARACTERISTICS OF CAEL TEST TAKERS	51
Introduction	51
Gender, Age and Student Status	51-53
First Language	53
Country and Region of Origin	53
Intended Area of Study	54
Length of Time in Canada	55
Summary	56
• Chapter Four:	56-70
SCORING CRITERIA, METHODS AND RELIABILITY	57
Introduction	57
Writing Performance	58-61
Listening Performance	61-63
Reading Performance	63-64
Speaking Performance	64-65
Overall Results	66-67
Comparability of Test Versions	67-70
• Chapter Five:	70-74
SCORE DISTRIBUTIONS	71
Introduction	71
Distribution of Overall Results	71
Distribution of Writing Performance Scores	72
Distribution of Listening Performance Scores	72
Distribution of Reading Performance Scores	73
Distribution of Speaking Performance Scores	73
Central Tendency and Variance of Score Distributions	74

• Chapter Six:	79-83
VALIDITY	75
Introduction	75
Ensuring Construct Representation	76
Investigating Construct Irrelevant Variance	77
Gathering Criterion Related Evidence of Validity	78-79
Consequences of Test Use	79
Test Taker Feedback: Post-test questionnaire	80-81
Test Impact Over Time	81-83
• REFERENCES	87

TEST SCORE AND USERS' GUIDE

The Canadian Academic English Language Assessment

The Canadian Academic English Language Assessment (CAEL) is a standardized test of English in use for academic purposes. It is designed to describe the level of English language of test takers planning to study in English-medium colleges and universities.

The CAEL Assessment tests ability to *use* English as it is used in Canadian universities and colleges. It allows test takers to experience what it is like to participate in a Canadian post secondary classroom.

The CAEL Assessment offers a unique alternative to other standardized tests in that it is integrated and topic-based, testing language in use for academic purposes within the narrow domain of college and university classroom study.

List of Tables	Page
Table 1.1: CAEL Assessment and TOEFL Score Comparability	20
Table 3.1: Descriptive Statistics for Age (Years) of CAEL Test takers	52
Table 3.2: Frequency Distribution of Ages (Years) of CAEL Test takers	52
Table 3.3: Test takers Most Commonly Reported First Languages	53
Table 3.4: Test takers Most Commonly Reported Countries of Origin	54
Table 4.1: Writing Performance Band Score Criteria	58
Table 4.2: Writing Performance Band Score Changes	60
Table 4.3: Listening Performance Band Score Criteria	61
Table 4.4: Reading Performance Band Score Criteria	62
Table 4.5: Speaking Performance Band Score Criteria	64
Table 4.6: Overall Result Band Score Criteria	65
Table 4.7: Overall Performance Band Score Changes	66
Table 4.8: Frequency (%) of Overall Result Band Scores for Four Test Versions	67
Table 4.9: Frequency of Band Score Changes in Overall Result	68
Table 5.1: Overall Results for Full Sample and by Gender	71
Table 5.2: Writing Performance Band Scores for Full Sample and by Gender	72
Table 5.3: Listening Performance Scores for Full Sample and by Gender	73
Table 5.4: Reading Performance for Full Sample and by Gender	74
Table 5.5: Speaking Performance Scores for Full Sample and by Gender	75
Table 5.6: Means & Standard Deviations for Sub-tests and Overall Results	76
Table 6.1: Grade Point Average and Letter Grade Equivalents	80
Table 6.2: Frequency Distribution of Grade Point Averages	81
Table 6.3: Agreement between CAEL Assessment Placement and Teacher Evaluations	84

The CAEL Assessment Philosophy of Language Testing

The CAEL Assessment's *Philosophy of Language Testing* introduces the *CAEL Assessment Test Score and Users' Guide* in order to affirm our commitment to quality, fairness, and principles of ethics and responsibility in language testing.

All of us who are associated with the CAEL Assessment will:

- continue to relate language testing to language teaching and language use;
- situate CAEL test development within current theoretical understandings of language performance;
- utilize the insights and practical experiences of teachers and learners in validating inferences drawn from test performance;
- provide every opportunity, given the constraints of the testing setting, to allow test takers to perform at their highest level of ability;
- regard *fairness* to be the primary consideration in test design;
- take every measure possible to involve test takers in the testing process by incorporating their feedback in test development decisions and valuing their contributions;
- provide test takers with as complete information about the test and testing procedures as possible;
- remain committed to the view that validation of test inferences is on-going and that arguments for the validity of inferences drawn from the use of the CAEL Assessment depend on the accumulation of evidence arising from the use of the test.

CHAPTER ONE

OVERVIEW:

The Canadian Academic English Language (CAEL) Assessment -- Test Audience and Purpose

The purpose of the CAEL Assessment is to test English in use for academic purposes (EAP). The CAEL Assessment is designed to describe the level of English language of test takers planning to study in English-medium colleges and universities.

The CAEL Assessment is a criterion-referenced, topic-based performance test, comprised of an integrated set of language activities. The language tasks and activities in the CAEL Assessment are systematically sampled from those that are commonly undertaken within the university academic community. The content for the tasks on the CAEL Assessment is drawn from introductory university courses at times when professors are introducing new topics to their students with the expectation that the students know little or nothing about the content. The test is comprised of representative tasks and performances that characterize academic study, for example:

- speaking about academic experience, information or understanding,
- listening to, taking notes, and transferring or applying information on a topic introduced or extended by an academic lecture,
- reading and selectively applying information from academic articles and texts about a topic introduced or extended by a lecture, and
- incorporating what has been learned from the lecture and readings in writing a formal, academic response to an academic task.

The CAEL Assessment was developed in 1987 at Carleton University in response to requests from the faculty and a motion of the University Senate to provide an alternative to other standardized tests. The CAEL Assessment was designed to more effectively identify students who could participate actively in academic study. Since its first use in 1989, test takers in many parts of the world have taken the CAEL Assessment as part of the process of admission to universities and colleges in Canada, Europe, and the United States.

How does the CAEL Assessment compare with other standardized tests of English?

In order to understand the distinctive nature of the CAEL Assessment it is useful to compare it with other standardized tests of English. The CAEL Assessment is *context specific*, testing language in use for academic purposes within the narrow domain of college and university study. Other standardized tests of English are *context general*. They test global language proficiency.

Global language proficiency is not necessarily relevant to the purposes for which English is used in an educational setting.

The CAEL Assessment reflects the *range and variety of responses* essential for academic study. The test identifies *reading, listening, writing and speaking for academic purposes* as key requirements of effective participation in a college or university setting.

The CAEL Assessment includes instances of test taker performance in, for example:

- extensive *reading* of a range of academic texts, for synthesis of key ideas, note-taking, paraphrasing, and responding in writing to questions that arise from the information;
- *listening* to lectures, note-taking, synthesizing the information and responding;
- *academic writing* across modes, (i.e., personal and argumentative), in speeded and non-speeded situations, and involving both extended and short answer responses;
- *speaking for academic purposes* across tasks (i.e., prepared and spontaneous), in speeded and non-speeded situations, involving both extended and short answer responses.

Other standardized tests do not test writing or speaking in any extensive way, and they rely mainly on multiple choice and other objective test formats.

Because the CAEL Assessment is *fully integrated and topic-based*, the test locates test taker performance on the test within meaningful and coherent content. CAEL test takers must position themselves within the content of the test in much the same way that students must situate themselves within the content of a course. Other standardized tests generally rely on discrete, independent items which are chance and disconnected.

The CAEL Assessment uses a range of response formats and tasks *selectively sampled from first-year university courses*. Other standardized tests depend mainly on multiple choice or objective response formats which have only limited use at the university or college level.

The CAEL Assessment provides *individualized, criterion-referenced profiles* of relative performance in academic listening, reading, writing and speaking as well as an overall score. This increases the meaningfulness and interpretability of the results for both test takers and test users. Other standardized tests generally report only a numeric score or series of scores, limiting both the meaningfulness and interpretability of the result.

The CAEL Assessment effectively identifies test takers who have language proficiency adequate for academic study. A full discussion of the validity of the inferences drawn from CAEL Assessment results is presented in the chapter on Validity, pp.75-83. Other standardized tests, principally because they under-represent (Messick 1989) the construct of English in use for academic purposes, often select test takers who are unable to cope with the demands of academic study. They identify test takers who can read and respond to multiple choice questions regarding English grammar and vocabulary, but who are unable to follow an academic lecture, speak effectively or write at a standard required for academic study.

BACKGROUND REGARDING THE CAEL ASSESSMENT: TEST DEVELOPMENT

CAEL Assessment Past

The initial development of the CAEL Assessment was undertaken in 1987 at Carleton University. At that time there was general dissatisfaction on the part of the faculty with regard to the effectiveness of language proficiency testing. In the view of the faculty, language proficiency screening was failing to identify applicants with necessary levels of English who were able to participate effectively in academic study at the university level. Graduate faculty supervisors were particularly interested in the identification of applicants who could participate in seminars, interact with other students and, in some cases, act as teaching assistants. As a result of a motion of the University Senate, a committee was struck to oversee the development of an alternative test of academic English proficiency which would replace other standardized tests used in the context of university admission. Other commonly used standardized tests had as their purpose to test *global* or *unsituated* language proficiency. The CAEL Assessment was designed specifically as a test of English in use for academic purposes, with the objective of more effectively identifying students who had the ability to participate in academic study.

In the initial stages of test development, EAP professors in the School of Linguistics and Applied Language Studies worked with professors in the Faculties of Science, Engineering, Social Sciences and Arts, collecting information about actual language performance requirements in introductory first-year classes. Initial needs analysis found that a number of language tasks and activities were consistent across all faculties (Young 1984, Freedman 1987, Fox 1995).

As a result of the initial needs analysis, CAEL test specifications were defined to relate academic language performance to criterion descriptions. Pilot testing of three initial CAEL versions was undertaken during 1988-89 and the first administration of the CAEL Assessment occurred in November 1989.

CAEL Assessment Present

Since its initial development, the CAEL Assessment has become the preferred test in many university and college admission contexts. It is frequently used in EAP program placement contexts in a number of institutions across Canada, Europe and Asia. In operationalizing the construct of language in use for academic purposes, the CAEL Assessment derives support from current theoretical understandings of language and learning in academic settings, from an intensive and on-going research program supervised by the Language Assessment and Testing Research Unit, which has managed the CAEL Assessment since 1994, and from the positive impact the test has had on ESL/EAP programs and learners. In the CAEL Assessment context, *teaching to the test* does not undermine student learning, rather it supports and complements it. This feature of the CAEL Assessment is discussed further in Chapter Six, which discusses Validity, pp.75-83. Institutions using the CAEL Assessment continue to report high satisfaction with the test. For example, an educational researcher at the World Maritime University in Malmo, Sweden, compared the CAEL Assessment to other standardized tests of English¹, commenting that “it is clear the CAEL was by far the best indicator” of a students' ability to meet

¹ TOEFL, IELTS and Oxford Placement Test.

the language demands of study in an English-medium context (Jackson, World Maritime University, Sweden, 1998).

CAEL Assessments are administered to as many as 10,000 test takers a year in 25 Canadian and international test sites with on-going test administrations across Canada including: Victoria; Vancouver; Edmonton; Calgary; Winnipeg; Hamilton; Toronto; Ottawa; Montreal; Saint John; Wolfville; and St. John's and in such international test sites as Beijing, Bangkok, Buenos Aires, India and Qatar.

CAEL Assessment Future

The CAEL Assessment will continue to accumulate evidence for the validity of the inferences drawn from its use as part of the on-going process of test development in order to fulfill its mandate by providing a reliable alternative standardized assessment of English in use for academic purposes.

ADMINISTRATION

Procedures

The CAEL Assessment is administered in three stages. Stage one involves completion of an online registration form (approximately 30 minutes). Stage two consists of a computer-mediated test of oral language proficiency in an academic context (approximately 20 minutes). Stage three consists of written responses to academic readings, an academic lecture and an essay prompt which asks test takers to agree or disagree with a statement (approximately 2 hours). The information elicited in these three stages is used to create a profile of the test taker's language proficiency. The profile is an essential part of the feedback provided to test takers regarding their strengths and weaknesses in academic English.

Stage One: Registration (approximately 30 minutes)

All test takers must register online prior to completing the second and third stages of the CAEL Assessment. This registration process serves a number of purposes:

- Verification of personal identification
- Collection of relevant background information (e.g., domain of study, admission status, educational background) and
- Distribution of test preparation information

Stage Two: The Oral Language Test (OLT) (approximately 25 minutes)

The OLT is a computer-mediated oral language test of spoken English in use for academic purposes. The OLT may be taken before or after the CAEL written test (see stage 3 below).

It consists of five tasks which sample typified and recurring speech acts within college and university settings:

Task 1

(2 minutes) *To make a short oral presentation.* Test takers are given the topic for the presentation upon registration in stage one and may take as much time as they like to prepare and practice their talk in advance of the test.

Task 2

(5 minutes) *To relay information obtained from a lecture.* Test takers are asked to provide key information to another student based on their understanding of a short tape-recorded segment of a professor's instructions.

Task 3

(5 minutes) *To relay information obtained from an academic document.* Test takers are asked to provide specific information based on their reading of a short, generic academic document.

Task 4

(5 minutes) *To relay information from an academic text.* Test takers are asked to read a section of the text in Task 3 aloud, for the purpose of discussing the text in the context of a group discussion or academic presentation.

Task 5

(8 minutes) *To explain a choice for participation in a group project.* First, test takers listen to a professor's instructions for a group oral presentation. Then they listen to other members of a group who explain their preferences for participation in the presentation.
After listening to the other group members, the test takers are asked to explain their own presentation choice and ask a question regarding an important detail related to the presentation, e.g., time allowed, marking criteria, available equipment, etc.

The scoring for the OLT is conducted by trained raters who listen to the test takers' recorded responses to the five tasks. Raters assign points to each task based on an analytic scoring rubric that focuses their attention on specific features of each task. The points for these tasks are summed to provide an analytic score. Trained raters then make a holistic assessment of the overall oral performance and allocate additional points to the test taker's score. The total possible score for the OLT is 54. Raw scores are converted to criterion-related band scores that range from 10 to 90. To review a copy of an OLT sample please see Chapter Two.

For details regarding the reliability of scoring of sections of the CAEL Assessment see Chapter Four, pp.57-70.

Stage Three: The CAEL Written Test (approximately 2 hours)

The CAEL Written Test is administered to groups of test takers in a university classroom setting. CAEL tests are currently available in a number of versions and typically present academic information on general issues, for example, food, health, environment, technology, globalization, culture, business or weather. At the beginning of a CAEL Assessment test takers are presented with an essay question which asks them to agree or disagree with a statement related to the topic of the test. It is assumed at this point in the test that the test takers know little or nothing about the topic of the essay. During the next two hours of the test, test takers are presented with a lecture, readings and tasks which define and develop the topic of the test.

After completing the listening and reading sections of the test, test takers are given 45 minutes to complete the essay question. Test takers may use the readings as well as their notes from the lecture in writing their essay responses. The readings and lecture supplement one another and provide the basis for the writing task. Test takers position themselves within the parameters of the input provided within the test in much the same way that students do within the parameters of a course. Input not only supports the responses of the test takers, it also constrains and defines the boundaries of their responses.

The following summary of test sections provides a general guide to the organization of the test. At the beginning of the written test, test takers are provided with an outline of the test which indicates the amount of time allowed for each component of the test as well as the number of points for each component.

CAEL Assessment Written Test: Summary of Test Sections

Reading 1	25 minutes	20 points
Listening	30 minutes	35 points
Reading 2	30 minutes	26 points
Writing	45 minutes	Band Score

For the reading sections, test takers are given two readings that are on the same topic as the lecture. The readings are taken from textbooks, newspapers, magazines, government publications, periodicals, etc., in other words, from any reading material that might be used in a first-year university course. Test takers are expected to use each reading in answering the content-related questions in their test booklets. Question formats for the reading sections generally consist of constructed response items. Scoring for the reading sections is similar to the listening section (see below). Scoring is undertaken by trained raters using a marking key. Marks are not lost for errors in spelling or grammar unless they interfere with meaning. The focus in the reading sections is on the test taker's comprehension. Partial credit is awarded when warranted and sample responses are provided in the marking key. Raw scores are converted to criterion related band scores that range from 10 to 90. For information regarding the reliability of scoring of sections of the CAEL Assessment see Chapter Four, pp.57-70.

For the listening (lecture) section, test takers listen to a tape-recorded lecture adapted from a first-year university course. The length of the lecture varies somewhat across the test versions, but is always between 15 and 17 minutes in length. The tape is played only once. Test takers are asked to respond to the content of the lecture while they are listening. Although the most frequently used test method is constructed response, question formats include short answer, filling in the blanks, completing tables, multiple choice, note-taking, transferring information to a flow-chart, etc. Test takers are given 3 minutes prior to the lecture to review the tasks and questions in the listening section. After the lecture, test takers are given 5 minutes to review their answers and make any changes or additions. Test takers are permitted to make notes on the test paper while listening to the lecture in the language of their choice, but their test answers must be in English. Scoring for the listening portion of the CAEL Assessment is undertaken by trained raters using a marking key. Partial credit is permitted on all of the questions and sample responses are provided in the marking key. Marks are not lost for errors in spelling or grammar

unless they interfere with meaning. The focus in the listening section is on the test takers' comprehension. Raw scores are converted to criterion-related band scores that range from 10 to 90. For information regarding the reliability of scoring of sections of the CAEL Assessment see Chapter Four, pp.57-70.

For the writing section, test takers are asked to write a response to an argumentative essay prompt. The essay prompt is provided at the beginning of the test so that test takers can consider the prompt when they are completing the listening and writing sections of the test. Test takers may use their responses to the listening and reading sections of the test to write their essays. They may also use the reading texts as sources of information and are instructed to use quotation marks in order to cite any *verbatim* phrases, sentences or passages from the texts that they include in their essays. The essays are typically one to two pages in length. Test takers are given 45 minutes to complete the essay section. Test takers are encouraged to plan their essays before they begin to write. Additional space is provided in the test booklet for plans and rough notes.

The marking process of the writing component of the CAEL Assessment is undertaken by a group of trained raters (separate from the raters who score the listening and reading sections). These raters are all experienced ESL/EAP instructors and have received formal training as CAEL essay raters. The marking process is collaborative. Raters meet for a standard-setting session at the beginning of each CAEL marking session. After reviewing the lecture and the readings, the raters mark selected benchmark-essays and the first five essays from the new CAEL administration. The raters follow a read-aloud protocol in marking (Fox & Soucy, 1990). One rater reads the essay aloud while each of the raters records his/her impression of the language use, organization and content of the essay. After listening to the essay, raters may ask for the essay to be read aloud again or may read the essay silently. Each rater then explains his/her mark to the group and a discussion follows. At the end of the discussion, the raters reach a consensus and a negotiated mark is assigned to the writing section. After the first five tests are marked - the baseline session - teams of three raters continue to use the negotiated, read-aloud marking protocol to mark the essays.

Writing scores are expressed as band scores and can range from 10 to 90. The writing component of the CAEL assessment is marked in this rigorous manner as this component of the test is seen as the most critical for academic English performance. A considerable quantity of research examining the reliability and validity of this marking protocol has been completed (see Chapters Four and Six for discussion of issues of reliability and validity, respectively).

Overall CAEL Assessment Score

An overall score is also provided for the test takers. This overall score is based, in part, on the scores for each section of the written CAEL Assessment but is not a linear-summation of the component scores. The overall score is determined by the test administrators and the testing coordinator based on all of the information available about the test taker. This includes the reading, listening, and writing scores, the oral score, and evaluation of the personal essay and the self-assessment component which was provided at registration. For information regarding the consistency and dependability of CAEL Assessment overall scores, please see the Chapter on reliability, pp.57-70. Sample tests for the CAEL Assessment (written test) and the Oral Language Test (OLT) are provided in Chapter Two. These are also available on the CAEL Assessment website at <http://www.cael.ca/taker/oral.shtml>

AVAILABILITY OF THE CAEL ASSESSMENT

Where the test is offered

The CAEL Assessment is administered in test centers across Canada and in a number of international centers including sites in Asia, the Middle East, South America and India. CAEL test takers are typically applicants to Canadian universities, although many universities in Europe and the United States have recognized CAEL Assessment scores for admission purposes. A list of universities that accept CAEL Assessment results in the context of university admission is available on the CAEL website.

Identification Requirements

Test takers are required to provide either a valid passport or two pieces of government-issued photo identification when registering online. On the day of the test, test takers must present the same identification as provided when registering and one current passport-size photograph.

The Cost of taking a CAEL Assessment

The fee for the CAEL Assessment is set in accordance with the requirements of the test location. Within Canada, the basic fee for a CAEL Assessment is \$185.00 CAD. This fee is subject to change however, and depending on the requirements of the specific test center, additional fees may apply. Test takers should contact the CAEL Assessment Office for specific information regarding the cost of taking a CAEL Assessment in a particular test location.

Measures to protect test security

Test security is maintained for the CAEL Assessment through a number of procedures. Test takers are required to present photo identification and provide a current photograph of themselves on the day of the test. The photograph is held on file to ensure that the applicant and the test taker are one and the same. All test papers are numbered before the test and are collected immediately following each test administration. Test papers are stored in a secure space and test scores are maintained in a secure database. Official test score reports are issued from the Carleton test center and are marked with an official seal. For testing locations outside of Carleton University, the test materials are sent by courier 6 days before the test date. Completed test booklets are sent back to Carleton University within one day of the test being completed. This procedure ensures that test materials are held in test centers for the minimum amount of time possible. For detailed information about training of administrators, test administration and the storing and handling of test materials, see *The CAEL Assessment Administration Manual*.

CAEL Assessment Test Administrators

Test administrators are trained directly by the CAEL Manager at Carleton University. All CAEL Test administrators are required to participate in a CAEL test administration and training program prior to their first administration of a CAEL Assessment. In addition, a written set of administrators' instructions available in *The CAEL Assessment: Administrators' and Monitors' Manual* is reviewed prior to every test administration.

These written instructions are strictly applied to the administration of the test in all test locations to ensure that the CAEL Assessment is administered in a consistent manner in each test location.

Testing Irregularities

If problems should arise in the course of a test administration, for example, equipment failure, unavoidable disruptions due to noise, a natural disaster, etc., test takers are given the opportunity of rewriting the CAEL Assessment at the earliest opportunity without additional cost.

In addition, if during the test a test taker becomes ill and wishes to cancel the test, he/she can consult with the CAEL Test administrator. Generally, test takers who wish to cancel a test result due to sudden illness are able to write the next available CAEL Assessment without additional cost. In some circumstances the test taker may be required to provide a medical certificate.

CAEL ASSESSMENT SCORE REPORTS

Information on the CAEL Assessment Score Report

The CAEL Assessment is a criterion-referenced test. The test score report provides band scores from 10 to 90 for each of the sub-tests of the CAEL Assessment in academic reading, listening, writing and speaking as well as an overall score. It is important to note that the overall score is not a linear summation of the sub-tests. Rather, each test taker's profile is reviewed by the CAEL Manager in determination of the final score, taking into account all of the information accumulated about the test taker. Additional information about interpretation of the CAEL band scores and performance criteria is provided in the chapter discussing issues of reliability, Chapter Four pp.57-70.

Release of Test Results

CAEL Assessment results are generally available within 15 days of the administration of the test. Official score reports are sent directly to the institutions that the test taker has indicated when they registered online. A score report is also emailed to the test taker.

Additional Score Reports

CAEL Assessment test results are considered most accurate for the period of two years (24 months) from the day of the test administration. During this period, test takers may request that official score reports be sent either to other institutions or directly to the test taker. Generally, when test takers register online, they are allowed to indicate ten institutions to which official score reports should be sent. If a test taker requires an additional score report, s/he is required to pay a supplementary fee of \$10 for each additional score report.

Confidentiality

CAEL Assessment test results are maintained in strictest confidentiality. Official score reports are released to other institutions only by written request of the test taker. To ensure the authenticity of CAEL score reports, institutions are urged to require official copies of CAEL Assessment score reports to be mailed directly from the CAEL Assessment Office.

All information retained in an individual's test file is considered private and confidential. All of those who work with the CAEL Assessment agree to safeguard information from unauthorized disclosure. Any information about an individual test taker or institution will be released only by prior agreement, or with the written consent of the individual or institution.

Official Test Reviews

Test takers who feel that there has been an error in their test result may request an official review of the CAEL Assessment for a cost of \$50.00 CDN. Official test reviews will take up to two weeks to complete and include a re-marking of the test. Results of the review are communicated to the test taker in person, where possible, or in writing.

Re-writing the CAEL Assessment

There are no restrictions on how frequently test takers may write a CAEL Assessment.

Test takers with Special Requirements

Test takers with special requirements are invited to contact the CAEL Office. Every effort is made to accommodate requests for special consideration. However, arrangements must be made in advance of the test administration.

Preparing to take the CAEL Assessment

There are several ways a test taker can prepare for the CAEL Assessment:

- Visit the CAEL Assessment website at www.cael.ca for a sample test. This enables test takers to become more familiar with the format of the test.
- CAEL preparation workshops are available at select testing sites. Test takers can contact the test location in advance for information regarding test preparation workshops. In addition, the *CAEL Assessment Test Taker Preparation Guides* can be purchased. For information, visit the CAEL website.

SOME FREQUENTLY ASKED QUESTIONS ABOUT THE CAEL ASSESSMENT

Question: The CAEL Assessment is a topic-based test. What if a test taker doesn't know anything about the topic?

Answer: It is not necessary to know anything about the topic to be successful in the CAEL Assessment. Everything a test taker needs to know to respond to the readings and lecture or write the essay is contained in the information provided within the test itself. Even if a test taker knows something about the topic of a CAEL Assessment by chance, the information necessary to respond to the writing task is specific to the information provided within the test. Research indicates that a positive predisposition toward the topic of a CAEL Assessment has little if any effect on test taker performance (see the summary of CAEL-related research studies in the Chapter on Validity, pp.75-83.)

Question: How often can a test taker take the CAEL Assessment?

Answer: There are no restrictions on how frequently a test taker can take the CAEL Assessment.

Question: Can a CAEL score be equated to a TOEFL score?

Answer: The CAEL Assessment is criterion-referenced and most of the response formats on the CAEL Assessment require constructed responses (i.e., open-ended, short or extended responses) as opposed to TOEFL which relies on discrete point responses (i.e., principally multiple choice).

In order to meet the needs of score users, however, we provide the following table to indicate the approximate equivalencies for a CAEL Assessment test result. This relationship is based on the official scores of 1579 test takers who took other tests within one month of each other. The table provides a guide to interpretation only.

***Table 1.1. CAEL Assessment Score Comparability²**

CBT	CAEL	CLB	iBT	IELTS
133-190 (450-520)	10 – 40	1-5	<60	1.0-5.0
190-213 (520-550)	50	6		6.0
213-237 (550-580)	60	7		6.5
237-280 (580-650)	70-80	8-9	86+	7.0-7.5
280+ (650+)	90	10-12		8.0-9.0

²Initial results of research suggest the above equivalencies. These conversions are for comparison only and have not been validated through extensive research. Research is ongoing.

Because of the very different nature of the tests it is both impossible and inappropriate to be more precise than this. Tracking studies of students admitted with 580+ on the paper-based TOEFL and 70+ on the CAEL Assessment and other studies relating to the

² Bands 10-40 and 70-80 are collapsed in the table because they are generally not critical cut-offs in most post-secondary programs. Each of the band scores is fully elaborated in the criteria descriptions provided in Chapter Four, pp. 57-70 of this *CAEL Assessment Test Score and Users' Guide and Manual*.

usefulness of inferences drawn from performances on the CAEL Assessment are reported in the chapter which examines validity, Chapter Six, pp.75-83.

The criteria which define and describe performance across the sub-tests provide an excellent source of information regarding what students at specific band levels are able to do. All of the criteria are elicited from and relevant to academic study.

Question: How can a test taker prepare for the CAEL Assessment?

Answer: A complete sample test is available online at: <http://www.cael.ca/taker/Rainforest.shtml>. The *CAEL Test Taker Preparation Guides* are available for purchase from the CAEL Assessment Office or consult the website at: <http://www.cael.ca/taker/prepguide.shtml>. The test taker may also be able to take advantage of CAEL test preparation workshops or courses offered in some locations in advance of the test. For information about test preparation workshops, test takers should check the CAEL Assessment website or contact their local test centre.

Strengths of the CAEL Assessment

The CAEL Assessment reflects the attitudes, beliefs, values and understandings of what it means to use language for academic purposes.

By virtue of its collaborative, in-house development which was undertaken by a test development team comprised of teachers in the English for Academic Purposes (EAP) program, genre and discourse specialists, and professors from various disciplines in the academy as a whole, the tasks and performances on the CAEL Assessment have been systematically identified to represent as fully as possible the target language use domain of the test - namely English in use for academic purposes within the setting of the university.

The CAEL provides test takers with a descriptive profile of their ability to use English for academic purposes within the setting of the university.

Overall, CAEL Assessment results are criterion-referenced and range from band scores 10 to 90. The results are reported as part of a *profile of academic language ability*. Test takers' performances in each of the CAEL sub-tests of reading, speaking, writing and listening are also criterion-referenced, allowing the test taker and score-user alike to interpret each sub-test result in relation to specific capabilities. The test provides information about a test takers' ability to use English for academic purposes across a number of academic situations (e.g. speeded and unspeeded, formal and informal test settings, etc.). Institutions and graduate supervisors who receive students with CAEL Assessment results may use the profile to make better informed admission decisions. In addition, English as a Second Language (ESL) teachers may use CAEL results for diagnostic purposes, and test takers may use their results to inform their study. In this way the CAEL Assessment provides what Delandshere & Petrovsky (1998), Moss (1994), Shohamy (1998) and others argue is a fundamental requirement for validity: sound, empirically-grounded interpretation and a full explanation of the final score.

The CAEL Assessment aims to amass evidence of systemic or ecological validity.

On-going development of the CAEL reinforces links between language and learning theory on the one hand, and testing, teaching and learning in practice on the other. The CAEL test development team is concerned with *systemic* or *ecological validity* or positive "washback over time" (Shohamy 1996).

The CAEL Assessment is affiliated with the School of Linguistics and Language Studies.

Affiliated with the very active School of Linguistics and Language Studies (SLaLS), the CAEL Assessment benefits from the input of graduate students, researchers and professors engaged in a wide range of inter- and intra-disciplinary research; from teachers engaged in EAP teaching and from visiting faculty. The CAEL Assessment benefits directly from the shared expertise within both SLaLS and the Language Assessment and Testing Research Unit (LATRU).

CHAPTER TWO

THE CAEL ASSESSMENT : Sample Tests

Chapter Two provides sample test versions of both the **CAEL Assessment** written test and the **Oral Language Test (OLT)**.

- *The CAEL Assessment Sample Written Test* includes a listening task, two reading tasks and a writing task all of which relate to the topic of *Deforestation*. As is the case in all CAEL Assessments, the prompt for the writing task is at the beginning of the test. Actual CAEL Assessments are comprised of three separate test booklets: one for responses to the listening and reading tasks, one for the reading texts, and one for the writing task. For convenience, the CAEL Assessment Sample Written Test is printed from pp.25-40.
- *The Oral Language Sample Test* provides test takers with practice in taking the five OLT tasks. MP3 versions of the OLT tasks are available on the website at: <http://www.cael.ca/taker/oral.shtml>

The Canadian Academic English Language (CAEL) Assessment

Sample Version

This booklet is designed to help you prepare for the Canadian Academic English Language (CAEL) Assessment. If you are coming to Canada from a different country, you may not be familiar with this kind of test. This is a sample with answers to the listening, reading and writing sections at the end of the Assessment. Please note, however, that the topic of this practice booklet is not the same as the one on the CAEL Assessment that you will take. This sample is shorter than the complete CAEL Assessment. It is intended to give you an idea of what to expect, and a chance to practice.

The CAEL Assessment is designed to determine whether or not you require further instruction in English in order to study at the University level. Because your goal is academic study, the CAEL assessment requires that you do basically the same kinds of tasks you would at university - listening, reading and writing - all involving the same topic. There will be one lecture/listening section, two or three articles in the reading section and an academic writing task. An oral test is part of the registration process.

TABLE OF CONTENTS

	PAGE
Sample Test (including readings)	26
Appendix 1 - Answer Key	41
Appendix 2 - Transcript of Lecture	46
Appendix 3 - Writing Sample for Essay Question	47

The Canadian Academic English Language (CAEL) Assessment

Sample Version

DEFORESTATION

In the CAEL Assessment, you will be asked to do the things you would be expected to do in a university course. You will read two articles which will provide you with background information about a topic and answer questions about them. You will listen to part of a university lecture which will deepen your understanding of the topic. The final task will be to write a response to a question about the topic using information from the readings and the lecture.

Summary

Time	Task	Points
30 minutes	Reading 1	15 points
30 minutes	Reading 2	27 points
15 minutes	Listening	19 points
45 minutes	Writing	Level placement

Total time: 2 hours

All the readings and the lecture in this sample CAEL Assessment are on the topic of deforestation. It is important to understand that at this point in the test you do not need to know anything about the topic of deforestation. All of the information you need to respond to the Writing Task is provided in the lecture and the readings.

You will be able to use information from the lecture and the readings to complete the **Writing Task**. At the end of the test, you will write an essay agreeing or disagreeing with the following statement:

Canada's largest natural resource, its woodlands, is in danger of being exhausted. Although part of the fault is due to natural causes, the main culprits are the forest industry itself and the government.

Reading 1: "Canada's Wasted Woodlands"

(30 minutes/15 points)

READ THE QUESTIONS BEFORE YOU READ THE ARTICLE. Now, read the article in order to answer the questions.

1. Which paragraph states the main concern of this article? _____
(1 point)

2. What is the purpose of the first six paragraphs?
(2 points)

3. According to this article, what are the two main reasons for the current state of the forestry industry?
(3 points)

4. If the forestry industry declines, what will the economic consequences be for Canada?
(4 points)

5. List two solutions to Canada's forestry problems suggested in this article.
(2 points)

6. Match the number of the paragraph in the article to the sentence below which best summarizes the paragraph. The first one has been done for you. (3 points)

- a) 17 Discusses the economic consequences of the industry's failure.
- b) _____ Describes the loss of productive forest area through both human and natural causes.
- c) _____ Talks about the industry response to the situation.
- d) _____ Suggests a possible solution to the political aspect of the forestry problem.

THIS IS THE END OF READING 1

Reading 2: “Exhausting the Inexhaustible”

(30 minutes/27 points)

READ THE QUESTIONS BEFORE YOU READ THE ARTICLE. Now look for the information in the article.

1. What is the main idea of this article. Put a check mark ✓ beside the correct answer. (1 point)

- the writings of John A. MacDonald
- the timber trade is Canada’s largest industry
- Canada’s timber resource is inexhaustible
- Canada is running out of trees

2. What percentage of Canada’s land is covered by trees? (1 point)

3. Fill in the chart below explaining why this percentage is deceptive. (6 points)

Problem	Details
Quality of trees	<ul style="list-style-type: none">• __________
Availability of suitable trees	<ul style="list-style-type: none">• _____• _____• _____
Type of mills	<ul style="list-style-type: none">• _____• _____

4. What evidence is there to support the following statement: **“Our forests are also shrinking.”** (1 point)

5. What is the purpose of paragraphs 5 and 6 in the article? (1 point)

GO ON TO THE NEXT PAGE

6. According to the article, in which area could the depletion of Canada's forests have devastating consequences? Put a check mark ✓ beside the correct answer.
(1 point)

- Environment
 Employment
 Exports

7. Fill in the chart below with information about loss of our forests due to fire and insects.
(4 points)

Year	Natural cause	Result
1980	• _____	• _____
1981	• _____	Eastern Canada • _____ Quebec • _____ _____

8. According to the article, which one of the following causes the most damage? Put a check mark ✓ beside the correct answer.
(1 point)

- Fire
 Disease
 Insects
 Human perceptions

9. List three ways in which the forests are neglected:
(3 points)

- _____
- _____
- _____

10. According to the article, who is responsible for the failure to manage the forest resource?
(1 point)

- _____

GO ON TO THE NEXT PAGE

11. Which level of government is responsible for each of the following areas of neglect?

Use **P** for Provincial and **F** for Federal³. The first one has been done for you.

(4 points)

LEVEL OF GOVERNMENT	AREA OF NEGLECT
<u> F </u>	lack of leadership
<u> </u>	has not enforced reforestation
<u> </u>	reduced CFS staff
<u> </u>	halved available research money
<u> </u>	has not developed plans for forest industry

12. What is now being done to protect the trees?

(3 points)

- _____
- _____
- _____

13. Are the forests still shrinking? Please put a check mark ✓ beside the correct answer.

(1 point)

Yes No

THIS IS THE END OF READING 2

³ **Definitions:**

Federal Government: National Government of Canada which passes laws concerning issues that affect all of the provinces.
Provincial Government: Regional Government of each province which passes laws concerning issues affecting only that province. There are ten provinces in Canada.

Listening Section

(15 minutes/19 points)

During this part of the test you will hear a taped lecture on deforestation. The tape will be played only once. You must answer the questions **while** you are listening to the tape. The questions follow the sequence of the lecture. Do not stop to re-write answers during the listening. At the end of the lecture, you will have 5 minutes to go over your answers. You have 3 minutes now to read the questions in this section before you listen to the lecture.

ANSWER THESE QUESTIONS WHILE LISTENING TO THE LECTURE.

1. What is this lecture about? Check one.

(1 point)

Logging in the Amazon rainforests

Pollution of the earth's atmosphere

Destruction of the world's forests

The forest industry since 1950

2. According to the speaker, name 3 reasons why "an all-out war is being waged against the forests of the world"?

(3 points)

a) _____

b) _____

c) _____

3. How much of the earth's surface was covered by forests in 1950? _____

(1 point)

4. How much is covered now? _____

(1 point)

GO ON TO THE NEXT PAGE

5. Fill in the chart below with the missing information.
(4 points)

Location	Degree of Deforestation
Malaysia	-half the forest logged in the past 20 years
• _____	-40% forest cut in past 10 years
Amazon forest	• _____
Asia and Africa	• _____
• _____	-more than 1/3 deforested

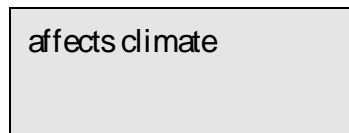
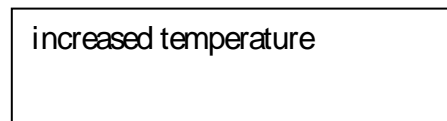
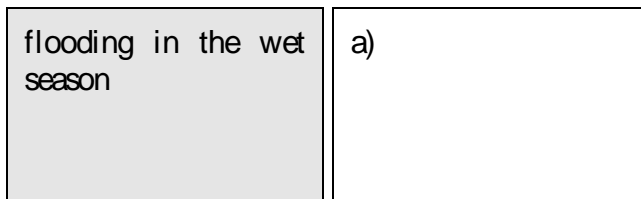
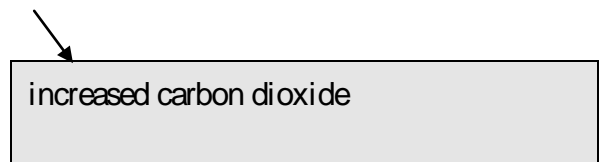
6. Fill in the chart below with information from the lecture.
(4 points)

RESULTS OF DEFORESTATION

LOCAL



GLOBAL



THIS IS THE END OF THE LISTENING SECTION

Reading 1

Canada's Wasted Woodlands

Because of poor management, its most treasured resource could be gone in a century.

By Douglas Martin
Victoria, British Columbia

1. For one last moment, the towering hemlock stands defiantly, as if unwilling to relinquish the idea of living after nearly three centuries of effort. Then, ever so slowly, it leans over Yellow Creek Mountain, about 80 miles north of here, and falls to earth with a flat thud.

2. It's one of about 130 trees that William Laplante - a French Canadian who has worked as a "feller" of trees for four decades - will saw down this sunny day. For him, the hemlock's death waltz is happily routine.

3. He makes more than \$160 for six and a half hours work each day; the labour is considerably easier than it used to be, and across the Pacific, there are plenty of eager buyers.

4. "The Japanese are crazy for that wood, eh?" Mr. Laplante said, pointing to the tree.

5. On the other side of the same mountain, his son Marcel, a feller for 17 years, is less sanguine.

6. "Pretty soon, all your big wood's going to be gone," he said in an emotionless voice.

7. In accord with many forestry experts, the younger man worries whether enough is being done to conserve Canada's biggest natural resource. But even if the woodlands are preserved, others fear that Canada's forests are proving uneconomical in a changing, more competitive world.

8. Essentially, the problem is that the industry wastes up to half of each tree it cuts; it also fails to replant much of what it harvests. Experts worry that, given present trends, the resource that in the nineteenth-century timber baron thought would last 700 years may disappear in less than a century.

9. At the same time, competitors are cropping up all over the globe - from the forests of Siberia to the jungles of Brazil. And these emerging industry rivals are, in some cases, producing new wood more quickly, more efficiently and more cheaply than their Canadian counterparts.

10. Awareness of the dilemma was first sparked by the migration of the environmental movement north of the

49th parallel; to politicians, expenditures for preservation and replanting, lagging far behind American outlays since Theodore Roosevelt's time, have now become as Canadian as Mounties and moose.

11. But the biggest impetus has come from the timber industry itself, which now finds it is using more costly fuel to carry trees to mills from increasingly remote cutting areas. Many now contend that if replanting had occurred as the trees were harvested, much of the current expense would have been avoided.

12. Moreover, the recession-wracked industry is enduring its fourth bad year in a row, and the effects are readily apparent in the country's unemployment and welfare rolls and in some corporate balance sheets. In fact, more than one-fifth of the timber industry's work force is unemployed; and in British Columbia alone, forest product companies lost \$300 million in 1982.

13. "Our forests are in a shambles, a mess," asserted Jack Walters, director of the University of British Columbia's big research forest: "It's a tragedy of the first dimension." And Michael Innes, manager of forestry for Abitibi-Price Inc., the world's largest producer of newsprint, said: "We're gradually running out of wood, and there's no more wood over the next hill."

14. This assessment is shared by Canada's chief forester, F.L.C. Reed: "Canada will run out of economically priced trees if it doesn't pull up its socks." He acknowledged that, as a result of poor management practices, "The next crop of trees is going to be smaller than the last one, and it's going to take longer to grow."

15. But the problem is not an obvious one - and with good reason. More than 40 percent of Canada is covered by trees - enough to build about 33 million houses. Flying across the country, it is almost impossible to avoid the impression that Canadians, in essence, live in the woods.

16. Indeed, one-tenth of the productive forest area in the world lies in Canada. The trees here on Vancouver Island represent one of the world's greatest exploitable supplies of mature large timber. Elsewhere in Canada,

smaller spruce trees make excellent pulp and paper products.

17. The economic consequences are immense. Canada leads the world in forest exports, with the value of shipments in excess of \$12 billion. Some 300,000 Canadians work in the forestry industry, with indirect employment estimated at more than 1 million. Only the former Soviet Union has more trees, but so far has been unable to capitalize on its advantage.

18. But the United States, which now buys large quantities of timber from Canada, faces no threat of a wood shortage; and new supplies are popping up in its Southeast. Meanwhile, a growing number of wood substitutes abound. And Michael Newton, professor of forest ecology at Oregon State University, said that if the latest in forest-management procedures were employed in the United States, "We could export timber like we're exporting wheat."

19. It is Canada that has the chips on the table: forest products represent one-seventh of all manufactured goods, last year accounting for \$23 billion in sales. They make a bigger net contribution to the nation's balance of payments than metals, agriculture, fisheries and the automotive industries combined.

20. In the face of these economic realities, it is obviously a bad time for Canada to be running low on trees, but that is exactly what experts say is happening - albeit slowly. They say that in the past the size of the resource was vastly overestimated, and that Canada may have only half of the productive forest it thought it had.

21. The mounting concern about Canada's timberlands comes against the increasingly clear perception that elsewhere in the world, wood is plentiful. Accordingly, the Forest Industries Advisory Committee, a high-level business and labour panel, recently warned that Canada's forestry industry future hinges on its ability to compete in today's tough marketplace.

22. Even if much of Canada's softwoods are of a higher quality than wood elsewhere, "the world seems to have an infinite capacity for making do, if the price is right," the panel said in a report to the Ottawa government. The panel asked for far more assistance from Canadian taxpayers. Its justification was stark.

23. "If the industry fails, in the future there will be a helluva lot fewer taxpayers," James Buchanan, president of the Canadian Paperworkers Union, said last week.

24. "In Canada, we are no longer a dominant producer of a scarce resource - we're now a large owner of an

abundant resource," observed C.C. Knudsen, chairman and chief executive of MacMillan Bloedel Ltd., Canada's largest forest product company and now a unit of Noranda Mines Ltd. "And we are beginning to price ourselves out of the market."

25. Indeed, MacMillan Bloedel Ltd., like many other Canadian companies, has already felt the pinch. Last year the company reported a \$57.3 million dollar loss, compared with a modest profit of \$3.3 million in 1981. But 1983's prospects have brightened a bit, with \$14.1 million in first-half profits, compared with a loss of \$21.5 million in last year's first half.

26. At the same time, Abitibi's fortunes have also declined sharply, as profits slipped to \$80.9 million in 1982, from \$117.6 million in 1981. Although the company reported a \$10.8 million profit in the first six months of 1983, this is well below the \$43.2 million it earned for the first half of 1982.

27. And it seems that the long-term outlook for many forest-product companies - both big and small - remains less than promising. The Science Council of Canada, a quasi-governmental organization that assesses scientific and technological resources, reports that about 6,000 square miles of forest land is destroyed or wasted annually. The council cites insufficient replanting, inability of some lands to regenerate naturally, and the effects of fire, insects, disease and wind.

28. "Natural regeneration is not keeping up with what you need," said Jack Munro, president of the International Woodworkers Union in Western Canada. "We're at a crossroads."

29. Apart from the timberland here in British Columbia - the last stop on the colourful westward trek of the loggers and saloonkeepers and painted ladies who followed them - the virgin forest is clearly gone. What is left is so scattered - or so remote and ecologically fragile - that it is not economical to exploit. New growths, in areas where the first generation of timber was long ago harvested, have fallen short of hopes.

30. Signs of diminishing resources are everywhere. Supplies near mills have been so depleted in some areas that it is necessary to transport logs hundreds of miles. Quebec and Ontario furniture-makers must import maple logs from the United States. And millions of acres of forest land in the Maritime provinces have been devastated by the spruce budworm.

31. The overall Government statistics are equally chilling. Each year Canada loses nearly 40,000 acres of forest land through lack of natural or artificial regeneration, one-fifth of what is harvested each year. And more than 30,000 acres fall prey annually to fires

and disease. Still more is lost to parks, farms and subdivisions. The result has left Canada's backlog of wasted forest lands at somewhere between 11.5 million and 70 million acres - with most analysts leaning toward the higher figure.

32. The upshot has been the renewal of a debate as old as the Dominion itself. "We are recklessly destroying the timber of Canada and there is scarcely a possibility of replacing it," declared Sir John A. Macdonald, Canada's first Prime Minister, in 1871.

33. Canadians are also worried about dwindling demand for the wood they have left - although such slackening might conceivably buy some time to replenish the forests. In the 1960s and early 1970s, global wood products orders were growing at nearly 5 percent a year. Analysts now believe the industry will be lucky to see a 3 percent annual growth rate in demand, even with a strong economic recovery.

34. In part, the industry response has been to seek more government support. The advisory committee last week, for example, recommended to the federal government that it assist the timber industry with a mix of tax incentives, trade development and heightened reforestation measures.

35. The Canadian provincial and federal governments, which own more than 90 percent of Canada's forest land and lease logging rights to companies, are addressing the timber resource problem. It is estimated that by 1985, the federal and provincial governments and industry will increase their combined annual forest regeneration spending to \$330 million from \$250 million five years earlier.

36. But much remains to be done. Mr. Newton, for instance, believes that Canada must spend tens of millions of dollars killing the economically worthless stands of hardwood, principally aspen, now blanketing what were formerly rich spruce forests of Canada.

37. Part of the problem, critics say, is that politicians are in the habit of thinking from election to election,

rather than in terms of the 60 years or so it may take to grow a mature tree. Another negative effect of the politicization of the forests, said Mr. Walters of the University of British Columbia, is that funds are not channeled to the best growing areas, but rather are apportioned among legislative members with trees in their districts.

38. A solution, Mr. Walters suggests, would be to put Canadian forests in private hands, as roughly three-fourths of the forest in the United States already are. But the Canadian public would probably never relinquish common ownership of its forests, observers say. Most companies believe the new provincial arrangements, which require more of companies but also give them a greater guarantee of tenure, protect long term investments.

39. Growing ecological awareness on the part of the general public has also added momentum. And the continuing spruce budworm, epidemic in the East, and several severe fires have underlined the limits of the already shrinking resource. Mr. Innes recalled a 300,000-acre fire in 1980 in Abitibi's forests near Thunder Bay, Ontario. He said that the company's board asked him, "If we ever had a second one of those things, what would happen?" His answer: "It would close three mills. Permanently."

40. In forestry, however, the answers are seldom simple. Experts agree that more forest land is better than less. However, proper management could cut in half the time needed to grow a commercial tree.

41. That might be impossible without vast resources of expertise and money. During the five years ending last year, the timber industry had capital outlays of \$15.3 billion, nearly twice the level of the preceding five-year period.

42. Although more money is needed to insure the continued productivity of Canada's forests, the question of how much more, the advisory panel said, may be less important than how new funds are used.

EXHAUSTING THE INEXHAUSTIBLE

A century ago, Canadians thought they could never cut all the trees in the country.

Now, as a freelance writer, M.L. Allen reports, our biggest industry is running out of raw material.

"The sight of immense masses of timber passing my windows every morning suggests to my mind the absolute necessity there is for looking into the future of this great trade. We are recklessly destroying the timber of Canada and there is scarcely a possibility of replacing it... It occurs to me that the subject should have been looked in the face and some efforts made for the preservation of our timber."

1. Those words were written by Sir John A. Macdonald in 1871 as he watched giant timber rafts move down the Ottawa River to the sawmills. He was certainly right that something needed to be done. However, it was 50 years before even the first small steps toward reforestation - planting new trees to replace what has been cut - were begun. It was another 50 years before many people started to think there might be a problem. It's only within the last few years that the full reality of the situation has sunk in - Canada is running out of trees.

2. That would seem to be impossible. Trees cover 70% of Canada's land area - 44 million km² of forests - an apparently inexhaustible supply. But, those figures are deceptive; the productive forest in Canada is only half as large as it appears to be.

3. A large portion of our trees are not the right kind of quality for harvesting. Of those that are suitable, many are not available; they are protected in parks and reserves, or they are in inaccessible areas such as mountains; or they are so remote as to make harvesting uneconomical. The situation is further complicated by the fact that many mills are built to handle only certain kinds of trees; thus, even if there are still other good trees close by, the mill might not be able to use them, and the mills that could use them might be too far away.

4. Our forests are also shrinking. A Science Council of Canada report states that "one-eighth of the country's productive forest area has deteriorated to the point where huge tracts of land lie devastated, unable to regenerate a saleable crop within the next 60 or 80 years." This amounts to a loss of over 200,000 km² each year. In practical terms, this means that in some areas, logs must be hauled hundreds of kilometres to mills; in other areas, even helicopters - at a cost of \$2,500 an hour - are used to get the trees. In some towns - Fort Nelson, B.C.; Dryden, Ontario; Noranda, Quebec; and Chatam, N.B. are examples - there isn't enough timber to keep the mills operating.

5. The depletion of Canada's forests, however, is far from a merely local problem. The forest industry here is not just big business, it is our biggest business - bigger than oil, bigger than agriculture, bigger than minerals and fisheries. Directly, and indirectly, one in ten Canadian workers is involved in forestry, making it the country's largest employer after government. In 1981, the forest industry produced \$23 billion in goods, and paid more than \$3 billion in taxes. Forest products account for nearly 15% of all goods manufactured in Canada (in New Brunswick it's 30%, and 50% in British Columbia). More than 300 communities across Canada are one-industry towns, solely dependent on forestry, not just for their livelihoods but for their very existence.

6. Internationally, Canada produces 9% of the world's forest products, the third largest producer after the United States and the former Soviet Union. However, our country is the largest exporter of manufactured forest products; 90% of our newsprint, 70% of our lumber, and 60% of our chemical pulp are exported; in all, nearly 30% of the manufactured forest products that enter the world market come from Canada. This export of wood products contributes a net gain of \$12 billion to the Canadian economy, more than the combined exports of farm products, fish, metal and coal.

7. Given the importance of forests to Canada, their depletion could have devastating consequences. It has been estimated that within the next ten years, 250,000 Canadians dependent on forestry could be out of work. At the same time, world demand for wood products is rapidly rising; it's expected to increase by 50% between 1975 and 2000. Thus, not only will existing jobs be lost, but new jobs and new revenue (perhaps as much as \$12 billion a year) will also be lost if Canada is unable to respond to the growing demand.

8. In part, the loss of our forests is due to natural causes. Each year, fire, insects and disease claim about 100 million m of wood. In 1980, for instance, the worst year in five decades for forest fires in Canada, five times more timber burned than was harvested.

In 1981, spruce budworm infestation covered 350,000 km² of forest in eastern Canada, severely affecting Newfoundland and Cape Breton; in Quebec, losses to the spruce budworm may end up totalling 150 million m of wood.

9. More than fire and insects, though, the depletion of Canada's forest resources is the result of mistaken attitudes and perceptions. For example, in the middle of the 19th century, a government survey estimated that the pine forests of the Ottawa Valley were sufficient to supply its mills for the next 600 years. Seventy-five years later, the forests were gone and most of the mills had closed down. Still, despite this and other examples to the contrary, the myth of the inexhaustibility of Canada's forests persisted for a very long time.

10. If the forest is thought to be inexhaustible - if you only have to go over the next hill to get more trees - it's not necessary to do much to preserve the woodlands. And that's exactly what's been done to this point - not much. The two levels of government receive huge tax revenues from forestry. However, only about 5 cents on the dollar goes back into the forestry management programs that help to ensure the continued health of the industry. In Canada today, there is one forester for every 800 km² of forest; the international standard is one for every 120 km², and in the United States there is one forester for every 80 km² of forest. Likewise, in Canada in 1978 only 0.6% of forestry sales were spent on research and development, compared to 1.5% of sales in the United States. By far, though, the most damaging aspect of the neglect has been insufficient reforestation. Canada's forest companies cut about 8,000 km² a year, and plant or re-seed about a quarter of that area. About 2,000 to 3,000 km² of the rest will restock reasonably well on its own. But that still leaves 3,000 to 4,000 km² that lie idle, or produce unusable scrub. Hence, the disappearance of the forests. The problem is not so much exploitation by the industry, which has been central to Canada's economy for 150 years. The real failure is that of government, which hasn't made sure that the trees are replaced, that our forest supplies - and the revenue that they generate - will continue into the future.

11. The federal government has not provided much leadership, and even cut the Canadian Forestry Service staff and research funds in half during the 1970s. The provincial governments, which control 90% of Canada's forests, have either ignored the issue of reforestation, or failed to come up with sensible plans for the forest industry.

12. Even where intentions are good, execution has fallen short. In 1977, Premier William Davis promised that Ontario would plant at least two trees for every one cut; in fact, in recent years, only 38% of land cut has

been replanted, and foresters think that only one third of the trees planted will survive. What should have been a great renewable resource is starting to look like a non-renewable one. As Les Reed, former assistant deputy minister in charge of the Canadian Forestry Service put it, "We have mined our forest when we should have been managing it."

13. Because trees are such a slow-growing crop, usually taking 60 to 100 years to reach harvestable size, our current problems have been a long time developing. "In this kind of situation, you don't fall off a cliff," Les Reed says, "you slide down a slippery slope." But eventually you'll get to the edge, and for the first time in the history of Canadian forestry, that edge is in sight still distant, perhaps, but very definitely there. "As we know it today, the forest industry is dying," John Walters, a University of British Columbia forestry professor says, "Action is urgently required if the forestry industry is to survive."

14. Blunt talk, but maybe - 113 years after Macdonald suggested it - that's what is needed to get the subject of preserving Canada's forests "looked in the face." People in the forest industry are speaking out about the problems. The federal government is getting more actively involved. Ontario is making agreements with the forest industry about reforestation programs. Research is developing faster-growing, stronger trees, and new ways to reduce waste and utilize trees that until now have not been useable.

15. So, some things are finally starting to be done. Whether it's enough won't be known for a while. Unfortunately, if it turns out that it's not enough, it will already be too late to make amends. Billions of dollars and hundreds of thousands of jobs are at stake. Meanwhile, the chainsaws are working. In the time it took to read this article, about three hectares of productive forest have been lost.

Writing Section

(45 minutes/ level placement)

You have 45 minutes to write an essay agreeing or disagreeing with the following statement:

Canada's largest natural resource, its woodlands, is in danger of being exhausted. Although part of the fault is due to natural causes, the main culprits are the forest industry itself and the government.

Use the information from the readings and the lecture to help you organize and plan your essay.

Please write at least 1 page.

DO NOT COPY DIRECTLY FROM THE READING ARTICLES. If you wish to use one or two sentences, please use quotation marks ("...").

You are advised to spend 15 minutes planning your essay and 30 minutes writing.

DO YOUR PLANNING IN THE SPACE BELOW.

Appendix 1 - ANSWER KEY

Reading 1: Canada's Wasted Woodlands (30 minutes/15 points)

1. Which paragraph states the main concern of this article?

(1 point)

7

2. What is the purpose of the first six paragraphs?

(2 points)

Introduction; Background

3. According to this article, what are the two main reasons for the current state of the forestry industry?

(3 points)

Half of each tree cut is wasted. The industry fails to replant what it cuts.

4. If the forestry industry declines, what will the economic consequences be for Canada?

(4 points)

Unemployment

Loss of export revenues

5. List two solutions to Canada's forestry problems suggested in this article.

(2 points)

Put Canadian forests back into private hands.

Put more money back into forests.

6. Match the number of the paragraph in the article to the sentence below which best summarizes the paragraph. The first one has been done for you.

(3 points)

a) 17 *Discusses the economic consequences of the industry's failure.*

b) 31 or 27 *Describes the loss of productive forest area through both human and natural causes.*

c) 34 *Talks about the industry response to the situation.*

d) 38 *Suggests a possible solution to the political aspect of the forestry problem.*

Reading 2: “Exhausting the Inexhaustible” (30 minutes/27 points)

READ THE QUESTIONS BEFORE YOU READ THE ARTICLE. Now look for the information in the article.

1. What is the main idea of this article. Put a check mark ✓ beside the correct answer. (1 point)

- the writings of John A. MacDonald
- the timber trade is Canada’s largest industry
- Canada’s timber resource is inexhaustible
- Canada is running out of trees***

2. What percentage of Canada’s land is covered by trees?

(1 point) 70%

3. Fill in the chart below explaining why this percentage is deceptive.

(6 points)

Problem	Details
Quality of trees	· <i>A large portion of our trees are not the right kind for harvesting</i>
Availability of suitable trees	· <i>Protected in parks and reserves</i> · <i>Are in inaccessible areas</i> · <i>Are so remote that harvesting is uneconomical</i>
Type of mills	· <i>Many are built to handle only 1 kind of trees</i> · <i>Suitable mills may be too far away</i>

4. What evidence is there to support the following statement: **“Our forests are also shrinking.”**

(1 point)

A Science Council of Canada Report

5. What is the purpose of paragraphs 5 and 6 in the article?

(1 point)

To show the economic importance of the forest industry in Canada

6. According to the article, in which area could the depletion of Canada’s forests have devastating consequences? Put a check mark ✓ beside the correct answer.

(1 point)

- Environment
- Employment***
- Exports

7. Fill in the chart below with information about loss of our forests due to fire and insects.

(4 points)

Year	Natural cause	Result
1980	forest fires	· 5 times more destroyed than harvested
1981	spruce budworm infestation	Eastern Canada · 350,00 km ² affected Quebec · 150 million m of wood lost

8. According to the article, which one of the following causes the most damage? Put a check mark ✓ beside the correct answer.

(1 point)

- Fire
- Disease
- Insects
- Human perceptions**

9. List three ways in which the forests are neglected:

(3 points)

- **Only 5% of tax revenues from forestry goes back into Forestry Management Program**
- **Canada has one forester for every 800 km²**
- **Insufficient reforestation**

10. According to the article, who is responsible for the failure to manage the forest resource?

(1 point)

the government

11. Which level of government is responsible for each of the following areas of neglect. Use **P** for Provincial and **F** for Federal. The first one has been done for you.

(4 points)

Level of Government	Area of Neglect
F	shows lack of leadership
P	has not enforced reforestation
F	reduced CFS staff
F	halved available research money
P	hasn't developed plans for forest industry

12. What is now being done to protect the trees?

(any 3 out of 6)

(3 points)

- **People in the forest industry are speaking out about the problems**
- **The federal government is getting more actively involved**
- **Ontario is making agreements with the forest industry about reforestation programs**
- **Research is developing - faster growing trees; stronger trees; new ways to reduce waste**

13. Are the forests still shrinking?
 (1 point)
Yes

Listening Section

(15 minutes/19 points)

1. What is this lecture about?
 (1 point)

The destruction of the world's forests

2. According to the speaker, name 3 reasons why "an all-out war is being waged against the forests of the world"?

(3 points)

- a) People are using forests for firewood.*
- b) Logging valuable lumber, such as mahogany or teak because logging provides much needed foreign exchange.*
- c) Clearing of forests to produce more agricultural land needed to feed growing populations.*

3. How much of the earth's surface was covered by forests in 1950?
 (1 point)

One quarter, 25%

4. How much is covered now?
 (1 point)

Less than 1/5 (less than one fifth, less than 20%)

Fill in the chart below with the missing information.

(4 points)

Location	Degree of Deforestation
Malaysia	- half the forest logged in the past 20 years
· <i>Thailand</i>	- 40% forest cut in the past ten years
Amazon forest	· <i>10,000 km² per year disappearing</i>
Asia and Africa	· <i>1/2 deforested</i>
· <i>Latin America</i>	- more than 1/3 deforested

Fill in the chart with information from the lecture.

(4 points)

- a) *water shortages***
- b) *"Greenhouse Effect"***
- c) *no increase in available agricultural land***
- d) *affects agriculture.***

Appendix 2 - TRANSCRIPT OF LECTURE

Listening Section

(15 minutes/19 points)

You should answer the questions in the listening section while you are listening to the lecture. If, however, you are not able to, the lecture is reprinted for you below. Find a friend to read the lecture aloud while you answer the questions which follow.

Before you begin, take 5 minutes to read over the questions about the lecture.

Text of the Lecture

At the present time, an all-out war is being waged against the forests of the world. People are using them for a variety of purposes. First, there is the problem of firewood. Nearly half the world uses firewood as its main fuel, and it is partly because of this that forests are being cut down. In poorer countries, the forests are also being logged for their valuable timber, such as teak and mahogany, so that this timber can provide these countries with the foreign exchange that is so badly needed. As well, they're being cleared by slash and burn techniques in the hope of producing the agricultural land needed to feed rapidly-growing populations.

It's really not so surprising, then, that the forests are losing the battle. In 1950, about one-quarter of the earth's land surface was covered with forests and today it is less than one-fifth. The greater part of this loss has occurred in the tropical rain forests, which are continuing to be destroyed at the rate of about 112,000 km² a year, this is in an area the size of Nova Scotia and New Brunswick combined.

I'm going to ask you to consider these statistics which are from around the world. In the last twenty years, half of Malaysia's forests have been logged; 40% of Thailand's forest has been cut in the last ten years; the Amazon forest, the greatest rainforest in the world, is now disappearing at a rate of even more than 10,000 km² a year. Altogether, about half of Asia and Africa and more than a third of Latin America have been deforested.

The results of such massive destruction are starting to alarm experts, because the consequences are dangerous on both local and global levels. On the local level, the loss of a forest frequently results in erosion, flooding during wet seasons, and water shortages in the dry seasons. Instead of creating agricultural land, deforestation often produces a desert, which is, of course, not an area known for the production of crops.

On a global basis, the loss of these forests could be even more catastrophic. Rain forests have been called "the lungs of the world." They take in carbon dioxide and release oxygen. An increase of carbon dioxide in the atmosphere causes what's known as the "greenhouse effect" - a gradual rising of global temperatures that could have a disastrous effect on climate and agriculture throughout the world. And there is another result which is indirect but extremely serious from this deforestation of the world's tropical forests. The tropical forests are the world's gene pool, containing more than 3.5 million species of plants and animals, of which only approximately 15% have been properly catalogued. The loss of the forest could result in the extinction of a million or more species and most of these would not even have been discovered, studied or analyzed for their potential benefits. Another area where we stand to lose is in the area of medicines. About one-quarter of our medicines are derived from the tropical forests, so the scope of the potential loss is enormous. And every minute, another 20 hectares of tropical forest disappears. The situation is truly alarming.

Appendix 3 - WRITING SAMPLE

Writing Sample for Essay Question

It is clear that although part of the fault ~~is~~ for the alarming decline of Canada's woodlands is due to natural causes ~~the main~~ - such as fires, insect attacks, etc. - the main sources of decline are the irresponsible actions of the forest industry itself, aided and abetted by federal and provincial governments.

According to the article "Exhausting the Inexhaustible", by M.L. Allen, "Canada's forest companies cut down about 8000 km² a year, and plant or re-seed about a quarter of that area." It doesn't take a mathematical genius to figure out why our forests are disappearing - if only a quarter of what is cut down is replanted. Soon - when we run out of virgin forest - we'll be cutting through the ¼ that was replanted. So, the total number of trees gets ~~less~~ smaller and smaller.

Mr. Allen ~~says~~ is right in claiming "the real failure is that of the government", the government has exercised little control over the forest industry. It's only now - after so much damage has been done - that there is a growing pressure for the government to act.

The 2nd article supports government awareness of the forest problems. It also sees "growing ecological awareness on the part of the general public has also added to the momentum."

If government had been acting responsibly over the past 100 years we would not now be facing such serious declines in our forests. ~~We should~~ It's right to be aware of the decline of rainforests in Brazil, but only if Canadians recognize the same decline and destruction in Canada.

Please note that for CAEL Assessments, you should write at least one page *handwritten*.

**CANADIAN ACADEMIC ENGLISH LANGUAGE ASSESSMENT
ORAL LANGUAGE PRACTICE TEST**

Please note: This is a practice sheet. Any reference to page numbers in the test booklet does not apply.

Sample Task 1

In this task you will talk about yourself. You may be asked to discuss either

- a. what you want to study at university and why
- or**
- b. your experience learning English and studying English for academic purposes.

Sample Task 2

When you take a university course you need to understand spoken instructions concerning assignments. Sometimes you will need to report this information to someone else. In this task you will listen to some recorded speech and follow the directions on the tape. You may write down any information here:

Sample Task 3

In this task you will be asked to give information about either

- a. an academic document
- or**
- b. a short article.

a. Academic Document:

Whenever you take a course at university, you will need to read some documents which inform you about the course. Below is a course outline:

<u>Anthropology 32: 124 - Cultural Anthropology Spring 2000</u>	
Monday 8:30 - 10:30 am Seminar 1 hour per week – to be assigned Room 408 Dunton Tower	Prof: John Robertson 314 Paterson Hall Office Hours: Mon. 10:30 - 11:30 Tel: 788-2013
<u>Course Description</u> Anthropology is not a science of the exotic carried on by scholars in ivory towers but a discipline with a lot to tell the public. One of its contributions is its broadening, liberating role in a college education.. This course will examine the development of this discipline as it now exists and look at its present day contributions to society as a whole.	
<u>Course text</u> The majority of readings will be taken from the following text. Other readings are on reserve in the library Kottak, C.P. <u>Anthropology the Exploration of Human Diversity</u> . New York: McGraw-Hill, Inc. 1991.	
<u>Grading</u> You will be expected to prepare two class presentations (topics to be discussed) during your seminar times. These presentations will be worth 20% of your total grade. In addition you will be required to submit a paper of 10 – 12 pages typed (topic to be assigned) which will be worth 40% of your total grade. There will be a final exam which will be worth 40% of your grade. If you have any questions or problems, please contact during office hours or call or leave a message and I will get back to you.	

b. Short Article:

Many times in your academic study you will need to read articles to summarize some points and express your opinion based on an article. In this task you will be asked to summarize in your own words the key points of a short article and to give your opinion on the issue discussed in the article.

Sample Task 4

Occasionally in your academic program, you will need to read aloud, for example, to quote from sources or to refer to part of a reading. In this task you will be asked to read either:

- a. the article you spoke about in **Sample Task 3**,
- or**
- b. an excerpt from an academic text.

b. Excerpt from Academic Text: *The Sub Disciplines Of Anthropology*

Cultural Anthropology

Cultural anthropologists study society and culture, describing and explaining social and cultural similarities and differences. In considering diversity in time and space, anthropologists must distinguish between the universal, the generalized, and the particular. Certain biological, psychological, social, and cultural features are *universal* - shared by all human populations. Others are merely *generalized* - common to several but not all human groups. Still others are *particular* - not shared at all.

Sample Task 5

In this task you will be asked to participate in a group discussion.

HANDOUT

32.325 - Violence in Society

Instructor: James Woods

**GROUP PROJECT WORK SHEET
VIOLENCE IN SOCIETY**

Your group will give an oral presentation on each of the following aspects of violence in society. Each member of the group will choose **one** topic listed below.

Choose your topic from the list below:

TOPIC	DETAILS
Violence in Movies	Effect on children Effect on male female relationship
Violence in families	Effect on children Problems with elderly people
Violence in the schools	Effect on learning Weapon control in high schools
Violence on the streets	Effect on ordinary citizens Cost for police, etc.

Questions: Due date, length, evaluation, written requirements, etc.

Chapter Three: Characteristics of CAEL Test takers

Introduction

Individual test takers take the CAEL Assessment for a variety of reasons. Most CAEL Assessment test takers write the test as part of the process of applying for admission to a university program. They take the test to demonstrate that their level of English language is adequate for university study. Some test takers take the test as part of their registration in an English as a Second Language (ESL) or an English for Academic Purposes (EAP) course. Occasionally test takers write the test in order to satisfy a personal desire to know more about their level of English language. Given the variety of reasons that test takers have for completing the test, we generally find that our test population has diverse characteristics.

CAEL Assessment test takers are asked to provide some background information as part of the registration process (see Chapter One, p.13). This information helps us to gain a better understanding of the diverse characteristics of our test population. We use this information to monitor the performance of the test for various groups of test takers and to guide the selection of pilot test groups used when developing new versions of the CAEL. The information is stored in a secure database and is summarized at regular intervals. In this Chapter we describe a sample of over 7000 test takers who completed their CAEL Assessment at the main test centre at Carleton University from January 2003 through January 2009. Information describing gender, age, student status, first language and country of origin is presented.

Gender, Age and Student Status

The sample of test takers described in this chapter is composed of 57.7% males and 42.3% females. Throughout the history of CAEL testing we have found this approximately 60/40 ratio to be quite consistent at the main test centre. By comparison, the full-time student population at Carleton University is 48.9% male and 51.1% female (OIRP, 2007). Clearly, the proportion of male students completing the CAEL Assessment is higher than the proportion of males registered for full-time study at Carleton University. Of course, when testing is undertaken at overseas test centres the proportion of males and females may differ considerably from that which is observed at the main test centre. An analysis of test performance by gender is presented in Chapter Five, *Score Distributions*, pp.71-74.

Table 3.1 provides some descriptive statistics for the age of this sample of CAEL Assessment test takers. The majority of test takers are students applying for entrance to a university program. However, from time to time students who are still studying at the high school level take the test in order to assess their level of English language. This explains the minimum age of 14 years shown in the 1997 – 1998 column of Table 3.1. Note that in the 2003 – 2009 column, the minimum age is 17, which better reflects the expected population of CAEL Assessment test takers. In addition, a number of adults returning to the university setting to take courses of interest complete the CAEL Assessment. This explains the maximum age of 67 years reported in both samples. The mean age of between 26.6 and 27.2 years reflects the fact that the CAEL Assessment is used for both undergraduate and graduate students. It may also suggest that students with English as a second language requirements tend to be a little older than other university applicants.

Table 3.1: Descriptive Statistics for Age (Years) of CAEL Assessment Test takers

Statistic	Age (Years)	Age (Years)	Age (Years)
	1997 - 1999	2003 - 2009	2010 - 2012
Mean	26.6	27.2	27.14
Median	25.0	25.0	24.0
Mode	19.0	25.0	21.0
Minimum	14.0	17.0	14
Maximum	67.0	67.0	71

While Table 3.1 provides an overall impression of the age of the population, it is also useful to look at the frequency distribution of ages. Table 3.2 provides a summary of the frequencies observed for various age ranges in both sample populations. The majority of CAEL Assessment test takers (81.3%) are between 17 and 29 years of age in both groups.

Table 3.2: Frequency Distribution of Ages (Years) of CAEL Test takers

Age Range (Years)	Freq.	%	Freq.	%	Freq.	%
	1997 - 1999		2003 - 2009		2010 - 2012	
14-23	905	42.8	2466	31.5	2210	45.4
24-29	877	26.8	3900	49.8	1408	29
30-35	384	18.3	669	8.6	468	9.6
36-41	176	8.4	433	5.5	309	6.4
42-47	56	2.7	221	2.8	221	4.8
48-67	24	1.0	144	1.8	221	4.8
Total	2111	100%	7833	100 %	4837	100%

When we refer to the student status of a test taker, we are referring to the level of study the student has been accepted into at a Canadian university. 79.3% of the test takers at the main test centre are identified as undergraduate. Approximately 9% of the test takers are identified as *Special* students. This category includes students who are not registered in a degree program but wish to take courses of interest at the university. This category also includes those students who are interested in registering for ESL/EAP courses only, as well as those who have not yet applied or been accepted to a university. 7.5% of the students tested at the main test centre at Carleton University have been accepted into a graduate program at the university level. This number has grown by approximately 3% since 2003.

First Language

Test takers are asked to indicate their first language on the registration form. The information reported has been summarized as Table 3.3. Because of the great diversity in first languages, over 70 different first languages have been reported, Table 3.3 presents the frequency of test takers as a percent, for the 20 most commonly reported first languages. These 20 languages include approximately 88% of the test takers for the years of 1997 – 1999, 95% of the test takers for the years of 2003 – 2009, and 94.6% of the test takers for the years 2010 - 2011.

Table 3.3: Test takers Most Commonly Reported First Languages

1997 - 1999			
First Language	Freq.(%)	First Language	Freq.(%)
Arabic	19.0	French	2.2
Cantonese	12.9	Vietnamese	2.1
Mandarin	11.4	Somali	1.9
Korean	5.6	Amharic	1.5
Farsi (Persian)	5.6	Urdu	1.3
Tamil	5.0	Turkish	1.2
Spanish	4.8	Romanian	1.1
Japanese	4.0	Indonesian	1.0
Russian	3.2	Bengali	.8
Serbo-Croatian	2.3	Polish	.8
Total (20 most common languages) = 87.7%			

2003 - 2009			
First Language	Freq.(%)	First Language	Freq.(%)
Mandarin	47.9	Urdu	1.0
Cantonese	12.5	French	.9
Arabic	11.3	Taglog	.7
Farsi (Persian)	5.3	Tamil	.7
Spanish	2.2	Bengali	.7
Russian	2.0	Romanian	.7
Korean	2.0	Amharic	.3
Vietnamese	1.9	Indonesian	.3
Turkish	1.2	Portuguese	.3
Japanese	1.1	Serbo-Croatian	.3
Total (20 most common languages) = 95.3%			

2010 - 2012

First Language	Freq.(%)	First Language	Freq.(%)
Arabic	43.1	Punjabi	1.4
Cantonese	17.2	Tagalog	1.3
Farsi	6.5	Vietnamese	.8
Mandarin	4.0	Turkish	.8
Spanish	3.9	Portuguese	.7
English	3.4	Japanese	.7
Korean	3.0	Romanian	.7
Urdu	2.1	Bengali	.6
French	1.8	Hindi	.5
Russian	1.7	Thai	.4
Total (20 most common languages) = 94.6 %			

The first language most frequently reported by CAEL test takers during the last 2 years at the main test centre is Arabic (43.1%). In comparison, the other language groups are not well represented. However, it is important to recognize that these distributions reflect test takers completing their test at the main test centre at Carleton University. Of course, when testing is undertaken in overseas locations the composition of first languages reported differs greatly.

Country of Origin

The country of origin reported by CAEL test takers is summarized in Table 3.4. To date, test takers at the main test centre have reported over 112 different countries of origin. Table 3.4 shows the frequency of test takers as a percent for the 20 most commonly reported countries of origin. The largest single country of origin is still China although the percentage has jumped from 16.3% to 58.2% (statistics for Hong Kong are reported separately). There has also been some shift in the 20 most commonly reported countries of origin which now account for about 87% of the test takers at the main test centre. In the last two years, however, there has been a slightly shift where the amount of test takers from Saudi Arabia has increased considerably.

**Table 3.4: Test takers Most Commonly Reported Countries of Origin
1999 - 1997**

Country	Freq (%)	Country	Freq (%)
China	16.3	Vietnam	2.2
Korea	5.6	Russia	2.2
Iran	5.6	Lebanon	2.1
Sri Lanka	5.5	Saudi Arabia	2.1
Hong Kong	4.6	India	1.9
Japan	4.0	Ethiopia	1.8
Taiwan	2.5	Yugoslavia	1.7
Kuwait	2.4	Jordan	1.6
Somalia	2.4	Palestine	1.6
Egypt	2.3	Pakistan	1.5
Total (20 most common countries) 69.9%			

2003 - 2009

Country	Freq (%)	Country	Freq (%)
China	58.2	Lebanon	1.1
Iran	5.1	Palestine	1.1
Saudi Arabia	2.3	Japan	1.0
Korea	2.0	Hong Kong	1.0
Vietnam	1.9	India	.9
Russia	1.2	Philippines	.9
Jordan	1.2	Libya	.9
Turkey	1.2	Sri Lanka	.9
Pakistan	1.1	UAE	.8
Taiwan	1.1	Egypt	.8
Total (20 most common countries) 82.7%			

2010-2012

Country	Freq (%)	Country	Freq (%)
Saudi Arabia	35.3	Iraq	1.2
China	19.7	Colombia	1.1
Iran	6.6	Egypt	1.1
Canada	3.6	Taiwan	.9
Korea	3	Turkey	.8
India	2.6	Vietnam	.8
Pakistan	2.3	Mexico	.8
Philippines	1.9	Jordan	.7
Libya	1.7	Japan	.7
Russia	1.3	Romania	.6
Total (20 most common countries) 86.7%			

Summary

The purpose of this chapter has been to describe the characteristics of typical CAEL test takers. Test takers complete the CAEL Assessment for a wide variety of reasons, and this means that their characteristics are quite diverse. Information is collected from test takers at the time of registration in order to help the test developers monitor test performance and determine the composition of pilot test groups. The data presented in this chapter were collected at the main test centre at Carleton University. In general, we can say that our test population is 60% male, 40% female. Most of the test takers are between 20 and 30 years of age and are applying for undergraduate study. The test population is largely composed of individuals whose first language is Mandarin, Cantonese, or Arabic. Most of the test takers have come from East Asia, the Middle East, or North Africa of course, the characteristics of test takers completing the CAEL Assessment in overseas locations are quite different.

Chapter Four: Scoring Criteria, Methods and Reliability

Introduction

In this chapter we present a detailed description of the scoring criteria, methods and reliability of the CAEL Assessment. A team composed of test developers and experienced ESL/EAP teachers at the Language Assessment and Testing Research Unit (LATRU) in the School of Linguistics and Applied Language Studies at Carleton University developed the score criteria for each sub-test of the CAEL Assessment. A concise but comprehensive description of the development of the CAEL Assessment can be found in Fox, Pychyl, & Zumbo (1993). The score criteria and methods are continually monitored and occasionally revised in order to maximize both efficiency and reliability. The score criteria are provided for each sub-test as well as for the Overall Result as a set of band scores (10 - 90) which range from *Very Limited Proficiency* to *Expert Proficiency*. A detailed description of the standard procedures used in scoring every CAEL Assessment is also presented. Since the scoring procedure is a little different for each sub-test, the score criteria and methods are presented separately for each sub-test of the CAEL Assessment (Writing, Listening, Reading, and Speaking) and then for the Overall Result.

In the *Standards for Educational and Psychological Testing*, reliability is defined as "The degree to which test scores are consistent, dependable, or repeatable, that is, the degree to which they are free of errors of measurement" (AERA, APA, & NCME, 1985). CAEL test developers endeavour to minimize all possible sources of error in the administration of the test and in the scoring and reporting of test results. Test developers recognize that the unique features of the CAEL Assessment necessitate the gathering of some very specific types of reliability evidence. Given that the CAEL Assessment is comprised principally of constructed-response items and that the score interpretations are criterion-referenced, the consistency of the scoring procedures is one important aspect of the reliability of this test. Constructed-response items require a greater degree of subjective rating than do selected response item types such as multiple choice items. For this reason, inter-rater reliabilities are routinely calculated for each of the CAEL sub-tests. Measures of the internal consistency of the test items (i.e., split-half reliabilities), often cited in test manuals, are not really meaningful given the test specifications, item types, and interpretations made from CAEL Assessment results (for a detailed discussion of this issue see Zumbo, Fox & Pychyl, 1993). Accordingly, this chapter presents the inter-rater reliability evidence that is collected on an on-going basis for each sub-test of the CAEL Assessment. The second unique feature of the CAEL Assessment that necessitates the collection of reliability evidence is the test is a topic-based integrated assessment available in a number of different versions. In accordance with the principles of fair and ethical testing as defined by the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1985), the CAEL test development team is committed to ensuring the comparability of these different test versions. This chapter presents evidence from a number of sources, which demonstrates the consistency of scores across different test versions.

Writing Performance - Score Criteria, Methods and Reliability

The writing band scores that have been developed for the CAEL Assessment are the result of a collaborative effort between CAEL test developers and experienced ESL/EAP teachers. The theoretical background knowledge and depth of experience of these ESL/EAP teachers is an invaluable resource for the test development team. These teachers are well versed in the curriculum and teaching methods for courses ranging from intensive English language courses for students with little or no prior English language training to the final stages of advanced academic English language courses. Test developers at LATRU work closely with the ESL/EAP teachers to develop pilot testing procedures that include students at various levels of English language proficiency. Pilot testing includes students from the full range of ESL/EAP course levels, as well as students admitted to first-year university courses who have already satisfied Carleton University's English language requirements. Some pilot testing is also conducted with students at the senior high school level and at other universities and colleges within Canada and abroad. This extensive pilot testing helps test developers to ensure that the full range of academic writing proficiency described in the score criteria is adequately sampled in the test development process. The collaboration between the ESL/EAP teachers and the test development team has resulted in well developed and articulated criteria for the Writing Performance sub-test of the CAEL Assessment.

The score criteria for Writing Performance is presented in Table 4.1. The band scores encompass a wide range of writing proficiency - from the *Very Limited Writer* (10-20) to the *Expert Writer* (80-90). These band score descriptions are included with the CAEL Assessment Score Report which is given to every CAEL test taker. This report provides the test taker with a description of their own performance as well as a description of the full range of band scores. This information is a valuable resource for CAEL test takers and the institutions which receive CAEL Assessment Score Reports.

Table 4.1: Writing Performance Band Score Criteria

Band Score	Score Criteria
10-20	<p>Very Limited Writer</p> <ul style="list-style-type: none"> • is generally unable to express ideas effectively in writing • uses very restricted and/or ungrammatical language • uses words randomly and without overall coherence
30	<p>Limited Writer</p> <ul style="list-style-type: none"> • attempts to write something which is related to the topic but the writing is not predictable • uses restricted and/or ungrammatical language • seems to understand the topic, but is unable to develop ideas because language constrains or distorts expression
40	<p>Marginally Competent Writer</p> <ul style="list-style-type: none"> • makes links among ideas and addresses the topic but the writing lacks clarity and cohesiveness • displays elements of control in the writing (e.g. a thesis statement, an introduction and conclusion) but internal coherence is lacking • uses little or no support (i.e. quotations, examples, etc.) to develop the thesis
50	<p>Competent but Limited Writer</p> <ul style="list-style-type: none"> • addresses the topic to a degree but with somewhat limited clarity and cohesiveness • uses some support to develop the thesis • control of the argument is limited by poor comprehension of the readings and lecture, and/or poor understanding of the requirements of academic writing
60	<p>Competent Writer</p> <ul style="list-style-type: none"> • develops a thesis using a range of support • uses language that is generally accurate but is constrained by a somewhat limited vocabulary • demonstrates general understanding of the requirements of academic writing
70	<p>Adept Writer</p> <ul style="list-style-type: none"> • responds readily to the demands of the topic and presents information clearly and logically • uses the readings and lecture effectively to support the thesis • demonstrates understanding of the requirements of academic writing
80-90	<p>Expert Writer</p> <ul style="list-style-type: none"> • demonstrates mastery of appropriate, concise, and persuasive academic writing • writes with authority and style

The scoring method used for the Writing Performance sub-test of the CAEL Assessment was developed by test developers within LATRU (see Fox & Soucy 1991) and is referred to as the "collaborative read-aloud marking protocol." Earlier experience with traditional methods of the holistic assessment of writing (Diederich, 1967) highlighted the impact of marker fatigue on the consistency of ratings (Fox & Soucy, 1991). CAEL test developers sought to devise a rating system that would minimize factors such as marker fatigue in the rating of the Writing Performance sub-test. Research investigating many factors that influence the consistency of ratings for writing was carefully reviewed and used as a guide in the development of the collaborative read-aloud marking protocol. The resulting protocol is a unique procedure that contributes to both the reliability and validity of the CAEL Assessment. For this reason, the procedure described in some detail is included in this chapter.

The writing samples produced by CAEL test takers in response to the essay prompt are typically essays of one to three pages in length. CAEL Assessment essays are marked 'blind'. That is, essays are identified using only the test taker's identification number and initials. Raters are not aware of the specific characteristics (gender, first language etc.) of the test taker. All CAEL Assessment Writing Performance raters are experienced ESL/EAP teachers who have received specific training in the marking of CAEL Assessment essays. Version-specific marking guide descriptors and sample essays are used in the training of new raters and are also available to the raters throughout the rating process.

A team of three trained raters meets to begin the marking protocol. The first essay is selected and one rater is chosen to read the essay aloud. Raters are trained to read the essays in as smooth and fluent a manner as possible without revealing in any way (by changing inflection in the voice, editorial comments, etc.) what the reader thinks of the piece being read. The writing samples are read aloud in order to minimize the impact of surface features in the text such as spelling and handwriting on the score assigned. The impact of these surface features has been researched extensively and can be substantial. Raters are encouraged to place greater emphasis on other more substantial features (e.g., meaningfulness, cohesiveness, flow, etc.) of the essays and to discount these surface features when assigning a score. After listening to the essay, each rater records brief descriptive comments indicating their impression of the content, language and organization of the writing sample. Each individual rater also records a score for the writing sample. At this time, individual raters may ask for the writing sample to be read again or they may ask to see the piece of writing. After each individual rater has recorded their mark, the marks are revealed. If there is consensus on the marks, then the raters go on to the next paper. If there is a difference in the scores assigned, a discussion begins. In this case raters describe their rationales for the marks they have given and a consensus is reached. When the consensus has been reached, the final writing band score is recorded for the test taker. This collaborative marking process results in a consistent application of the score criteria that have been developed for the CAEL Assessment. Rarely an essay is encountered for which the original marking team cannot reach a consensus. When this occurs a second team of raters is asked to review the writing sample.

The inter-rater reliability of this process is constantly monitored. A number of previously scored writing samples are re-marked at every rating session. The correlation between the first and second marking occasions is calculated. Because the scale used in marking the writing samples is an ordinal scale, a Spearman rank-order correlation co-efficient (ρ) is reported. The most recent correlation co-efficient is .962 (n=178). Because the inter-rater reliability data is collected on an ongoing basis, a number of teams of raters and all current test versions are included in this sample.

In addition to Spearman's ρ , the distribution of score changes for these 178 writing samples is shown in Table 4.2. The frequency of scores changed is shown as a percent. No change indicates that the band score assigned by the second team of raters was exactly the same as the band score assigned by the original team of raters. A change of one indicates that the band score assigned by the second team of raters was either one band score above or one band score below the band score assigned by the original team of raters. A similar description applies for the changes of 2 and 3 band scores. Table 4.2 indicates that 87.1 % of the writing scores are exactly the same across the two groups of markers, or they are within one band score of each other. Given the number of factors influencing the rating of essays, as reported in a number of research studies, CAEL test developers are confident that the reliability of the collaborative read-aloud marking protocol meets or exceeds the reliability of essay rating processes used in other contexts.

Table 4.2: Writing Performance Band Score Changes (n=178)

Band Score Difference	Freq. (%)
No change	87.1
Change of 1 band score	10.1
Change of 2 band scores	2.2
Change of 3 band scores	0.6

Listening Performance - Score Criteria, Methods and Reliability

The score criteria for the Listening sub-test are provided in Table 4.3. These criteria have been developed collaboratively by test developers and ESL/EAP teachers at LATRU in much the same manner as they were for writing performance, see description above. The principal difference between establishing the Listening Performance criteria and the criteria for Writing Performance lies in the fact that an initial raw score is calculated when the Listening Performance sub-test is marked. Raw scores are not generated for the Writing Performance. The raw scores for each version of the CAEL Assessment are then converted into standard band scores on the basis of the results from the pilot testing process. As is the case for the Writing sub-test, extensive pilot testing of students across a broad range of English language proficiency is conducted. This procedure enables test developers to make accurate and consistent conversions of the raw scores generated from each version of the test into standard band scores.

The band scores shown in Table 4.3 reflect a wide range of listening proficiency from the *Very Limited Listener* (10-20) to the *Expert Listener* (80-90). As with the Writing Performance sub-test, these band score descriptions are included in the CAEL Assessment Score Report which is

given to CAEL Assessment test takers. This report provides the test taker with a description of their own Listening Performance as well as a description of the full range of band scores.

Table 4.3: Listening Performance Band Score Criteria

Band Score	Score Criteria
10-20	Very Limited Listener Demonstrates very limited comprehension of lectures takes some meaning from individual words overall understanding is sketchy and random
30	Limited Listener Demonstrates limited and inconsistent comprehension of lectures makes sense of some sections of lectures by guessing overall understanding is limited
40	Marginally Competent Listener Demonstrates uneven comprehension of lectures is able to identify the meaning of some unfamiliar terms overall understanding is restricted
50	Competent but Limited Listener Demonstrates somewhat limited comprehension of lectures is able to process most lecture sections for general ideas, but misses or misinterprets specific details from time to time overall understanding is still somewhat restricted
60	Competent Listener is able to understand information regarding both main ideas and supporting details in lectures may lack some flexibility and miss some information compensates at times for missed information by drawing on overall understanding of what is being said
70	Adept Listener is able to understand lectures with apparent ease compensates strategically for unfamiliar vocabulary or terminology overall understanding is flexible and consistent
80-90	Expert Listener Demonstrates comprehension of lectures which is equal to that of experienced academic listeners Understands both main ideas and supporting details with ease is fully engaged by and interacts with the information being presented

Scoring of the listening component is conducted by trained raters using detailed marking keys. The marking keys provide explicit examples of acceptable responses and indicate the number of points to be allotted for each response. The marking keys are occasionally revised to reflect the range of responses that are provided by CAEL test takers. Partial scoring including half marks is allowed for all of the constructed-response items. A raw listening score is determined by summing the points for each of the items on this test component. The raw listening score obtained on each version of the CAEL Assessment is converted into a band score so that CAEL Assessment scores can be compared across versions.

Scoring of the Listening Performance sub-test of the CAEL Assessment is more objective than the scoring of the Writing Performance. However, a certain amount of rater subjectivity is always involved in the rating of constructed-response test items. For this reason, the inter-rater reliability of the Listening Performance is constantly monitored on a version-by-version basis. A number of listening tests are re-marked by individual raters at regular intervals. Because the raw scores are converted to a band score which is considered to be an ordinal scale of measurement, a Spearman rank-order correlation co-efficient (ρ) correlation co-efficient is calculated as a measure of the extent to which the two groups of raters provide similar scores. The most recent investigation of the inter-rater reliability of the Listening Performance sub-test resulted in a Spearman rank-order correlation co-efficient (ρ) of .973 (n=178). With regard to the distribution of band score changes, 93.8% of the band scores remained the same. Given the consistently high correlations that result from this process, CAEL test developers are confident that the scoring procedures for the Listening Performance sub-test are reliable.

Reading Performance – Score Criteria, Methods and Reliability

The score criteria for Reading Performance are provided in table 4.4. The development of these criteria is very similar to the process described above for the Listening Component. As with the Listening Component, raw scores for Reading Performance have been converted into standard band scores on the basis of the results of extensive pilot testing involving students with a wide range of English language proficiency. The score criteria range from the *Very Limited Reader* (10-20) to the *Expert Reader* (80-90).

Once again, these band score descriptions are included in the CAEL Assessment Score Report which is given to every CAEL test taker. This report provides the test taker with a description of their own Reading Performance as well as a description of the full range of band scores.

As described earlier, each CAEL Assessment typically includes two readings, both of which are related to the topic of the test. Responses to the items for each reading are scored using detailed scoring keys in much the same manner as the scoring for the Listening Performance sub-test. The raw scores for the two readings are added together and the total raw score for Reading Performance is then converted to a band score.

Table 4.4: Reading Performance Band Score Criteria

Band Score	Score Criteria
10-20	Very Limited Reader is unable to read effectively takes some meaning from pictures, titles, random words, etc. may understand the main idea at times but misses almost all of the supporting details
30	Limited Reader reads with limited accuracy and fluency reads with some understanding of the main ideas but is unable to identify specific, relevant details is often unable to identify the meaning of unfamiliar terms from context

40	Marginally Competent Reader is unable to understand main ideas is restricted by limited vocabulary and a lack of familiarity with textural conventions reads more slowly than most academic readers
50	Competent but Limited Reader reads with understanding of the main ideas and is able to identify some relevant details reads more slowly and with greater effort than most academic readers may misinterpret information at times
60	Competent Reader Understands main ideas and is able to identify most relevant details reads more slowly and with greater effort than some academic readers is able to interpret information with some flexibility
70	Adept Reader reads academic texts with ease provided sufficient time is available demonstrates comprehension of academic texts which approaches that of experienced academic readers interprets information with flexibility
80-90	Expert Reader reads academic texts with ease demonstrates comprehension of academic texts which is equal to that of experienced academic readers understands both main ideas and supporting details with ease

The inter-rater reliability of the scoring for this section of the test is also monitored by re-marking a certain number of tests. After the raw score is converted into the band score, a Spearman rank order correlation co-efficient is calculated on the band scores from the two raters. The most recent investigation of the inter-rater reliability of the Reading Component resulted in a Spearman rank order correlation co-efficient of .949 (n=178). Frequency distribution of the number of band changes show that 91.6% of the remarks resulted in the same band score. Based on this information, CAEL test developers are confident that the scoring methods used for the Reading Performance produce reliable results.

Speaking Performance - Score Criteria, Methods and Reliability

Test developers at LATRU and ESL/EAP teachers at the School of Linguistics and Language Studies developed the score criteria for the Speaking Performance in much the same manner as for the other components of the CAEL Assessment. As with the Listening and Reading Performance, a raw score is initially calculated and is then converted into the standard band scores. This conversion is conducted on the basis of results from extensive pilot testing of test takers with a wide range of speaking proficiency.

The score criteria are shown in table 4.5 and encompass a broad range of speaking proficiency from the *Very Limited Speaker* (10-20) through to the *Expert Speaker* (80-90). This information is included in the CAEL Assessment Score Report provided to every test taker who has taken the Oral Language Test.

As described earlier, the Oral Language Test component of the CAEL Assessment is a tape-mediated process. The test takers responses for each item in the Oral Language Test are tape-recorded. The tape-recorded responses are then rated by a trained rater using a detailed (analytic) scoring criterion. The raw score obtained from this procedure is converted into a band score as for all components of the CAEL Assessment.

The inter-rater reliability of this procedure is monitored in essentially the same manner as for the other sub-tests. At regular intervals a selection of test taker response tapes are re-marked, the scores are converted to band scores, and a Spearman rank order correlation co-efficient is computed between the two sets of scores. The most recent correlation for a sample of n=178 oral language tests resulted in a correlation of .952. CAEL test developers are confident that the scoring methods used for the Oral Language Test produce reliable results.

Table 4.5: Speaking Performance Band Score Criteria

Band Score	Score Criteria
10-20	Very Limited Speaker speaks with great difficulty and many long pauses mispronounces many words manages to communicate some information
30	Limited Speaker speaks with some difficulty; hesitations or false starts mispronounces some words searches for words or provides studied and careful responses
40	Marginally Competent Speaker speaks with some fluency but without flexibility speed of response (either too fast or too slow) sometimes limits communication communicates information adequately but with noticeable effort
50	Competent but Limited Speaker speaks with some fluency and flexibility speaks unevenly – at times there is a natural and easy quality to the response and at other times the response breaks down
60	Competent Speaker speaks fluently, flexibly and with a degree of ease compensates strategically for limitations communicates most required information clearly
70	Adept Speaker speaks with ease presents information clearly and logically communicates required information effectively
80-90	Expert Speaker speaks with authority on a variety of topics demonstrates flexibility and controls nuance speaking is characterized by spontaneity and comprehensibility

Overall Result - Score Criteria & Methods

The Overall Result provided to CAEL test takers is neither a summation nor an average of the four sub-tests. Rather, a placement team meets to consider the entire score profile as well as factors such as the performance in specific sub-tests before assigning an Overall Result⁴.

The band scores for the Overall Result are listed in table 4.6 along with a description of the meaning of each Overall Result. This information is provided to test takers on the CAEL Assessment Score Report. The Overall Result criteria reflect the standards of proficiency which are accepted at Canadian universities who use the CAEL Assessment. Thus, the Overall Result criteria do not offer a description of English language performance such as is provided for the individual sub-test criteria.

Table 4.6: Overall Result Band Score Criteria

Band Score	Score Criteria
10-40	needs to increase the level of academic English before admission requirements for Canadian University degree programs are met
50	may meet academic English language requirements for admission to a few Canadian degree programs
60	meets academic English language requirements for admission to some Canadian University degree programs
70-80	meets academic English language requirements for admission to most Canadian University degree programs
80-90	meets academic English language requirements for admission to Canadian University degree programs

As indicated above, the overall score on the CAEL Assessment results from a review of all of the evidence collected about a test taker's ability to use English for academic purposes during the CAEL testing procedure. This includes:

- performance on the Oral Language Test,
- performance on the CAEL sub-tests of reading, listening and writing,
- responses to the personal essay and self-assessment, and
- other information which has been shown to have an effect on student performance in academic settings (e.g., amount of study in an English-medium school, intended program of study, intended level of academic study, etc.)

⁴ Overall Results are closely tied to the proficiency levels of the ESL/EAP courses at Carleton University. The criteria of the scale map directly onto the continuum of proficiency which is denoted by course levels. In making decisions related to the overall score, the placement team considers which course would most closely match the test taker's English language learning needs. The link between language learning and proficiency is direct and contributes to validity arguments based on construct representation (Messick, 1989.)

The CAEL Assessment review procedure is undertaken by the Testing Coordinator, the Coordinator of the Intensive English (ESL), the Academic English (EAP) Programs Coordinator and/or the CAEL Head Administrator. Because the review procedure involves judgement, the reliability of the procedure is constantly monitored. On a regular basis, 90-100 randomly selected and previously evaluated score summaries are re-evaluated by the CAEL Assessment review committee and a final score re-calculated. The correlation between the overall scores awarded at the first and second occasions is calculated.

Using a Spearman rank-order correlation co-efficient (ρ), the most recent re-evaluation of 178 score summaries results in a correlation of .948. The distribution of score changes for these 90 scores is shown in Table 4.7 below. The frequency of score changes is indicated as a percent. No change indicates that the band score assigned by the second review committee was the same as the band score assigned by the original committee. A change of one indicates that the band score assigned by the second review committee was either one band score above or one band score below the original band score assigned by the first committee. There were no changes greater than one band score above or below the original band score. CAEL test developers are confident that the CAEL Assessment review procedure demonstrates adequate reliability.

Table 4.7: Overall Performance Band Score Changes (n=178)

Band Score difference	Frequency: Number and (%)	
No change	(n=148)	83.1%
Change of 1 band score	(n=29)	16.3%
Change of 2 band scores	(n=1)	0.6%

Comparability of Test Versions

For security reasons test versions are administered on a rotating basis. Although no two versions of any test may be considered perfectly comparable in all respects, versions of the CAEL Assessment are developed in a manner that meets recognized standards for comparability and satisfies principles of fair and ethical testing practice. Evidence of the comparability of the various versions of the CAEL Assessment is continually being gathered.

Evidence demonstrating comparability can be found from four sources:

- an examination of test development procedures;
- a statistical analysis of the score distributions for each version of the test;
- a statistical analysis of the scores of test takers who have taken more than one version of the test in a short time period; and
- specific research studies designed to investigate the comparability of the test versions.

Each of these sources of reliability evidence is described in this chapter.

In terms of test development, a number of procedures have been adopted in an effort to ensure the comparability of the test versions. First and foremost in this process was the development of test specifications for each component of the CAEL Assessment. These test specifications were

developed through an iterative, consensus-based method involving both test developers at LATRU and experienced ESL/EAP teachers at the School of Linguistics and Language Studies. The use of these test specifications maintains comparability across versions of the CAEL Assessment both in terms of the specific tasks contained in the test and in terms of the overall domain coverage of each sub-test.

Another test development procedure that helps to ensure comparability is the pilot testing process. Through extensive pilot testing, the performance of students with a wide range of proficiency on a test version under development can be compared with the performances of a similar group of students on an existing version of the test. The results of this comparison may be used to adjust individual test items that do not seem to be discriminating among test takers in the expected manner. As a final step, before releasing a new test version, test takers may be randomly assigned to complete one of two test versions and the resulting score distributions can be compared. Any difference in the performance of the two versions can be corrected before the new version is released for use at a test centre.

A statistical analysis of the score distributions is conducted on a version-by-version basis at regular intervals to ensure that all test versions function in a comparable manner. Given that there are currently 14 different versions of the test in use, it is not feasible to present all possible comparison distributions in this manual. However, the distributions can be compared for a subset of test versions. In Table 4.8, the Overall Result distributions are provided for four different versions of the CAEL Assessment.

These four versions were selected because, in every case, there were more than 400 recent Overall Result scores available. For reasons of test security, the versions are simply identified as A, B, C, and D. The Overall Results displayed in Table 4.8 represent CAEL test takers who completed the test at the main test centre at Carleton University in the period from January 2006 through September 2009.

Table 4.8: Frequency (%) of Overall Result Band Scores for Four Test Versions

Overall Result Band	Version A (n=434)	Version B (n=704)	Version C (n=762)	Version D (n=818)
10	1.4	2.8	1.6	2.3
20	16.1	16.9	15.9	14.3
30	30.0	26.4	30.7	34.5
40	27.6	24.0	28.5	27.8
50	15.2	18.3	12.1	8.9
60	5.3	7.8	7.1	6.4
70	4.4	3.3	3.9	5.0
80	-	-	.3	.9
90	-	-	-	-
Total	100	100	100	100

An examination of Table 4.8 reveals that, while there are some fluctuations in the band scores, the distributions for the Overall Result follow essentially the same pattern for each of the test versions.

Further evidence of the comparability of the different test versions can be found in an examination of the results of test takers who take two different versions of the test over a relatively short time period. Apart from test takers who agree to participate in research studies, it is difficult to obtain data from actual test takers who have taken two different versions of the test. The majority of test takers only complete one version of the test. For those who complete two or more versions of the test there is often a long time delay and considerable language learning which occurs between the two testing occasions. However, we do have one source of data available to us from routine test administrations. Occasionally individual test takers make a request to the test administrator to re-write the CAEL Assessment. The test taker may feel that their performance on their initial test day was adversely affected by fatigue, illness, a recent arrival in Canada, or the particular version of the test which they wrote. Each of these requests is reviewed on an individual basis before permission to re-write the test is granted. If permission is granted to re-write the test then the test taker may complete a second CAEL Assessment. We track the performance of test takers who complete re-writes and the information which is gathered can shed some light on the reliability of the different test versions.

An analysis of the score results from those test takers who requested re-writes is reported in this chapter. The sample includes all test takers who wrote two different versions of the CAEL Assessment within a maximum period of four weeks. We have not included test takers for whom the interval is greater than four weeks because it is quite likely that a substantial amount of language learning could occur as the time interval between the test dates increases. For some test takers the interval between test dates is as short as one week. For others, an interval of two, three, or four weeks occurred. Table 4.9 offers a summary of the differences found in the Overall Result between the first and second testing occasion for these test takers (n=107). This data was collected over the period from January 1997 through to September 1999 and includes all of the test versions currently in use.

Table 4.9: Frequency of Band Score Changes in Overall Result

Overall Result - Band Score Change	Freq	%
2 band scores lower	6	5.6
1 band score lower	7	6.5
No change	47	43.9
1 band score higher	30	28.0
2 band scores higher	17	15.9
Total	107	100

It is clear from an examination of Table 4.9 that about 44% of test takers who take a second version of the test within a four-week time period score exactly the same in terms of their Overall Result. Further, 78.4% of the test takers have a final result within one band score of their original Overall Result. While 28% of the test takers score at one band score higher on the second testing experience, these results are essentially to be expected. There are three principal reasons for this modest gain in performance. First, all of the test takers in this sample were dissatisfied with their initial performance. That is, these test takers requested a re-write because they felt that they had been adversely affected by factors such as fatigue or illness during their first test experience. This means that we can expect some improvement in their test scores because they were feeling better at the time of the second testing. The second factor which undoubtedly accounts for some of the improvement in these test scores is termed a 'practice effect'. That is, the test takers were much more familiar with the CAEL Assessment testing process. This familiarity is likely to result in a reduction in test anxiety and an improvement in test performance. The third factor which may account for the improvement in test scores is that some language learning has occurred between the first and second testing occasions. Very few test takers, only 6.5% found a reduction of one band score in their Overall Result on the second testing occasion. This analysis of test takers who have requested to write a second version of the test provides an additional source of evidence for the comparability of the different test versions currently in use.

More evidence of the reliability of the various versions of this test can be found in specific research studies which have been undertaken with the CAEL Assessment. An early study which compared the performance of test takers across two versions of the CAEL Assessment is reported in Zumbo, Fox, & Pychyl (1993). In this study the performance of 80 test takers who completed two versions of the test within a three week time period is examined. Measures of reliability, which are suitable for use in criterion-referenced testing situations, were computed. The results indicated that the decision consistency of the Overall Result for the CAEL Assessment was very high.

Jennings, Fox, & Shohamy (1999) investigated the potential existence of a topic-effect for the CAEL Assessment. In this study test takers (n=254) were randomly assigned to two groups. One group was given the opportunity to choose from among five versions of the CAEL Assessment. The other group was given the regularly scheduled version of the test. A detailed explanation of the experimental design and the analysis of the resulting data is available in the article (see Jennings et al, 1999). The results demonstrated that there was no effect of topic on the scores of the test takers. That is, the performance of the test takers that were given a choice of test version did not differ from that of those who had no choice. This evidence supports the contention made by CAEL test developers that the versions of the CAEL Assessment function in a comparable manner.

Chapter Five: Score Distributions

Introduction

Score distributions have been generated for the population of test takers who completed the test at the main test centre located at Carleton University during the period from January 1997 through September 1999 (Table 5.1.a) and from January 2003 through January 2009 (Table 5.1.b). Score distributions are reported for the Overall Result as well as for each sub-test (Writing, Listening, Reading & Speaking). The number of test takers achieving each band score is shown both as a frequency count and as a percent. Summary statistics in the form of means and standard deviations are also provided in order to describe both the central tendency and extent of variance in the test taker population.

These score distributions are presented in an effort to provide a clear description of the performance of typical CAEL test takers. These distributions should not be interpreted as representative of norms for use at other institutions. The performance of test takers at any institution will be strongly influenced by factors such as the background characteristics of the test takers, the amount and type of English language training, and the level and intended course of study of the test takers. For this reason, the score distributions of test takers at other test centres may vary considerably from the scores reported here.

Distribution of Overall Results

Table 5.1.a presents the 1997-1999 Overall Results for the full sample and also offers a comparison by gender. Most of the test takers tend to cluster around the middle range of proficiency. The most frequently occurring Overall Result is a band score of 50 with 56% of test takers achieve a band score of 40 or 50. There are relatively few test takers (7.6%) scoring in the lowest band scores of 10 and 20. This distribution likely reflects the fact that CAEL test takers generally have some previous English language training prior to applying to study at an English language university.

Table 5.1.b presents the 2001-2008 Overall Results for the full sample and by gender. In this group, the most frequently occurring Overall Result is band score 30 (34.4%) with the majority of test takers achieving an Overall Result of a band score of 30 or 40 (58.5%). 22% of test takers fall into the lowest band levels of 10 and 20. This change in distribution possibly reflects the significant change in country of origin and the fact that more students are beginning their studies at English language universities with little or no previous English language training.

Table 5.1.c presents the 2009-2012 Overall Results for the full sample and by gender. In this group, the most frequently occurring Overall Result is band score 50 (28%) with the majority of test takers achieving an Overall Result of a band score of 40, 50 or 60 (67.6%). 6.7% of test takers fall into the lowest band levels of 10 and 20. This change in distribution possibly reflects the significant change in country of origin and the fact that more students are beginning their studies at English language universities with little or no previous English language training.

The highest Overall Results achieved by test takers in the three cohorts at the main test centre are at the 70, 80, and 90 band levels. The number of test takers to achieve this level has dropped from 8.6% in the 1997-1999 cohort to 2.6% in the 2003-2009 cohort but increased to 18.7% in the 2010-2012 cohort. It is these test takers who have achieved a sufficient level of English language proficiency to register for full-time courses at the university without further ESL/EAP training. Test takers who achieve a band score of 40 or below require intensive ESL training before they can begin their university studies. This cut point is currently under review at Carleton University and may be lowered to 30. Presently at Carleton University test takers who achieve band scores of 40, 50 or 60 may be permitted to take between one and three university courses in addition to their ESL/EAP courses. Again, the lower cut point is under review and may be lowered to band score 30. This gradual admission policy is unique to Carleton University and does not reflect the English language proficiency requirements at other universities which accept CAEL Assessment scores. A full description of the Overall Results of band scores is provided in the chapter entitled *Scoring Criteria, Methods & Reliability*.

Table 5.1.a: Overall Results for Full Sample and by Gender 1997-1999

*Overall Result Band Score	Full Sample		Males		Females	
	Freq	%	Freq	%	Freq	%
10	28	1.3	15	1.2	12	1.4
20	133	6.3	87	6.9	46	5.4
30	208	9.8	122	9.6	86	10.1
40	568	26.8	347	27.4	221	26.0
50	620	29.2	384	30.3	236	27.7
60	383	18.0	212	16.7	171	20.1
70	174	8.2	95	7.5	77	9.0
80	9	0.4	6	0.5	2	0.2
Total	2123	100	1268	100	851	100

*information concerning gender was not available for some test takers

Table 5.1.b: Overall Results for Full Sample and by Gender 2003-2009

*Overall Result Band Score	Full Sample		Males		Females	
	Freq	%	Freq	%	Freq	%
10	167	2.1	120	2.7	46	1.4
20	1564	19.9	1018	22.7	539	16.4
30	2697	34.4	1604	35.7	1075	32.7
40	1892	24.1	1003	22.3	881	26.8
50	886	11.3	438	9.7	439	13.4
60	404	5.2	200	4.5	200	6.1
70	180	2.3	93	2.1	86	2.6
80	27	0.3	7	0.2	19	0.6
Total	7817	100	4494	100	3285	100

*information concerning gender was not available for some test takers

Table 5.1.c: Overall Results for Full Sample and by Gender 2010-2012

*Overall Result Band Score	Full Sample		Males		Females	
	Freq	%	Freq	%	Freq	%
10	50	1.0	47	1.6	3	0.1
20	278	5.7	245	8.3	33	1.7
30	554	11.4	411	14	143	7.6
40	942	19.4	627	21.2	315	16.8
50	1357	28.0	787	26.6	570	30.3
60	978	20.2	518	17.5	460	24.4
70	575	11.8	286	9.6	288	15.3
80	96	2.0	31	1.0	65	3.4
90	15	.3	8	0.2	7	0.4
Total	4845	100	2963	100	1889	100

*information concerning gender was not available for some test takers

Distribution of Writing Performance Scores

Tables 5.2.a, 5.2.b, and 5.2c present the distribution of writing scores observed for these populations. Both the frequency and percent of test takers scoring at each band score is provided for the full sample and by gender. As for the Overall Results, the majority of test takers score at band scores of 40 or 50 (46.3%) in the 1997-1999 cohort, at band scores of 20 or 30 (62.2%) in the 2003-2009 cohort, and at band scores of 40, 50 or 60 (66.5%). While females score marginally higher than males on the Writing Performance sub-test, the pattern of results is quite similar across cohorts.

Table 5.2.a: Writing Performance Band Scores for Full Sample and by Gender 1997-1999

*Writing Performance Band Score	Full Sample		Males		Females	
	Freq	%	Freq	%	Freq	%
10	34	1.6	23	1.8	11	1.3
20	248	11.7	147	11.6	101	11.9
30	344	16.2	202	15.9	142	16.7
40	539	25.4	336	26.5	203	23.9
50	443	20.9	271	21.4	171	20.1
60	309	14.6	170	13.4	139	16.3
70	86	4.1	54	4.3	29	3.4
80	116	5.5	63	5.0	53	6.2
90	4	0.2	2	0.2	2	0.2
Total	2123	100	1268	100	851	100

*information concerning gender was not available for some test takers

Table 5.2.b: Writing Performance Band Scores for Full Sample and by Gender 2003-2009

*Writing Performance Band Score	Full Sample		Males		Females	
	Freq	%	Freq	%	Freq	%
10	282	3.6	194	4.3	86	2.6
20	2603	33.2	1614	35.9	975	29.7
30	2277	29	1299	28.9	966	29.4
40	1224	15.6	635	14.1	585	17.8
50	708	9.0	385	8.6	315	9.6
60	366	4.7	183	4.1	179	5.4
70	216	2.8	107	2.4	105	3.2
80	119	1.5	54	1.2	64	1.9
90	22	0.3	12	0.3	10	0.3
Total	7817	100	4483	100	3285	100

*information concerning gender was not available for some test takers

Table 5.2.c: Writing Performance Band Scores for Full Sample and by Gender 2010-2012

*Writing Performance Band Score	Full Sample		Males		Females	
	Freq	%	Freq	%	Freq	%
10	98	2.0	87	2.9	11	0.6
20	268	5.5	221	7.5	47	2.5
30	603	12.4	448	15.1	155	8.2
40	1012	20.9	644	21.8	368	19.5
50	1263	26.0	720	24.3	543	28.8
60	939	19.6	524	17.7	415	22.1
70	506	10.4	246	8.3	260	13.8
80	137	2.8	59	2.0	77	4.1
90	19	0.4	11	0.4	8	0.4
Total	4845	100	2960	100	1884	100

*information concerning gender was not available for some test takers

Distribution of Listening Performance Scores

Tables 5.3.a, 5.3.b, and 5.3.c present the distribution of Listening Performance results for these samples. Both the frequency and percent of test takers scoring at each band score is provided for the full sample and by gender.

Table 5.3.a: Listening Performance Scores for Full Sample and by Gender 1997-1999

Listening Performance Band Score	Full Sample		Males		Females	
	Freq	%	Freq	%	Freq	%
10	71	3.3	45	3.5	26	3.0
20	190	8.9	107	8.4	83	9.7
30	262	12.3	161	12.6	101	11.8
40	429	20.1	260	20.4	169	19.7
50	587	27.5	362	28.4	225	26.3
60	453	21.2	266	20.9	185	21.6
70	90	4.2	51	4.0	39	4.6
80	42	2.0	19	1.5	21	2.5
90	10	0.5	3	0.2	7	0.8
Total	2134	100	1274	100	856	100

*information concerning gender was not available for some test takers

Table 5.3.b: Listening Performance Scores for Full Sample and by Gender 2003-2009

Listening Performance Band Score	Full Sample		Males		Females	
	Freq	%	Freq	%	Freq	%
10	475	6.1	308	6.9	164	5.0
20	1271	16.2	798	17.8	467	14.2
30	1793	22.9	1055	23.5	727	22.1
40	1673	21.4	936	20.9	723	22.0
50	1648	21.1	895	20.0	742	22.6
60	751	9.6	403	9.0	345	10.5
70	142	1.8	63	1.4	79	2.4
80	59	0.8	24	0.5	34	1.0
90	5	0.1	1	0.0	4	0.1
Total	7817	100	4483	100	3285	100

*information concerning gender was not available for some test takers

Table 5.3.c: Listening Performance Scores for Full Sample and by Gender 2010-2012

Listening Performance Band Score	Full Sample		Males		Females	
	Freq	%	Freq	%	Freq	%
10	95	2.0	85	3.0	10	0.5
20	335	6.9	274	8.3	61	3.2
30	678	14.0	483	16.5	195	10.3
40	945	19.5	602	20.5	343	18.2
50	1188	24.5	714	24.3	474	25.1
60	1065	21.8	565	19.2	500	26.5
70	396	8.2	186	6.3	209	11.1
80	110	2.3	42	1.4	68	3.6
90	38	.8	10	0.4	28	1.5
Total	4850	100	2961	100	1888	100

*information concerning gender was not available for some test takers

Distribution of Reading Performance Scores

Table 5.4.a, 5.4.b, and 5.4.c present the distribution of Reading Performance results for these samples. Both the frequency and percent of test takers scoring at each band score is provided for the full sample and by gender.

Table 5.4.a: Reading Performance for Full Sample and by Gender 1997-1999

Reading Performance Band Score	Full Sample		Males		Females	
	Freq	%	Freq	%	Freq	%
10	47	2.2	27	2.1	20	2.3
20	140	6.5	92	7.2	48	5.6
30	257	12.0	172	13.5	85	9.9
40	423	19.8	268	21.0	155	18.1
50	596	27.9	347	27.2	247	28.9
60	492	23.0	265	20.8	226	26.4
70	120	5.6	71	5.6	49	5.7
80	55	2.6	30	2.4	24	2.8
90	6	0.3	4	0.3	2	0.2
Total	2136	100	1276	100	856	100

*information concerning gender was not available for some test takers

Table 5.4.b: Reading Performance for Full Sample and by Gender 2003-2009

Reading Performance Band Score	Full Sample		Males		Females	
	Freq	%	Freq	%	Freq	%
10	139	1.8	100	2.2	38	1.2
20	741	9.5	489	10.9	249	7.6
30	1577	20.2	985	22.0	582	17.7
40	1870	23.9	1072	23.9	790	24.0
50	1938	24.8	1089	24.3	836	25.4
60	1184	15.1	577	12.9	597	18.2
70	268	3.4	130	2.9	137	4.2
80	90	1.2	37	0.8	51	1.6
90	10	0.1	4	0.1	5	0.2
Total	7817	100	4483	100	3285	100

*information concerning gender was not available for some test takers

Table 5.4.c: Reading Performance for Full Sample and by Gender 2010-2012

Reading Performance Band Score	Full Sample		Males		Females	
	Freq	%	Freq	%	Freq	%
10	123	2.6	119	4.0	4	0.2
20	355	7.3	306	10.3	49	2.7
30	688	13.8	469	15.8	199	10.5
40	894	18.4	565	19.1	329	17.4
50	1010	20.8	590	20.0	420	22.2
60	1123	23.1	587	19.8	536	28.4
70	509	10.5	257	8.7	251	13.3
80	145	3.0	57	1.9	88	4.7
90	23	0.5	11	0.4	12	0.6
Total	4850	100	2961	100	1888	100

*information concerning gender was not available for some test takers

Distribution of Speaking Performance Scores

Table 5.5.a, 5.5.b and 5.5.c present the distribution of Speaking Performance results for this sample. It should be noted that the number of test takers reported for the 1997-1999 Speaking Performance sub-test (n=1779) is smaller than that reported for the Overall Result (n=2123) and the other sub-tests because the current version of the Oral Language Test was brought into use in the spring of 1997. By contrast, the distributions reported for the Overall Result and the Writing, Listening and Reading Performances of this sample were collected from January of 1997.

Table 5.5.a: Speaking Performance Scores for Full Sample and by Gender 1997-1999

Speaking Performance Band Score	Full Sample		Males		Females	
	Freq	%	Freq	%	Freq	%
10	4	0.2	2	0.2	2	0.3
20	25	1.4	17	1.6	8	1.1
30	275	15.5	177	16.5	98	14.0
40	398	22.4	244	22.7	154	21.9
50	500	28.1	310	28.9	189	26.9
60	241	13.5	133	12.4	108	15.4
70	123	6.9	71	6.6	50	7.1
80	182	10.2	101	9.4	80	11.4
90	31	1.7	18	1.7	13	1.9
Total	1779	100	1073	100	702	100

*information concerning gender was not available for some test takers

Table 5.5.b: Speaking Performance Scores for Full Sample and by Gender 2003-2009

Speaking Performance Band Score	Full Sample		Males		Females	
	Freq	%	Freq	%	Freq	%
10	47	.6	35	.8	12	.4
20	361	4.6	238	5.3	120	3.6
30	1426	18.2	901	20.1	516	15.7
40	1729	22.1	1079	24.1	639	19.4
50	2129	27.2	1143	25.5	977	29.7
60	1261	16.1	674	15.0	579	17.6
70	350	4.5	164	3.7	184	5.6
80	460	5.9	226	5.0	231	7.0
90	54	0.7	20	0.4	30	0.9
Total	7817	100	4480	100	3288	100

*information concerning gender was not available for some test takers

Table 5.5.c: Speaking Performance Scores for Full Sample and by Gender 2010-2012

Speaking Performance Band Score	Full Sample		Males		Females	
	Freq	%	Freq	%	Freq	%
10	57	1.2	51	1.7	6	0.3
20	138	2.8	121	4.1	17	1.0
30	328	6.8	238	8.1	90	4.8
40	663	13.7	462	15.6	201	10.8
50	1077	22.2	716	24.3	361	19.3
60	980	20.2	581	19.7	399	21.3
70	1055	21.7	533	18.1	522	27.9
80	484	10.6	226	7.7	258	13.8
90	38	0.8	23	0.7	15	0.8
Total	4820	100	2951	100	1869	100

*information concerning gender was not available for some test takers

Central Tendency and Variance of Score Distributions

Table 5.6 presents the means and standard deviations for the Overall Results and each of the sub-tests for both samples of test takers –the cohorts 2002-2007 and 2010-2012 respectively. This information is provided for the full sample and then separately for males and females. The highest overall mean in both samples is achieved for the Speaking Performance sub-test. While Writing Performance has the lowest overall mean in the 2002-2007 cohort, it remained somewhat paired with listening and reading in the 2010-2012 cohort. The sub-test with the greatest amount of variance in the 2002-2007 sample is the Speaking subtest as indicated by the standard deviation of 15.35 for the full sample, while in the 2010 – 2012 sample the greatest variance is in the Reading subtest. The least amount of variance for both samples is reflected in the Overall Results with a standard deviation of 12.89 and 14.90, respectively.

Table 5.6: Means & Standard Deviations for Sub-tests and Overall Results

Statistic	Overall Result		Writing		Listening		Reading		Speaking	
	02-07	10-12	02-07	10-12	02-07	10-12	02-07	10-12	02-07	10-12
Mean (Full Sample)	34.9 N=7817	48.75 N=4845	32.77 N=7817	47.99 N=4845	37.54 N=7817	47.22 N=4850	42.61 N=7817	47.69 N=4850	47.06 N=7817	55.67 N=4820
Std Deviation (Full Sample)	12.89	14.90	15.05	15.54	15.04	15.78	14.32	16.57	15.35	16.26
Mean (Females Only)	36.6 n=3285	52.90 n=1884	34.28 n=3285	51.78 n=1884	39.0 n=3285	51.47 n=1888	44.49 n=3285	52.44 n=1888	49.03 n=3228	59.74 n=1869
Std Deviation (Females Only)	13.04	13.40	15.43	14.21	15.13	14.90	14.24	14.60	15.42	14.80
Mean (Males Only)	33.63 n=4483	46.09 n=2960	31.61 n=4483	45.56 n=2960	36.47 n=4483	44.51 n=2961	41.20 n=4483	44.66 n=2961	45.59 n=4480	53.09 n=2951
Std Deviation (Males Only)	12.61	15.20	14.63	15.87	14.89	15.72	14.2	17.00	15.09	16.62

A comparison of male and female scores indicates that female test takers score marginally higher on all of the sub-tests as well as on the Overall Result. The sub-test with the greatest difference between males and females in the 2003 – 2009 sample was the speaking performance, while in the 2010-2012 sample, the reading sub-test showed the greatest difference. Continued tracking of the performance of test takers by gender may reveal the extent to which these observations are related to difference in sample size, characteristics of the test takers themselves, or a result of the test items. Further tracking of CAEL test scores is necessary before a conclusion can be drawn. In addition, a comparison of the performance of the CAEL Assessment across gender with other tests of English for academic purposes would also be insightful.

Chapter 6: Validity

Introduction

CAEL test developers have adopted an approach to validity that is consistent with the AERA/APA/NCME Standards (1985) and also reflects the writings of Samuel Messick (1989). Thus, CAEL test developers support the notion that "validity is an evolving property and validation is a continuing process" (Messick, 1989, p13). Through the ongoing process of test development and in-house research studies, evidence of the validity of the CAEL Assessment is continuously gathered. From time to time additional validity evidence is also provided in research studies, masters' theses and doctoral dissertations conducted at academic institutions outside the Language Assessment and Testing Research Unit (LATRU) at Carleton University. Readers interested in obtaining these documents may do so by contacting the Testing Coordinator (for full details see page one of this manual).

CAEL test developers embrace the definition of validity set out in the AERA/APA/NCME Standards document:

Validity is the most important consideration in test evaluation. The concept refers to the appropriateness, meaningfulness, and usefulness of the specific inferences made from test scores. Test validation is the process of accumulating evidence to support such inferences. Validity, however, is a unitary concept. Although evidence may be accumulated in many ways, validity always refers to the degree to which that evidence supports the inferences that are made from the scores. The inferences regarding specific uses of a test are validated, not the test itself. (1985, p. 9)

Traditionally, validity evidence has been gathered in three distinct categories: content-related, criterion-related, and construct-related evidence of validity. More recent writings on validity theory stress the importance of viewing validity as a 'unitary concept' (see Messick, 1989; AERA/APA/NCME Standards, 1985). CAEL test developers support the notion that validity is a unitary concept. Thus, while the validity evidence is presented in separate categories, this categorization is principally an organizational technique for the purposes of the presentation of research in this manual. The reader is encouraged to recognize that there is considerable overlap from one category of evidence to another. For example, a study which principally investigates criterion-related sources of evidence may also have real merit as a discussion of the construct-related validity of the test inferences. Finally, in keeping with the suggestion of Messick (1985), effort has also been expended to investigate the social consequences of the use of CAEL Assessment scores.

The evidence which has been gathered to date is presented in four categories: Ensuring Construct Representation, Investigating Construct Irrelevant Variance, Gathering Criterion-Related Evidence of Validity, and Consequences of Test Use.

Ensuring Construct Representation

As identified by Messick, one of the principal threats to the validity of any test is that of construct under-representation. Construct under-representation occurs when "the test is too narrow and fails to include important dimensions or facets of the construct" (Messick, 1989, p.34). Great care was taken to address this issue when the test specifications were developed for the CAEL. The process began with an in-depth 'needs analysis' of the English language requirements of first-year students in a variety of faculties at Carleton University. The purpose of the needs analysis was to attempt to define the reading, writing, speaking and listening requirements of students in the academic setting. The needs analysis included surveys and interviews with professors and students as well as an examination of the assignment and exam methods used in courses throughout the university. Test developers attended lectures and reviewed the lists of articles and textbooks which students were assigned to read. The results of this analysis are described in greater detail in Fox, Pynchyl & Zumbo, 1993 and in Fox 1995). Among the most important observations revealed in the needs analysis was that the extensive use of multiple choice items was found to occur only in the departments of Psychology and Economics and then only in first-year courses. The remainder of departments placed more emphasis on the use of short-answer and essay responses. Even the department of Mathematics used some extensive writing as students were required to describe the process they used in solving problems.

This information has heavily influenced the construct of English for academic purposes which is measured in the CAEL Assessment. The results of the needs analysis were the principal reason for adopting a constructed-response question format in the CAEL Assessment. Item types including filling in the blanks, labeling of diagrams, completing tables, note-taking, short answer and longer essays are used in the CAEL Assessment because these tasks are common within the Canadian university context.

More detailed test and item specifications were developed when the results of the needs analysis were complete. Each detailed CAEL Assessment specification includes the domain, title and general description of the criterion. A prompt or task definition is included to explain exactly what the student will be required to do for each test item. Item specifications also provide a response definition explaining what the student will do in responding to the prompt. This includes a description of successful and unsuccessful responses. A sample item or task is included with every item specification and additional information defining the characteristics of the context of use is provided. The use of these detailed test and item specifications in the development of new versions of the CAEL Assessment helps to ensure that the construct validity of the test remains true to the original concept of the test.

Follow-up studies are conducted at regular intervals to determine the extent to which faculty members at the university find the language requirements of the CAEL test to be consistent with the English language use in their courses.

Investigating Construct-Irrelevant Variance

The second principal threat to the validity of any test is the potential for construct-irrelevant variance. Construct-irrelevant variance exists when the "test contains excess reliable variance that is irrelevant to the interpreted construct" (Messick, 1989, p.34). This construct-irrelevant variance is viewed as a contaminant with respect to score interpretation. For this reason, CAEL test-developers attempt to investigate and minimize all sources of construct-irrelevant variance. Two kinds of construct-irrelevant variance can be identified. Construct-irrelevant difficulty occurs when "aspects of the task that are extraneous to the focal construct make the test irrelevantly more difficult for some individuals or groups" (Messick, 1989, p. 34). Conversely, construct-irrelevant easiness occurs when "extraneous clues in item or test formats permit some individuals to respond correctly in ways irrelevant to the construct being assessed" (Messick, 1989, p. 34). In general then, construct-irrelevant difficulty leads to lower scores for some test takers while construct-irrelevant easiness leads to higher scores for some test takers. CAEL test developers are constantly examining the test specifications, test items, administration and scoring techniques in an effort to reduce the impact of these threats to validity.

Efforts to investigate sources of construct-irrelevant variance generally take the form of research studies which are presented at academic conferences, peer reviewed and published. This process enables test developers to scrutinize various aspects of the test and to receive critical reviews from professional language testers throughout the world.

Probably the single greatest potential source of construct-irrelevant variance for the CAEL Assessment results from the integrated topic-based nature of the written test (the Oral Language Test is task-based). Because each version of the written test discusses one specific topic, it is possible that some test takers may be advantaged or disadvantaged in terms of their performance because of the topic of the test they wrote. That is, individual test taker's background knowledge, interest and opinions concerning a topic may impact their performance. The test specifications for the CAEL Assessment recognize this potential source of construct-irrelevant variance. For this reason, care is taken to ensure that all the information the test taker requires to formulate responses is contained within the testing materials provided. The content of the readings and the lecture provide a rich context, which test takers can draw upon in making their responses. Attempts to build the type of *context-free* items which are often claimed by other English language proficiency test developers are not seen as appropriate for the construct of English for academic purposes which the CAEL Assessment endeavours to measure.

Nonetheless, CAEL test developers take seriously the responsibility to demonstrate that the topic of the test is not a source of construct-irrelevant variance. For this reason, studies have been conducted and published investigating this potential validity threat. The most extensive study of this issue is described in Jennings, Fox, Graves & Shohamy (1999) and has already been described in some detail in the '*Comparability of Test Versions*' section of Chapter Four, page 55. The results of this study which involved a comparison of test taker performance across six different versions of the test found no impact of topic on the scores of test takers (n=254). CAEL test developers are confident that the topic of the test is not a source of construct-irrelevant variance.

Gathering Criterion-Related Evidence of Validity

Gathering criterion-related evidence of validity is an important task for all language testers. This task is particularly difficult for the CAEL test given some of the unique features of the CAEL Assessment. That is, the use of constructed-response test items in a topic-based fully integrated language test is essentially a unique approach to language testing at the present time. These aspects of the CAEL Assessment strengthen the claim made by test developers that the CAEL Assessment is a reasonable approximation of the language demands of English for academic purposes, particularly in Canadian university contexts. However, the essentially unique nature of the test means that gathering criterion-related evidence of validity is problematic.

CAEL test scores have been compared with the performance of test takers on the Test of English as a Foreign Language (TOEFL). However, the TOEFL is clearly measuring English language proficiency in a very different manner. Does this mean, then, that a correlation of the two test scores provides criterion-related evidence in support of the CAEL test? Clearly, the establishment of an appropriate criterion is always a challenge when gathering this type of validity evidence.

One procedure that has been adopted in an effort to gather more meaningful criterion-related evidence of validity is to conduct follow-up studies of CAEL test takers who score at various proficiency levels. One such follow-up study was conducted for this manual. In this study the university course performance of 79 test takers who achieved an Overall Result at a band score of 70 or greater was collected. The basic design was to determine the grade point averages (GPA) of these students in their first full term of study after achieving an Overall Result of 70 or greater on the CAEL Assessment. The score of 70 was selected for this study because test takers who achieve this score are permitted to register for regular courses at the university without any further ESL/EAP training. Data was collected for each test taker for the term immediately following their CAEL Assessment in an effort to avoid measuring the impact of language learning which occurred after the test was completed.

A six-point scale was used for the GPA as shown in Table 6.1.

Table 6.1: Grade Point Average and Letter Grade Equivalents

Letter Grade	Grade Point Value	Letter Grade	Grade Point Value
A+	6.0	C+	3.0
A	5.5	C	2.5
A-	5.0	C-	2.0
B+	4.5	D+	1.5
B	4.0	D	1.0
B-	3.5	D-	.5
		F	0

In the sample of 79 test takers the range of GPA's was from 0 through to 6.0. The frequency distribution of GPA's for this sample is shown in Table 6.2.

Table 6.2: Frequency Distribution of Grade Point Averages

Grade Point Average Range	Freq	%
0 to 0.49 (F)	2	2.6
0.5 to 1.9 (D +/-)	6	7.7
2.0 to 3.4 (C +/-)	15	19.1
3.5 to 4.9 (B +/-)	28	35.4
5.0 to 6.0 (A +/-)	28	35.4
Total	79	100

We can begin to interpret this information in absolute terms. It would appear that the vast majority of test takers who took the CAEL Assessment and were permitted to register in university courses as a result of their strong performance have been able to successfully complete their courses. The mean GPA for the entire sample was 4.11. This indicates a GPA at the letter grade level of a B. While table 6.2 indicates that 2 of the 79 students achieved a GPA equivalent to an F, it is very difficult to determine the extent to which their difficulties were related to English language proficiency. However, it is clear that the vast majority of the students are performing in an acceptable manner after completing the CAEL test.

Consequences of Test Use

Messick (1989), among others, highlights the importance of considering both the intended and unintended consequences of test use when accumulating evidence of the validity of inferences for a given test. This emphasis on the consequences of test use is also included in the principles of fair and ethical testing described in the Standards document referred to earlier in this manual (AERA, APA, & NCME, 1985). CAEL test developers are committed to ensuring that the information provided to test takers and the institutions which use CAEL Assessment Score Reports is a consistent and meaningful measure of the candidates' proficiency in the use of English for academic purposes. The use of the CAEL Assessment in lieu of other currently available measures of English language proficiency arguably provides the test taker with a more accurate measure of the construct of English for academic purposes. However, it is important that the test takers and test score users have the same confidence in the validity of the inferences made from CAEL Assessment results as is held by the test development team. In an effort to determine the extent to which test takers and test score users share this impression of the CAEL Assessment, feedback is solicited through a number of means.

CAEL test developers gather on an on-going basis the feedback of test takers, score users and members of the university community who are impacted by the CAEL Assessment. Efforts to gather and review this feedback help to minimize any potentially negative consequences of the use of CAEL Assessment scores.

Test Taker Feedback: Post-Test Questionnaire

At the end of every CAEL Assessment, test takers are asked to complete a one-page questionnaire. The questionnaire is designed to allow test developers and administrators to gather the comments and opinions of test takers concerning the CAEL Assessment. At present, there are three types of questions on the questionnaire. In the first type of question, test takers are given statements concerning the CAEL Assessment and asked to respond by circling agree, disagree or no opinion. The second question asks test takers to rank five factors in terms of which factor they find most important in their test experience. Finally, the questionnaire includes three open-ended items in which test takers are given the opportunity to indicate what they like about the test, what they dislike about the test, and what aspects of the test might be changed in order to improve their performance. The test takers' responses to the questionnaire are summarized at regular intervals and are considered carefully when tests are being developed and when changes are being made to the test administration process.

In response to the statement 'This test was a good experience', 83.2% of test takers agree, 5.6% disagree, 8.4% indicate that they have no opinion and 2.9% did not respond to the item.

The responses from a recent summary of 790 test taker questionnaires collected from the period from January 1997 through September 1999 is presented in this chapter to illustrate the type of information which has been collected. In response to the first 'agree/disagree' statement 'This test reflects my true knowledge of English.' 39.1% of test takers agree, 32.0% disagree, 23.8% indicate that they have no opinion and 5.1% did not respond to the item. While the largest group of test takers indicates that they feel the test does reflect their knowledge of English there is still a relatively large proportion of test takers who disagree with the statement. Some further insight may be gained from the open-ended items on the questionnaire. Some test takers felt that they lacked sufficient information to assess their own knowledge of English. These test takers tended to respond in the 'no opinion' category. Other test takers clearly felt that their performance on the test was not as strong as it could be and that their knowledge of English was greater than that revealed by their test performance. These test takers tended to disagree with the statement. It is hoped that continued research will provide evidence as to the extent to which this tendency to feel the test does not reflect 'true knowledge' is an artifact of all testing settings and the extent to which it is specific to the CAEL Assessment experience.

In response to the statement, "This test was fair," 56.1% agree, 13.7% disagree, 24.7% indicate that they have no opinion and 5.6% did not respond to the item. Clearly, while some test takers feel that the test did not reflect their true knowledge of English, most test takers feel that the CAEL Assessment was fair.

In the second question, test takers are asked to rank the importance of five factors in their test performance. The five factors are the amount of time allowed to complete test items, the physical comfort of the testing environment, the availability of sample tests, the sound quality of the taped-lecture, and the topic of the test which the test taker completed. These factors have been identified by test takers in response to the open-ended items of the questionnaire and have also been reported on in a number of research studies.

A recent summary of 790 test taker questionnaires indicated that 29.5% of the test takers did not answer the ranking question. Further analysis of these results indicated that over 50% of those test takers who did not answer the item had achieved Overall Results with band scores at 40 or lower. This may indicate that the format of the item is too difficult for test takers with lower levels of English language proficiency. For this reason, the pattern of responses with respect to this item more closely reflect the opinions of test takers with intermediate or advanced levels of English language proficiency as assessed by their CAEL performance. The factor that was most frequently identified as the most important by these test takers was the amount of time allowed to complete the test items. In fact 25.9% of test takers indicated that this was the most important factor in their test score performance. This finding is consistent with the results of other research studies which examined the test takers' responses to testing conditions (Norton & Starfield, 1997). The factor which was identified the second most frequently as the most important factor in the test takers testing experience was the topic of the test. From this sample 22.7% of the test takers felt that the topic of the test was the most important factor in their performance.

The responses to the open-ended sections of the post-test questionnaire seem to support the observations from the ranking item. In the three open-ended items on the questionnaire test takers are asked to indicate what they liked about the test, what they disliked about the test and what changes to the test would help them most. It is difficult to summarize the results from these open-ended questions in a manner which is brief enough for inclusion in the test manual. However, the essential findings were that the majority of open-ended comments re-iterated a need for more time to answer the items. The second most frequently appearing comment was that the topic of the test was problematic.

Test Impact over Time

CAEL test scores have important consequences for test takers, as is the case with all high-stakes language tests. Achieving a certain score on a CAEL Assessment may mean the difference between being accepted to study in a university program or being denied admission. Therefore, it is important to examine the adequacy of CAEL test decisions through the on-going collection of evidence regarding test takers whose lives are affected by their performance on the CAEL Assessment.

In order to examine the relationship between CAEL test performance, which occurs within the highly constraining context of a testing setting, and actual language ability demonstrated in a classroom setting over time, data are collected on an on-going basis within the English for Academic Purposes program at Carleton University. Within the Carleton context, a CAEL test taker's overall band score is related to a level of English language ability that is linked to and reflective of one of a range of EAP courses offered by the School of Linguistics and Language Studies⁵. Thus a band score of 40 allows a test taker to register for a specific,

⁵ As is the case in a number of other universities, Carleton has always been described as having a "gradual admission" policy. Students are allowed to begin their university program provided they include a required EAP course during the first term/s of study and a limited number of courses in their chosen discipline, according to a formula defined by the School of Linguistics and Language Studies. Thus, students testing at band score 60 on the CAEL Assessment would be required during their first term of study to register in ESLA 1900, an advanced EAP course, and allowed to take three other courses specific to their programs of study. The students would earn credits toward their degree from the EAP course as well as the other academic courses.

introductory course in English for English for Academic Purposes and take one additional course in their discipline. A band score of 50 allows a test taker for a specific, intermediate course in English for Academic Purposes and to take two additional courses in their academic discipline. The organization of the EAP program and its relationship to performance on the CAEL Assessment creates a context wherein the adequacy and usefulness of the CAEL scores may be examined over time.

For example, a 1999 study of test impact utilizes the overall CAEL Assessment scores of 120 test takers who wrote the CAEL Assessment during July and August, 1999 at Carleton University and were later registered in one of the 12-week EAP courses for the Fall term (September-December, 1999). At the end of ten weeks of instruction, a questionnaire was circulated to the 12 teachers who received these students. The questionnaires identified the test takers in their classes who had been placed there on the basis of the test⁶, and asked the teacher's evaluation over time. Nine (9) of the teachers responded regarding the performance of ninety five (95) students. The correlation between the level of language ability demonstrated by the test taker performance on the CAEL Assessment and the language ability demonstrated by the student in class is significant ($<.01$), with a correlation co-efficient of .863 based on a Spearman rank-order correlation.

In addition to Spearman's rho, the distribution of differences between CAEL test performances and the teachers' evaluations for these 95 test takers is reported below in Table 6.3. No change indicates that the band score on the CAEL Assessment was, in the teachers' opinions, an accurate indication of the students' actual English language ability as demonstrated by their in-class performance during the 10 week period. A change of one either indicates the CAEL test performance under- or over-estimated the students' English language ability.

Table 6.3: Agreement between CAEL Assessment Placement and Teacher Evaluations (n=94)

Grade Point Average Range	Freq	%
No change	76	81%
Within One/half to One Band	15	16%
Within Two Bands	3	3%
Total	94*	100

*One test taker withdrew during the 10-week period.

As table 6.3 indicates, there is considerable agreement between the test taker's performance on the CAEL Assessment and the EAP teachers' evaluation of ability to use English based on 10 weeks of evaluation in the context of the classroom.

⁶ Although all students test into an EAP course on the basis of a CAEL test score, they may satisfy the language requirement by successfully completing the highest EAP course, ESLA 1900.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1985). Standards for Educational and Psychological Testing. Washington, DC: American Psychological Association.
- Diederich, P. (1967). Cooperative preparation and rating of essay tests. English Journal, 55, 435-449.
- Fox, J. & Soucy, A. (1991). Collaborative assessment of writing. National Testing Network on Writing, 4, 10-11.
- Fox, J., Pychyl, T.A., & Zumbo, B.D. (1993). Psychometric properties of the CAEL Assessment, I: An overview of development, format, and scoring procedures. Carleton Papers in Applied Language Studies, Vol. 10, Carleton University, Ottawa, Canada, 1-12.
- Fox, J., Pychyl, T.A., & Zumbo, B.D. (1995). An Investigation of Background Knowledge in the Assessment of Language Proficiency. In A. Huta, V. Kohonen, L. Kurki-Suonio & S. Luoma (Eds.), Current Developments and Alternatives in Language Assessment. Jyväskylä: University of Jyväskylä and University of Tampere, 367-383.
- Freedman, A. (1987). Learning to Write Again: Discipline-Specific Writing at University. Carleton Papers in Applied Language Studies, Vol. 4, Carleton University, Ottawa, Canada, 95-115.
- Jackson, S. (1998). Unpublished report on CAEL Assessment comparability with TOEFL and Oxford Proficiency Tests. Malmö, Sweden: World Maritime University.
- Jennings, M., Fox, J., Graves, B. & Shohamy, E. (1999). The test takers' choice: an investigation of the effect of topic on language test performance. Language Testing, 16(4), 426-456.
- Messick, Samuel. (1989). Validity. In R.L. Linn (Ed.), Educational Measurement, Third Edition. Macmillan, New York, 13-103.
- Young, L. (1984). Needs Assessment Surveys. Carleton Papers in Applied Language Studies, Vol. 1, Carleton University, Ottawa, Canada, 51-60.
- Zaluska, D.T. Editor, (1997) Carleton University Data Book (1996-97). Office of Budget Planning, Carleton University Graphic Services, Ottawa, Canada.
- Zumbo, B.D., Fox, J., & Pychyl, T.A. (1993). Psychometric properties of the CAEL Assessment, II: An examination of the dependability/reliability of placement decisions. In Carleton Papers in Applied Language Studies, Vol. 10, Carleton University, Ottawa, Canada, 13-27.

Related Articles:

- Artemeva, N. & Fox J. (2010) Awareness vs. Production: Probing students' antecedent genre knowledge. *Journal of Written and Business Communication*.
- Berman, R. and Cheng, L. (2001). English academic language skills: Perceived difficulties by undergraduate and graduate students and their academic achievement. *Canadian Journal of Applied Linguistics*, 4, 1, 25-40.
- Fox, J. Taking a diagnostic turn: Reinventing the portfolio in EAP classrooms (with P. Hartwick), In D. Tsagari & I. Csepes (Eds.), *Classroom-based language assessment*. Frankfurt: Peter Lang, in press.
- Fox, J. Exploring the testing process: Three test takers' observed and reported strategy use over time and testing contexts (with C. Doe). *Canadian Modern Language Review*, in press.
- (2008). Moderating top-down policy impact and supporting EAP curricular renewal: Exploring the potential of diagnostic assessment. *Journal of English for Academic Purposes*, 8, 26-42
- (2005). Rethinking second language admission requirements: Problems with language residency criteria and the need for language assessment and support. *Language Assessment Quarterly: an International Journal*, 2, 2, 85-115
- (2004). Test decisions over time: Tracking validity. *Language Testing*, 21, 4, 437-468.
- (2003). From product to process: An ecological approach to bias detection. *International Journal of Testing*, 3, 1, 21-47.
- (2002). It's all about meaning: L2 test validation in and through the landscape of an evolving construct. PhD Dissertation. Ottawa: Carleton University.
- (2001). The role of reflection in the Canadian Academic English Language (CAEL) Assessment. CAEL Assessment Testing Office, Carleton University, Ottawa, Canada.
- Fox, J. & Fraser, W. (2008). Comparative analysis Qatar Nursing Applicant Test Scores: CAEL Assessment and University of Calgary Admission Test. College of the North Atlantic - Qatar.
- Fox, J., Pychyl, T. & Zumbo, B. (1997). An investigation of background knowledge in the assessment of language proficiency. *Current Developments and Alternatives in Language Assessment*. A. Huhta, V. Kohonen, L. Kurki-Suonio and S. Luoma (Eds.). Jyväskylä: Finland. University of Jyväskylä Press, 367-383.
- Fox, Zumbo & Pychyl (1993). Psychometric properties of the CAEL Assessment II: An examination of the dependability/ reliability of placement decisions.. *Carleton Papers in Applied Language Studies*, X:13-26.

Fox, Zumbo and Pychyl (1993). Psychometric properties of the CAEL Assessment I: Overview of development, format, and scoring procedures (with T. Pychyl and B. Zumbo). Carleton Papers in Applied Language Studies, X:1-12.

Fraser, W and Fox, J. (2007) Working for bi-directional washback: Tapping the potential of learner perception and classroom practice to improve test quality (with J. Fox), in T.S.C. Farrell (Ed.), *Teacher Research Series, Vol. 3, Language Teacher Research in the Americas*, (H. McGarrell, Ed.). Washington, D.C.: TESOL Publications.

(2005). Reflective item types in English for academic purposes testing: and examination of method effect. MA Thesis, School of Linguistics and Applied Language Studies, Carleton University, Ottawa, Canada.

(2001). The role of reflection in the Canadian Academic English Language (CAEL) Assessment. CAEL Assessment Testing Office, Carleton University, Ottawa, Canada.

Fraser W. and Fox, J. (2007). Studying classroom practices and learner perceptions to improve test quality. In H. McGarrell (Ed), *Language Teacher Research in the Americas*. TESOL, 69-82.

Luomo, S. (2004). *Assessing speaking*. Oxford: Cambridge University Press.

Theses and Dissertations:

Aminkova, Elena (MA) Test-Takers' Perceptions: Behind the Construct of a Test. 2004, Carleton University, Ottawa, Canada.

Doe, Christine (MA) The role of preparation in test takers' strategy use and test performance. 2007. Carleton University, Ottawa, Canada.

Fox, J. (PhD) It's all about meaning: L2 test validation in and through the landscape of an evolving construct. 2002, McGill University, Montreal, Canada.

Fraser, Wendy (MA) Reflective Item Types in a High-Stakes Assessment of English for Academic Purposes: An Examination of the Method Effect. 2005, Carleton University, Ottawa, Canada.

Hungerland, Ruth (MA) The Role of Contextual Factors in Mediating the Washback of High-Stakes Language Assessments. 2004, with distinction Carleton University, Ottawa, Canada.

Kennedy, Lauren (MA) Expanding Test Specifications with Rhetorical Genre Studies and Activity Theory Analyses. 2008. Carleton University, Ottawa, Canada.

Youyi, Sun (MA) The Influence of the Social Interactional Context on Test Performance: A Sociocultural View. 2005, with distinction Carleton University, Ottawa, Canada.