

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI[®]

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Needs Assessment and Formative Evaluation in the Development of a College
Level English Second Language Placement Test

Lydia Barsalou Froio

A Thesis

in

The Department

of

Educational Technology

Presented in Partial Fulfilment of the Requirements
for the Degree of Master of Arts at
Concordia University
Montreal, Quebec, Canada

September 1997

© Lydia Barsalou Froio, 1997



National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

Our file Notre référence

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

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

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

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

0-612-40162-6

A B S T R A C T

Needs Assessment and Formative Evaluation in the Development of a College Level English Second Language Placement Test

Lydia Barsalou Froio

Instructional design and test development models converged in the development of a valid placement test for college level students in English second language courses in Quebec. The realities of instructional design within the context of budget cuts and time constraints often result in the curtailment of the needs assessment and formative evaluation phases of design. The development of the placement test relied heavily on these two essential phases of instructional design despite constraints. The needs assessment and formative evaluation phases provided the necessary insights for the production of a test that corresponds to the learning objectives of the curriculum and the expectations of the test users. Adhering to these models provided the framework for the test design and ensured that the test items produced for the large scale field testing were functioning well in terms of their level of difficulty and their ability to discriminate. This thesis deals with the needs assessment that led to the development of items for a placement test and the formative evaluation of both the individual items and the test itself. $n= 299$ and $n= 118$, respectively

This thesis is dedicated to my husband

Paul

with gratitude for the past

and to our son

Anthony

with hope for the future

A C K N O W L E D G E M E N T S

The work that this thesis represents required the support and the collaboration of many people professionally, academically and personally. My heartfelt appreciation goes to my colleagues in the ESL community who supported with enthusiasm and involvement our work on the development of the placement test. To the many students, teachers, deans and support staff who participated in one or several of the development phases, I say thank you. It was a pleasure to work with you.

A special thank you to the two people with whom I worked closely for two years on the English second language placement test: Charles Pearo and Michel Laurier. I learned a great deal from both of them and would gladly repeat the experience. Thank you, Chuck, for convincing me to do this. Thank you, Michel, for your time and your diligence in what was, at times, a race against the clock.

I would also like to thank the *Programme d'aide à la recherche sur l'enseignement et l'apprentissage (PAREA)* and Collège André-Laurendeau and Collège de Maisonneuve for providing the necessary funding. A particular thank you to Hélène Lavoie at the Ministry of Education.

My appreciation goes to Dr. Richard Schmid, my thesis advisor, who provided valuable insights and feedback on my work. He made himself available even under difficult conditions. Thank you to Dr. Gary Boyd and Dr. Robert Bernard for accepting on such short notice to replace the original committee members. A special wink and a thank you to Anne Brown MacDougall, the administrative assistant, who always was there with the right information at the right time.

As always, a load of appreciation goes to Katharine O'Flynn, an ESL colleague, a wonderful proof-reader and consummate professional.

Last and by no means least, I would like to thank my husband and my son. There are no words to express my deep appreciation for their support of my work. Combining studies and a full time job is an extraordinary undertaking but their help and encouragement eased the way making the workload seem almost ordinary. Thank you, Mom and Dad.

TABLE OF CONTENTS

List of Tables.....	viii
List of Figures.....	ix
Introduction.....	1
Educational Context.....	1
Background on ESL in Quebec.....	1
The Reform of the College Curriculum.....	3
The Mandate from the Ministry of Education.....	3
Objective.....	4
Methodology.....	6
Needs Assessment Phase.....	7
Needs Assessment Literature Review.....	7
Design of the Survey Instruments.....	9
Formative Evaluation of the Needs Assessment Questionnaires.....	13
Results of the Needs Assessment.....	17
Testing Instruments and Cut-off Scores.....	18
Need for Placement and a New Test.....	22
Standardization of Placement Test and Cut-off Scores.....	23
Level of Satisfaction with Present Tests.....	24
Problems Related to Present Tests.....	25
Test Content.....	26
Administration of the Test.....	27
Outcome of the Needs Assessment Phase.....	29
Test Design Phase.....	30
Testing Literature Review.....	30
Purpose of Tests.....	30
Diagnostic.....	31
Achievement.....	31
Proficiency.....	31
Placement.....	32
Test Qualities.....	33
Reliability.....	33
Validity.....	33
Practicality.....	34
Methods and Forms of Testing.....	35
Criterion-referenced vs. Norm-referenced Tests.....	35
Direct vs. Indirect.....	36
Discrete vs. Integrative.....	37
Objective vs. Subjective.....	38
Pencil and paper vs. Computer adaptive.....	38
Test Design.....	39

Formative Evaluation Phase	45
Formative Evaluation Literature Review.....	45
Methodology.....	46
Sample	47
Results of the Pilot Testing.....	51
Conclusion	58
Concluding Remarks on the Importance of Needs Assessment.....	58
Concluding Remarks on the Importance of Formative Evaluation	59
Epilogue.....	60
References	62
Appendix A: Needs Assessment Questionnaires	66
Appendix B: Language Skill Matrices.....	83
Appendix C: Test Site Documents.....	88
Appendix D: Formative Evaluation Questionnaires and Quantitative and Qualitative Data.....	93

LIST OF TABLES

Table 1: Placement tests used in Quebec colleges.....	2
Table 2: Needs assessment questionnaire topic summary and question numbers.....	12
Table 3: Response rate.....	18
Table 4: Placement tests in use.....	19
Table 5: Cut-off scores for colleges using the Comprehensive English Language Test for Speakers of English as a Second Language.....	20
Table 6: Cut-off scores for colleges using the Michigan Test of English Language Proficiency.....	20
Table 7: Cut-off scores for colleges using TCALS 100 (Test de classement d'anglais langue seconde - 100 questions).....	21
Table 8: Cut-off scores for colleges using TCALS 50 (Test de classement d'anglais langue seconde - 50 questions).....	21
Table 9: Cut-off scores for colleges using TCALS 150 (Test de classement d'anglais langue seconde - 150 questions).....	21
Table 10: Question # 15a & 1: Do you want a new test?.....	22
Table 11: Question #15b & 3: Should all colleges use the same placement test?.....	23
Table 12: Question #15c & 4: Should cut-off scores be the same for all colleges?.....	24
Table 13: Question #20a & 5: Who should be responsible for administering a placement test?.....	27
Table 14: Question #20b & 6: Who should correct a placement test?.....	28
Table 15: Question #15d & 7: Would you be interested in using computers for placement test-taking?.....	28
Table 16: Question #16 & 8: Type of test preferred.....	29
Table 17: Design of the placement test.....	44
Table 18: Student population by region.....	48
Table 19: Sample by college and by region.....	49

Table 20: Descriptive statistics for the 9 sub tests in form 1	53
Table 21: Descriptive statistics for the 9 sub tests in form 2.....	53
Table 22: Descriptive statistics for the 9 sub tests in form 3.....	54
Table 23: Sample Item 1.....	55
Table 24: Sample Item 2.....	56
Table 25: Sample Item 3.....	56

LIST OF FIGURES

Figure 1. Test design.....	42
----------------------------	----

I N T R O D U C T I O N

The thesis describes the design and evaluation of items for an experimental version of a college level English second language (ESL) placement test for the Ministry of Education in Quebec. The thesis also includes a formative evaluation of the test conducted exclusively by the author. The results of the formative evaluation are discussed as well as the ways in which these results were used to prepare the items and the instructions for the eventual large scale field test of the experimental version of the placement test with two thousand students at fourteen test sites.

The thesis consists of:

- the development of a needs assessment questionnaire
- the analysis of the needs assessment questionnaire
- a preliminary test design
- pilot testing of preliminary items
- the analysis of data from the pilot testing
- formative evaluation of the test instrument
- analysis of the formative evaluation data.

The thesis does not involve content validation or reliability studies since these will be conducted after the large scale field testing for which the work described is the basis.

E D U C A T I O N A L C O N T E X T

Background on ESL in Quebec

Historically, second language study in Quebec has always formed an important part of the college curriculum, mostly as an elective course. In only a few programmes was it compulsory: Arts, Office Studies and Technology and more recently in the Computer Science programme. All students registered in English second language (ESL) courses were required to take a placement test in order to be put into an appropriate level. Students were then placed in a course at one of six levels of difficulty ranging from false beginner to very

advanced. In Quebec, between grade 4 and finishing high school in grade 11, students receive approximately 500 hours of formal English instruction.

In their survey, Pearo and Rock (1988) described the instruments used for placement at 53 of 60 public and private colleges in Quebec. At the time, 60% of respondents (29 colleges) used TCALS (*Test de classement d'anglais langue seconde*) developed by the Provincial Coordination Committee and implemented in 1976. Eight colleges used The Michigan Test of English Language Proficiency and three colleges used the CELT (Comprehensive English Language Test). The remaining colleges used various other methods for classifying students. In addition, the survey also reported on the cut-off scores institutions had established for each of the levels of English second language (ESL) courses. In many cases, institutions which used the same tool, often did not have the same cut-off scores for the same level.

Fournier's 1995 report on the new courses updated the Pearo and Rock survey on the state of affairs of placement tests in the colleges. Table 1 reports on the forty-two colleges which were included in the report.

Table 1
Placement tests used in Quebec colleges

Placement test	Colleges
TCALS	21
Michigan	6
CELT	3
Other instruments and methods	12

Other problems in placement testing were cited in the report, most important of which was the tendency for students at the advanced and the high intermediate levels to back place themselves in order to be assigned a lower level of English and, therefore, an easier course. Students reason that taking an easier course ensures a higher z-score and a better profile for admission to particular universities and particular programmes.

The Reform of the College Curriculum

In 1993, the Minister of Education, Lucienne Robillard of the Liberal party, passed a law reforming the college curriculum. Among other major changes, the law made second language courses compulsory to obtain a *Diplôme des études collégial*. This means that students attending francophone colleges must now complete two courses in second language, one a general English course, the other, an ESL course adapted to each programme of studies. The Reforms also reduced the number of levels offered from six to five. In fact, the reduction is more substantial since there are really only four credited English courses numbered 104¹, 101, 102 and 103 at the elementary, intermediate, advanced and very advanced levels respectively. The fifth course is a remedial non-credit course. Understandably, this has led to wider spread of abilities among the students in the groups.

The Mandate from the Ministry of Education

Although the Ministry of Education had long hoped to use high school leaving scores in English for placement purposes, several attempts to use these scores for placement at various colleges failed disastrously. The final high school grade is made up of the results of a provincial exam and test scores obtained throughout the year. Both are graded by the high schools. Despite a common curriculum, there is no evidence that the teachers are focusing on the same abilities and share the same requirements. This raises the question of validity. Also, since 50% of the mark is based on local attribution and there are no procedures or guidelines to ensure inter-rater reliability, it is not surprising that Secondary V scores in ESL are not good indicators of college level placement. In addition, the high school curriculum offers only two levels of English, which means a large variance in the language abilities of the students within the groups.

¹At the time of publication of this thesis, the Ministry of Education changed the number of the elementary English course to English 100.

In early 1995, the Ministry decided that a new placement test was needed to ensure correct placement of college students in the compulsory ESL courses. Since over half of the colleges were using a test that was almost twenty years old, this was a timely decision since "norms change with time as the characteristics of the population change, and therefore such tests must be renormed" (Henning 1987, p. 8).

Hughes (1989) does not recommend purchasing a test from a commercial test developer since no one placement test will work for every institution. The most successful tests are those constructed for particular situations since effective placement depends on the identification of the key features of the different levels of teaching in the institution and requires that the test be tailor-made. Harrison (1983, p.2) concurs and states that "a good test, like a good suit, should be made for the individual customer."

The Ministry of Education mandated a three-member research and development team to produce within a two-year period a placement test for the francophone colleges. The team consisted of two ESL teachers from two Montreal colleges and an expert in testing and evaluation from the Université de Montréal. The author of this thesis was a member of the research and development team - working primarily as an instructional designer. The Ministry did not establish guidelines or a specific model for the test development. This was to be determined by the development team under the guidance of the expert in testing and evaluation.

The first year of development included needs assessment, curriculum analysis, test design, item writing and pilot testing. The test items had to be pre-administered before the large scale trials could be undertaken in the second year of development.

O B J E C T I V E

This thesis describes the preliminary phases of development of an ESL placement test. The realities of instructional design within the context of budget cuts and time constraints often result in the curtailment of the needs assessment and formative evaluation phases of design. This development project relied

heavily on these two essential phases of instructional design despite constraints.

The purpose of this thesis was to design, develop and pilot test a preliminary version of a college level placement test based on information from a needs assessment obtained from teachers of ESL and college deans and an analysis of the curriculum. The preliminary test was also formatively evaluated by the test takers. The test must assess the proficiency of the students in communicating in English and correspond to the entry skills for each of the four credited levels of English and the one remedial non-credit course. The new placement test could be used by approximately 150,000 students in over 56 colleges across Quebec.

The design and development phases of the test was based on the Carroll, & Hall (1985) development model detailed below.

Design

- definition and description of the test takers
- specifications of the setting and needs
- statement of tasks and topics

Development

- construction of the drafts of the actual test
- trials of draft test on a suitable sample to get indications of its validity and stability and of any practical problems that may not have been foreseen
- analysis of trials
- test revision

Operation

- introduction of the test for practical use
- development of the final version of the test
- use of the test on a sizeable population

Monitoring

- follow up on test use
- establishment of test measurement characteristics
- preparation of an on-going revision schedule.

M E T H O D O L O G Y

The design and development phases of constructing this placement test are founded on the iterative use of instructional design models that consist of: needs assessment, test design, item writing, field tests and formative evaluations. Needs assessment, an essential phase in the instructional design process, is absent from the above test development model. Although time-consuming, needs assessment allowed the researchers to ask those involved in placement testing in the colleges for their expert judgement. The mandate did not require this phase and the test could have been designed solely by the language specialists on the team who had over 40 years of experience between them. However, the author's academic background in instructional design led to the development of the placement test using a consultative approach. In the development of the test, various questionnaires were designed and used in each phase. During the two year project, specific questionnaires were addressed to students, teachers, language department coordinators or college deans at different times and for different purposes.

In the autumn of 1995, the first phase of development of the new ESL placement test began. The needs assessment included three survey questionnaires addressed to teachers, department coordinators and academic deans in all the public and private francophone colleges in Quebec. This was followed by the test design phase which was based on an analysis of the Ministry standards and objectives for the five levels as well as the quantitative and the qualitative data from the needs assessment. The third phase, formative evaluation, involved field testing the placement test items and conducting a formative evaluation of the test.

These three phases were the essential steps in the preparation of the placement test items for large scale field testing that would take place in subsequent phases. Following the data analysis, decisions about the final design of the test would be taken and actual production of the placement test begun. These later stages do not form part of this thesis.

N E E D S A S S E S S M E N T P H A S E

This chapter deals with the literature on needs assessment, the design and formative evaluation of the needs assessment questionnaires and the results obtained.

The needs assessment attempted to clarify the actual test setting in the colleges and any problems regarding the placement of students in the ESL courses. For instance, Fournier (1992) reported that correction of the present placement test is done by staff other than language faculty in all but eight colleges. It would seem unlikely that the colleges would accept a direct test using subjective scoring methods. Given the number of students that need to be tested, this would put an enormous strain on language departments since the subject matter experts would necessarily have to handle the correction of the test. The needs assessment phase was expected to clarify issues such as these.

In order to identify the optimal and actual (Rossett) of the present placement testing situation in Quebec, three questionnaires were developed, piloted and mailed to ESL teachers, department coordinators and the Academic Deans of the francophone public and private colleges in Quebec. These needs assessment questionnaires were designed to elicit the views of the ESL community regarding the design and definition of the test and were pilot tested with a group of volunteers from each of the groups mentioned.

Needs Assessment Literature Review

Usually a needs assessment is conducted prior to the development of tests or instructional materials to obtain information on

- optimal - what respondents think ought to be going on; how the system should work; what they know about it;
- actual - the details of performance; the way the system is operating; whether respondents perceive a problem;
- feeling - how people feel about the topic; how they think others feel; confidence regarding the topic; whether they value/like the topic;
- cause - what is causing the problem, if any;

- solutions - options for solving the problem or implementing innovations (Rossett, 1987).

According to Rossett (1987) and Berdie, Anderson & Niebuhr (1986), questionnaires are a cost-effective method of getting the necessary information from many people at the same time. As well, questionnaires are usually easy to respond to and give respondents time to reflect on their answers.

The major steps involved in developing and designing effective surveys are:

1. deciding what information is needed and from whom
2. writing effective items
3. writing good directions
4. writing good cover letters
5. piloting the instruments.

Well-designed questionnaires yield a higher response rate and are easy to score and analyze. Questions need to be worded simply and chosen carefully. Rossett, commenting on wording and content of the questions, states that the "foremost challenge in producing a questionnaire is getting it right for public consumption." In addition, a great deal of attention should be paid to the wording and the clarity of the instructions (Carroll & Hall, 1985; Dick & Carey, 1990; Rossett, 1987; Smith & Ragan, 1993).

Pre-testing the questionnaire is an essential step to determine whether each item is clear to respondents. Questionnaires can be piloted in two ways: intimate or expanded. The intimate approach involves asking a few potential respondents (representative of the population) to respond and react to a draft of the questionnaire, encouraging them to make suggestions and comments. If a questionnaire is piloted using the expanded approach to a wider range of respondents, the questionnaire must be sent by mail and this adds to the time to complete the development of the questionnaire.

Survey questionnaires, tests or instructional materials must be clearly presented with white spaces and formatting features that facilitate reading and respect the principles of page layout and readability (Rossett, 1987; Berdie, Anderson & Niebuhr, 1986).

Questionnaires must also be sent to the right people: either to all the people concerned with the issue or to a random sampling. If the questionnaire is sent to a random sample, then the recipients have to be truly representative of the entire population; they must have the same characteristics as the larger group.

Design of the Survey Instruments

Each of the three survey questionnaires is divided into two main topics, one dealing with the present placement test (actuals) and the other focusing on the administrative and pedagogical aspects of the new test (optimals). During the initial development phases, a list of objectives or topics was drawn up and some preliminary questions developed. In addition, because of the number of topics to be addressed, decisions were made concerning who, of all the players in the placement test scenario, could best provide the needed information.

The questionnaires covered the following topics to each of the target groups.

Topics of survey questions addressed to ESL teachers:

- attitudes concerning placement testing and provincial standardization of tests and scores
- level of satisfaction with their present test
- identification of problem areas in their present test
- expectations regarding the new test
- relative importance of language skills
- relative importance of various test components
- question formats (item types) for the new test
- request for volunteers
- familiarity with the present test
- demographic data

Topics of survey questions addressed to ESL Department Coordinators:

- information on the test being used
- attitudes concerning placement testing and standardization
- identification of problem areas in their present test (actuals)
- expectations regarding the new test
- practicality issues concerning the present test
- administration issues concerning the present test
- demographic data

Topics of survey questions addressed to deans:

- need to classify students in languages
- attitudes concerning placement testing and provincial standardization of tests and scores
- level of satisfaction with the administrative aspects of their present test
- identification of problem areas in their present test (actuals)
- expectations regarding the new test
- practicality issues concerning the present test

The needs assessment questionnaires addressed to the ESL teachers and department coordinators were written in English; the Deans questionnaire in French. The three questionnaires, the quantitative data and the accompanying letters are in Appendix A. The data obtained provided important information to the test developers and aided in the design of the test. For comparative

purposes, Table 2 lists the topic areas and the corresponding questionnaire (Dean, CD, teacher) in which the exact questions can be found in order to compare the wording and the data for each of the questions for all the topics. The column head "CD" refers to department coordinator (*coordonnateur du département*).

Table 2

Needs assessment questionnaire topic summary and question numbers

Topic areas	Dean	CD	Teacher
• need to classify in languages	14		
• attitudes concerning placement testing and standardization	14, 15b, 15c, 18		3, 4
• level of satisfaction with their present test (ease and accuracy)	13		14, 17
• need for a new test	15a		1
• expected length of test	19		2
• responsibility for administering and correcting the new test	20a, 20b		5, 6
• responsibility for administering and correcting the present test	1, 2, 3	13, 14, 15 8, 9	
• time, place, number of students tested	4, 5, 6, 7, 8, 9		
• relative importance of language skills; • favoured question formats • relative importance of various test components			9
• type of test (computerized or conventional)	15d, 16, 17		7, 8
• number and type of computers and labs	21, 22, 23, 24	16, 17	
• test in use		2, 3, 4, 5, 6	13
• change of course levels	10, 11		
• policy on cheating (back placement)	12	12	
• failure rate		7	
• reasons for mis-classification		11	15
• number of students misclassified		10f	16
• requests for volunteers to write items, test items with students, and read a draft of the test			10, 11, 12
• familiarity with the present test			18, 19, 20
• demographic data		1, 10	last item

Formative Evaluation of the Needs Assessment Questionnaires

Developing a valid and useful questionnaire requires feedback from a representative sample from the group that will eventually be asked to complete the questionnaire. Feedback is an invaluable asset to the instructional designer in the process of developing materials (Berdie, Anderson & Niebuhr, 1986; Rossett, 1987). Validation of the questionnaires was conducted with participants from each of the groups concerned. In addition to providing feedback on the questionnaire that would eventually be sent to their colleagues, these focus group participants also read and reacted to the questionnaires that were intended for the other interest groups. For example, the teachers commented on the teacher questionnaire as well as on the department coordinator and deans' questionnaires.

Focus groups can open up options and determine a range of alternatives in optimal, solutions and causes (Rossett, 1987). The focus group sessions enabled the test developers to build a reliable questionnaire that would consistently convey the same meaning to all people in the population being surveyed with questions that represent a single meaning (Berdie, Anderson & Niebuhr, 1986). One-to-one and small group evaluations were conducted in the development phase of each of the questionnaires. Once necessary changes were made, the questionnaires were ready to send out to the colleges.

The limited number of colleges made it feasible to send the needs assessment survey questionnaires to all colleges and not just a representative sample. The ability to gather information from the entire population enabled the researchers to get a complete picture of the opinions of all those involved in placement testing ESL in Quebec. Given the number of colleges that responded, the researchers were able to make decisions that would conform to the expectations of the eventual users of the test.

The three survey questionnaires were piloted using the intimate approach (Rossett, 1987). Besides being a faster approach than the expanded, this approach kept the developers in contact with the ESL community who became involved in the test development.

Focus group meetings were held in Montreal, Quebec City, Chicoutimi and Laval. Twenty-six teachers and department coordinators representing 15 different colleges attended these focus group sessions. Others who participated in the validation process in a one-to-one setting were an Academic Dean, an assistant Academic Dean, and individuals from two colleges who have handled the administration of the test for a number of years.

The purpose of the focus group sessions was to

- validate the three separate questionnaires
- obtain information on feelings
- sensitize teachers to item writing concepts
- verify wording and obtain feedback on alternative options for some of the questions
- discuss placement test issues in general to ensure total coverage of all placement test issues in the three questionnaires
- discuss the English course standards and objectives
- obtain feedback on the order of the questions and the coherence of the items and the questionnaire in general
- inform department coordinators about distributing the survey questionnaires to all the ESL teachers in their department
- discuss sample selection and availability for field testing in the spring.

The focus group meetings were held over a four week period in a number of venues. One of the focus group sessions was held during the SPEAQ (*Société pour la promotion de l'enseignement de l'anglais, langue seconde, au Québec*) Annual Convention. Since teachers from various regions in Quebec attend the conference, it was an opportunity to get direct feedback from these teachers who, otherwise, would not have had the financial resources to participate in the other focus group sessions.

The participants discussed all three survey questionnaires and provided important feedback on the questions that were proving to be the most difficult to develop. The major areas of difficulty in the development of the questionnaires centered around these concerns:

- focus group participants favouring open-ended questions rather than a closed checklist type of question

- negative reactions to questions about how familiar teachers were with their present test
- feelings elicited in the person responding to the questionnaire
- wording and clarity of some of the questions
- number of questions and response time
- relevance of the questions.

The focus group participants were also invited to discuss general questions on whether placement test problems have increased since English became compulsory; how their institution has changed placement testing, either by modifying the test, by adding elements or by using other means to place students. Also addressed were issues concerning placement verification and changes of level in cases of mis-classification. Finally, the possibility of modifying an existing test was explored and rejected. It was decided that a completely new test was needed.

The wording of the questions was discussed at length. Questions with only one unequivocal meaning are difficult to formulate. Everyone should understand the same thing when responding to a question. The questions need to be clear and succinct. It was discovered that terms such as course content, programme and curriculum do not hold the same meaning for all teachers. Another problematic area was the wording of the question that dealt with the appropriateness of the current tests. Because the concept of this question confused some respondents, the open question was changed to a fixed response list.

Questions that required further clarification were those related to computer adaptive testing. Computerized testing seemed to hold some meaning for most people but the "adaptive" aspect was more elusive. This form of computerized testing is not well-known in the non-computer community. Most participants understood the concept of a computerized test as simply a paper and pencil test put on computer. A computer adaptive test (CAT) is more than that. The test is "adapted" to the level of the respondent. The differences between the two types are discussed in the chapter on the test design phase.

The questions dealing with student mis-classification involved placement concerns and administrative issues. Question 15, which deals with the reasons

for mis-classification, started off as an open question and was changed to a multiple choice question using the reasons the focus group participants cited: human error, intentional mistakes and not trying hard.

Initially, it was thought that only two questionnaires would suffice; however, in time, it became clear that the teacher questionnaire could be streamlined by putting those questions that a department coordinator could best respond to into a separate questionnaire. Questions that asked for facts and not an expression of an opinion were put into the department coordinator questionnaire. The topics addressed in this questionnaire were more administrative in nature. Similarly, to decrease response time, the deans were not asked to name the test they used. This decision to exclude a question met with criticism from one administrator who wrote that an important question had not been asked.

In some cases, the same questions were asked on two questionnaires. This was done to ensure that the researchers would receive information without too many follow-up calls. However, one of the focus group participants suggested that this redundancy might be interpreted as a lack of confidence in the respondents.

Open-ended questions provide more information but are more difficult to compile. The focus group discussions permitted the developers to change questions from open-ended to multiple choice and checklist type questions. A checklist facilitates responding for new teachers who might be less familiar with the context and decreases the time it takes to complete the questionnaire. In spite of some objections, checklist type questions were adopted and the focus group participants furnished additional options to the checklist. To accommodate those who preferred open questions, a space for comments was included after each of the questions.

It is important not to inadvertently antagonize respondents with some questions. After some adverse reactions to the first five demographic questions, it was decided to amend and move these from the beginning to the end of the questionnaire. There was reluctance on the part of the teachers to include these questions yet the test developers were not willing to do without the information these questions provided.

Changes in format, wording, order and relevance were made to the questionnaires. Some of the other changes consisted of softening the wording of some questions; for instance, "open to" replaced "favour". Redundant or irrelevant questions were eliminated.

The questionnaires underwent numerous changes between September 7 and the mailing date, October 16, 1995. Each of the nine drafts incorporated changes suggested by each of the focus groups. The numerous revisions ranged from those of wording to complete changes of question format and to the elimination of some questions.

Once completed, the questionnaires were mailed to all the Deans and the ESL Department Coordinators of the 56 colleges. Department coordinators were asked to make photocopies of the questionnaire and distribute one to each of the English teachers in their department. A two-week period was allowed for responses before follow-up calls were initiated. When the completed questionnaires were received, each question was coded for computer data entry. The data were then entered into a statistical programme and the results obtained.

Results of the Needs Assessment

Clearly, the development of a placement test elicited much participation and interest. The number of responses surpassed expectations and provided the researchers with the necessary support in the work of development. The promptness of the Deans to respond indicated their interest in the test development and attests to the value of needs assessment in any development project. Table 3 details the responses received from the 56 colleges for each of the questionnaires.

Table 3

Response rate

Colleges	Academic Deans	Department Coordinators	English second language teachers
Public	39	39	213 responses
Private	9	8	26 responses
	48 (85.7%)	47 (80.3%)	239 questionnaires received ²

Testing Instruments and Cut-off Scores

Colleges are presently using a variety of tests for placement purposes. As seen in Table 4, the most widely used test is the *Test de classement d'anglais langue seconde* (TCALS), followed by the Michigan Test of English Language Proficiency and the Comprehensive English Language Test for Speakers of English as a Second Language (CELT).

The Michigan Test of English Language Proficiency is designed as a battery of tests to "estimate student's linguistic readiness to pursue academic study at the university level" (Jenks, 1987 p.58). The Comprehensive English Language Test for Speakers of English as a Second Language is designed for high school, college and adult learners. It measures the English language proficiency of non-native speakers of English at the intermediate and advanced levels (Oxford, 1987, p.22).

² The percent of responses received could not be calculated since the total number of ESL teachers was unknown.

Table 4

Placement tests in use

Test in use	Number of colleges	Percentage
TCALS (50 questions)	2	(4.4%)
TCALS (100 questions)	22	(48.9%)
TCALS (150 questions)	2	(4.4%)
Michigan	5	(11.1%)
CELT	3	(6.7%)
In-house	4	(8.9%)
Other tests	3	(6.7%)
No test	4	(8.9%)
TOTAL	45	100%

Prior to the development of the new placement test, it had always been assumed that all the colleges using TCALS were, in fact, using the same test. In reality, three versions were being used.

Of the 20% who said their test had been modified, most had added a composition component which is not counted as part of the placement test score but used to verify placement. Two other colleges made more substantial changes. One college added some linked listening items and another had added a longer reading text with questions. It is not known if these changes are reflected in the placement test scores of the students. In the case of the former college, the modification was made to an in-house test and, therefore, does not affect any comparison with other colleges. But in the latter case, the change is of concern because the college uses TCALS 100, and given the significance of the change, their data cannot be compared to the other colleges who also use that test.

The colleges were asked to report whether their method of correction had been changed. One college changed the scoring method to eliminate or lessen the effect of student guessing. Six other colleges stated they could not determine if they calculated the scores as recommended in the manual since they no longer had the manual. Another three interpreted the question to mean modifications of the recommended cut-off scores. The question was meant to elicit information about how the correct and incorrect number of responses was calculated. Obviously, the wording of the question created difficulty for the respondents.

Cut-off scores were supplied by the department coordinator in all but three cases. Colleges which use the same test do not necessarily use the same cut-off scores to place the students at one of the five levels of English. In fact, there is a wide variance in the cut-off scores at the different levels even among those using the same test. Colleges using a multi-part test like the Michigan Test of English Language Proficiency may be using different parts.

Tables 5, 6, 7, 8 and 9 report on the cut-off scores for each of the 5 standardized tests used in the colleges as reported by the department coordinators. The tables include cut-off scores for each of the five levels of English courses offered. The column head gives the numbers of the credited courses (from 104, elementary English, to 103, very advanced). Remedial refers to the non-credit level.

Table 5

Cut-off scores for the 3 colleges using CELT

Remedial	104	101	102	103
	0-40	41-65	66-79	80-100
1-39	40-55	56-78	79-98	99-100
	51-75	76-115	116-156	157-

Table 6

Cut-off scores for the 5 colleges using the Michigan Test of English Language Proficiency

Remedial	104	101	102	103
0-23	24-38	39-62	63-85	86-100
0-25	26-38	39-62	63-85	86-100
0-19	20-44	45-64	65-85	86-100
0-40	41-54	55-70	71-87	88-100
0-35	35-60	60-110	110-135	135-150

Table 7

Cut-off scores for the 23 colleges using TCALS 100 (*Test de classement d'anglais langue seconde - 100 questions*)

Remedial	104	101	102	103
0-14	15-29	30-59	60-84	85-100
0-14	15-30	31-59	60-84	85-100
0-44	45-57	58-70	71-84	85-100
0-39	40-55	56-70	71-85	86-100
n/a	0-60	61-72	73-85	86-100
0-39	40-59	60-74	75-89	90-100
0-33	34-51	52-70	71-90	91-100
0-40	41-60	61-75	76-90	91-100
0-39	40-59	60-75	76-90	91-100
0-39	40-59	60-75	76-90	91-100
0-39	40-59	60-75	76-90	91-100
0-39	40-59	60-75	76-90	91-100
0-39	40-59	60-75	76-90	91-100
0-39	40-59	60-75	76-90	91-100
0-39	40-59	60-75	76-90	91-100
0-39	40-55	56-70	71-90	91-100
0-36	37-59	60-75	76-90	91-100
0-39	40-59	60-75	76-90	91-100
0-39	40-59	60-74	75-90	91-100
0-39	40-59	60-75	76-90	91-100
0-45	46-59	60-75	76-90	91-100
0-42	43-59	60-75	76-90	91-100
0-39	40-59	60-75	76-90	91-100
0-35	36-70		71-94	95-100

Table 8

Cut-off scores for college using TCALS 50 (*Test de classement d'anglais langue seconde - 50 questions*)

Remedial	104	101	102	103
0-19	20-31	32-42	43-47	48-50

Table 9

Cut-off scores for the 2 colleges using TCALS 150 (*Test de classement d'anglais langue seconde - 150 questions*)

Remedial	104	101	102	103
0-37	38-47	48-64	65-79	80-100
0-39	40-60	61-75	76-90	91-100

This diverse range of tests and placement scores does little to ensure any form of standardization across the province. A student classified at a 101 level in one college could be placed in a higher or lower level at a neighbouring college. This creates complications when students transfer from one college to another and is one of the reasons respondents gave for wanting a new test.

Need for Placement and a New Test

The Academic Deans clearly support the need for placement testing. A resounding 89.6% said that it was necessary to place students into the different courses according to their level of ability. Teachers were not asked this question since it was assumed that language teachers want to have fairly homogeneous groups. Those who were not in favour of placement testing still expressed support for adequate placement according to the level of difficulty of the courses using alternative means such as high school scores in English second language.

Although there is an expressed need for placement, Table 10 reports on the respondents' opinion about the need for a new test. Some of the uncertainty can be explained by a desire on the part of the deans to place students using less time-consuming means; new teachers likely felt unsure of the need for a new test since they would not be familiar with the current test. Teachers who have been in the system for a number of years suggest that the test in use still suits their purposes.

Table 10

Question # 15a & 1: Do you want a new test?

	Academic Deans	Teachers
Yes	70.8%	77.1%
No	12.5%	2.1%
Uncertain	16.7%	20.8%

Reasons cited in favour of a new test included the need for a more valid and reliable test that could evaluate both oral and written competencies, a test that would be more closely linked to the new standards and objectives. In addition, the Deans would like a test that is short and easier to administer and correct.

A surprising 20.8% of the teachers were uncertain about the need for a new test. None of these respondents provided an explanation for their answer. It

would seem that these teachers are satisfied with the way their present test places the students. Yet, only 73.9% said they had read the test and only 39.4% had also done the test.

Standardization of Placement Test and Cut-off Scores

In addition to agreeing on the need for a new test, teachers and deans believe that standardization of the placement test and of the cut-off scores is a desirable goal. Table 11 reports that both teachers and Deans favour standardizing the placement test.

Table 11

Question #15b & 3: Should all colleges use the same placement test?

	Academic Deans	Teachers
Yes	85.1%	81.0%
No	6.4%	8.9%
Uncertain	8.5%	10.1%

Teachers in favour of standardization cite the need for the different cut-off levels to correspond to the objectives and standards of the new Ministry courses. Without this standardization, reaching the same objectives becomes more difficult. Standardization of levels is desirable not only for students who change colleges but also for teachers working on evaluation, pedagogical materials or course content. Some teachers stated that without uniform levels, setting exit level profiles and developing assessment instruments for each level become impossible. Deans favour standardizing the test and the cut-off scores for similar reasons. Teachers who oppose using the same test in all colleges cite regional differences, maintaining that language levels and skills differ by region.

More respondents are reluctant to standardize cut-off scores than to standardize the test. Table 12 reports on the Deans and the teachers opinions concerning standardizing the cut-off scores.

Table 12

Question #15c & 4: Should cut-off scores be the same for all colleges?

	Academic Deans	Teachers
Yes	78.7%	65.5%
No	10.6%	20.0%
Uncertain	10.6%	14.5%

Teachers fear that standardizing cut-off scores would lead to an unbalanced number of groups of higher level students in the Montreal region as compared to the regions outside the city. Some fear that fewer students scoring into the higher levels would mean that the advanced level might not be offered. In effect, the distribution of the number of groups at a given level would certainly be different from one region to another. But this would not affect the total number of levels offered across Quebec.

By far the most often cited reason for different cut-off scores is that English in Montreal is different from English in other regions. It might turn out that standardization would mean that the very advanced courses would not be offered in some regions or would have a small number of students in the groups whereas in Montreal large numbers of students would place at the very advanced level.

Level of Satisfaction with Present Tests

There was not much certainty among teachers about their level of satisfaction with their present test; 48% expressed satisfaction while 37% expressed dissatisfaction. However, 15%, a rather high percent, did not answer this question, stating that they were new in the system and were not capable of judging. Fifty percent of the teachers who responded to the questionnaire had three or fewer years of experience at the college level - a result of the recent reforms of the college curriculum which made ESL compulsory and required colleges to hire more teachers. Many were not familiar with the test used in their college (48.3%).

Although satisfied with their present test, some teachers still favoured the idea of a new placement test. They found their present tests appropriate in the listening and grammar components but less so for the other parts of the test.

The Deans (83.3%) stated they found their present test easy to administer. This had a direct impact on the design of the test since the practical issues of testing have to be taken into consideration.

Problems Related to Present Tests

Mis-classification is the biggest problem related to placement testing. According to the teachers, two main reasons account for mis-classification: students purposely answering incorrectly to get an easier course (back placement, 78%) and the test does not measure the four language skills (76%).

The test content does not correspond to the course content is cited as the third cause of mis-classification as (39.8%). It should be noted that a placement test measures the level of language proficiency and is not linked to course content. Rather, it answers the question as to whether students can handle the type of material that will be used in the course (Alderson, Clapham, & Wall, 1995).

Department coordinators added clerical errors and errors in data entry as other reasons for mis-classification. Teachers did not rank this as one of the major reasons for mis-classification since they do not usually handle the administration of level changes.

A relatively large number of students are mis-classified . According to teachers, 10% of the students are mis-placed and only 50% of these students change to a more appropriate level. This is due mainly to administrative constraints: scheduling conflicts, delays in level changes and short deadlines. Even when colleges (93.8%) agree to change students' level, these changes must be made within a week (48%) or two (33%) - too short a time. Changes are subject to student timetable constraints since all the courses have already been scheduled into the timetable with few, if any, open periods. The registrar's office does not allow for schedule changes that involve other subjects. Also, some students refuse to change level and others arrange not to be "discovered". Mis-classified students present problems not only to the teacher but also to themselves.

In effect, certain students think that by placing themselves at a level below their true level of proficiency, they have less work to do and are assured high

marks. However, the effects of poor placement on the course are numerous. Mis-classified students cause others who are properly classified to feel discouraged and, at the same time, are themselves bored because they are not learning anything. These students disrupt classes. Those who actually belong in the level begin to feel inadequate. These situations benefit no one: neither the teachers nor the students.

Sixty-nine percent of the departments surveyed have no policy regarding this form of "cheating". However, the reasons described above augur in favour of colleges' adopting a policy that would disallow students from deliberately mis-representing their level of ability and impose a stringent consequence on offenders, the effect of which could reduce and might eliminate the occurrence of back placement.

One college, which requires students to be bilingual, does not experience mis-classification problems. The problem occurs in situations where students have no incentive to take a more difficult course even if they have the necessary language skills to handle the course. These students are more concerned about getting admitted into the university and the programme of their choice. A high z score is the motivating force, not learning. Students believe that following a lower level course will increase their z scores and thereby also increase their chance of being accepted into the university of their choice.

Comments from both teachers and Deans indicate that the choice of the word cheating on the questionnaires was misunderstood. Several stated that cheating was not possible because the test was supervised. The term "back placement" should have been used instead.

Test Content

In responding to questions regarding the composition of the test, and the relative importance of the four language skills, well over 120 teachers want the new placement test to:

- test the four language skills of writing, reading, speaking and listening;
- include a grammar component and audio segments;

- correspond to the objectives and standards of the new English second language courses.

Teachers stressed that it was important that the test be adapted to the target population in terms of age, interests and culture and that it be able to detect borderline cases as well as those who deliberately back place themselves. Some suggested opting for some form of student self-assessment. At the same time, the test must be corrected by the Registrar's office and not take more than 55 minutes. (54.6 minutes according to teachers and 50.7 minutes according to deans)

Administration of the Test

Besides questions dealing with the composition of the test, the surveys also addressed the issues related to the administration of a placement test. Presently, in 68.8% of the colleges, the Registrar handles the administrative aspects of the test and in 75% of the colleges, the test is corrected by members of the administrative staff. The correction of the test is done manually in 54.2% of the colleges. In departments which have added a writing component, the English teachers correct it. Deans want to maintain control over placement testing (63%) and accept the correction of the test as their responsibility in 56.2% of the colleges. If testing must be done, they want to be in charge of it. These figures must take into consideration the fact that some colleges see no need to formally test students to place them in a language course.

Tables 13 and 14 describe the respondents' attitudes to the administration and correction of a new placement test. The option "other" refers to administrative staff other than the personnel in the registrar's office.

Table 13

Question #20a & 5: Who should be responsible for administering a placement test?

	Academic Deans	Teachers
Registrar	63.0%	46.8%
Language Department	15.2%	38.4%
Both of the above	6.5%	9.3%
Other	10.9%	5%

Table 14

Question #20b & 6: Who should correct a placement test?

	Academic Deans	Teachers
Registrar	41.3%	44.7%
Language Department	26.1%	38.3%
Both of the above	2.2%	5.1%
Other	26.1%	11.5%

In many cases, although it is the Registrar who looks after testing, the correction of the test is done by computing services in colleges where the correction is computerized. This accounts for 26.1% of the responses.

An average of 1141 students are tested annually in each college. Most students do the test in a large group setting, such as an auditorium, before their first semester (85.4%). This presents logistical problems to deans who have to notify the new students of the time and dates of the test. They would like to facilitate the process and reduce costs in both financial terms and human resources.

From the outset, the research and development team believed in the feasibility and the necessity of developing a computer adaptive test. When queried about a computerized test, the respondents seemed fairly open to the idea. Table 15 sets out the data regarding computerized testing, which is somewhat surprising given the limited hardware in some colleges.

Table 15

Question #15d & 7: Would you be interested in using computers for placement test-taking?

	Academic Deans	Teachers
Yes	44.7%	57.6%
No	27.7%	16.5%
Uncertain	27.7%	26%

When asked to indicate the preferred type of test, the deans opted for a pen and paper format whereas teachers were almost evenly split between a computerized test and a pen and paper format. In describing feelings, 46.8% of the Deans responded that they had no reservations concerning a computerized placement test. Some teachers expressed concern that the lack of keyboarding skills would hinder students' placement.

The quantitative data reported in Table 16 also includes a not applicable (n/a) cell. The "neither" option was eliminated on the teachers questionnaire since it was assumed that all teachers would favour one form of testing or another.

Table 16

Question #16 & 8: / Type of test preferred

	Academic Deans	Teachers
Pen and paper	52.2%	41.8%
Computerized	39.1%	48.4%
Both	4.3%	9.8%
Neither	4.3%	n/a

Since the development team is especially interested in computer adaptive testing, the college deans were asked to detail the equipment available in their respective colleges in order to judge the feasibility of developing such a test. Many colleges do not have the facilities to conduct large scale computerized testing. Smaller colleges have a limited number of computers and larger colleges say the computers are used exclusively for courses. If computerized testing were implemented, colleges who test large numbers of students at a time would have to re-organize the timing and the scheduling of the testing sessions.

Outcome of the Needs Assessment Phase

Without the needs assessment phase, many of the attitudes and opinions about placement testing would not have been uncovered. Attitudes toward standardization of the placement test and the cut-off scores, for example, clearly revealed the positions of teachers and deans.

Working from the data received and within the imposed time constraints to complete the test (2 years with no extension), the team moved to develop a paper and pencil test rather than a computer adaptive test. The data indicate that respondents want a test that is easy to administer and correct and takes less than an hour to complete. The only type of test that respects those criteria is a multiple choice test. Given the large number of students that must be tested annually, the time and budget constraints, and the insufficient number of computers available in some colleges, it is the only practical solution.

In addition, if staff other than language teachers correct the test, then the production skills of speaking and writing cannot be included. Testing these skills would add considerably to the time it takes to complete the test and correct it and the correction would have to be done by the language teachers themselves.

The needs assessment questionnaire addressed to the teachers was also used as a means to solicit volunteers for various aspects of the development - including item writing and reading drafts or testing items with students. However, less than 25% of the 30 teachers who volunteered to write items submitted any.

The surveys proved to be an effective and productive method of consulting the ESL community and also proved to be a gold mine of information for the researchers. The data presented a clear picture of the present testing situation in the colleges and people's expectations regarding a new test. The data facilitated some of the decision-making in the test design phase of development.

T E S T D E S I G N P H A S E

This chapter discusses the literature on testing, providing some background on testing, and on computer adaptive tests and how these were linked to the needs assessment data leading towards a preliminary test design.

Testing Literature Review

Testing is at the heart of any course and of all teacher-student relationships. It is not separate from teaching and learning but is closely linked to both providing useful information to teacher and students. Testing can take various forms: observations, simulations and pencil and paper tests (Carroll, & Hall, 1985; Harrison, 1983; Smith, & Ragan, 1993).

Purpose of Tests

Although there are many purposes for testing, Smith and Ragan summarize the concept of testing purpose by stating that there are really only two "rather diverse reasons for assessing learners' achievement: (1) to determine level of

competence and (2) to compare or rank learners' abilities." (p.102). Testing may also be used to determine if the objectives of the course have been met; it can be used to improve the quality of the instruction. These purposes allow teachers and administrators to make very different kinds of decisions and can be further divided into more specific categories: diagnostic, achievement, proficiency and placement.

Diagnostic

Diagnostic tests are designed to provide formative evaluation to the students in terms of their strengths and weaknesses. They are based directly on course content and therefore, can inform students as to the progress that has been made and what parts of the learning need to be remedied with further instruction or help. They provide critical information to students, teachers and deans. Progress tests measure what a student has learned so far and as such can be a type of diagnostic test (Alderson, Clapham, & Wall, 1995; Davies, 1990; Harrison, 1983; Henning, 1987; Hughes, 1989).

Achievement

Achievement tests are also referred to as summative tests. The achievement test looks back over a longer period than the diagnostic test (Harrison, 1983). It is concerned with measuring what has been learnt of what has been taught in a content area (Davies, 1990; Henning, 1987). Achievement tests are directly related to language courses and establish how successful individual students have been in achieving objectives. In statistical terms, the best distribution on these tests is a negatively skewed curve which indicates that all students have learned the material (Alderson et al. 1995; Hughes, 1989).

Proficiency

A proficiency test is designed to measure a person's overall ability in a language, regardless of any training in that language. It is based on a specification of what candidates have to be able to do in the language in order to be considered proficient. Proficient means having sufficient command of the language for a particular purpose. These tests are not based on courses but on the ability to use the language in a particular situation, for example, academic

study. These tests are primarily used for placement or selection (Alderson et al. 1995; Davies, 1990; Harrison, 1983; Henning, 1987; Hughes, 1989).

Further to proficiency tests are language aptitude tests which measure a person's language learning ability. These tests seek to indicate capacity for growth. In some cases, they are used in conjunction with placement tests for admission to language programmes.

Placement

Placement tests are not achievement tests, nor are they an exit test from one level of instruction to another. Instead, they test a variety of language-related skills to determine the overall competence of the test-taker. Their main function is as screening, level-finding tests. In some situations, those whose performance is very good are excluded from further instruction. Placement tests are also used for selection purposes to decide who should be admitted to a particular programme of instruction. They are based on instruction that is to come after (Davies, 1990; Henning, 1987) and are "never geared to the learning that went before" (Harrison, 1983). They usually must be done quickly so that teaching can begin (Alderson et al. 1995; Harrison, 1983; Hughes, 1989). Although these tests are useful as an admission and placement tool, they are not designed to be a predictor of academic success (Pike, 1979).

All assessment is comparison and discrimination of one student with others or of one student pre- and post-tested. "The more efficiently a placement test discriminates between students, the easier it is to divide them into groups." (Harrison, 1983).

In a language learning context, these types of tests are designed to provide information to place students according to their level of language ability so that they can be placed into homogeneous learning groups, that is courses that are most appropriate to their abilities. The degree of homogeneity is determined by the number of levels offered within the curriculum. The greater the number of levels, the smaller the range between the different students. The opposite is also true. The fewer the levels, the greater the range of abilities among the students. One thing is certain, mis-classified students do not benefit from instruction, whether it be at a level that is too easy or too difficult for them.

Placement tests are developed by selecting items based upon difficulty and upon discrimination factors which automatically create variability, a desirable characteristic of a placement tool (Dick & Carey, 1990). The normal curve is acceptable in placement testing.

However, the test should be designed with items on an incline of difficulty. Heaton (1988) recognizes the need to ensure that placement tests not be too difficult for beginners nor too easy for advanced students. "Tests need to discriminate among the different test- takers and to reflect the differences in the performance of the individuals in the group."

Test Qualities

A good test must be reliable, valid, and practical in addition to being able to compare and discriminate (Harrison, 1983).

Reliability

An assessment instrument is reliable if it consistently measures what it claims to measure and if we have a high degree of confidence in the scores that it produces. The student's score should be the same regardless of which version of the test is taken. Reliability requires consistent circumstances in which the test is taken and consistency in the way it is marked. The latter is what makes constructed responses such as essays so difficult to assess, compared to more objectively scored test formats such as multiple choice, true-false statements or matching. Inconsistency in scoring, for example, can cause unpredictable error in a test instrument making it unreliable. Other factors that can contribute to error in an instrument are length, clarity of the items, clarity of the directions or problems in the administration of the assessment (Alderson et al., 1995; Harrison, 1983; Smith & Ragan, 1993).

Validity

Validity describes the "extent to which the test measures what it is intended to measure." (Harrison, 1983). The test should have "valid items which stimulate accurate and relevant data." (Berdie et al., 1986). Of the various forms of validity, construct, content and concurrent are among the most important and are necessarily the first to be addressed in any validation study of a test.

Construct validity ties the content of the test to the underlying theory of what is being assessed. It involves assessing "how well a test measures the constructs it is based on." (Alderson et al. 1995). This type of validity assesses how well the test structure matches with the theoretical conceptual framework of what is being measured, language ability, for instance.

Content validity involves what is in the test based on specifications and the purpose of the assessment (Dick & Carey, 1990; Oller, 1978). It verifies that the content of the test is comprehensive and adequately samples the domain being tested (Alderson et al. 1995; Henning, 1987).

Although not a real type of validity (Laurier, 1997), face validity assesses what teachers and students think of the test. It answers to the acceptability of the test in the eyes of teachers, parents, students or deans as to whether it is a reasonable way of assessing students and the relative level of difficulty of the test (Henning, 1987). It is pre-scientific and based on perceptions.

Validation studies of assessment instruments come from a variety of sources both internal to the test and external to it. Internal validation studies compare the various sub tests and items to each other. External validation studies include subject matter expert review or a comparison with another test that has the same approaches to measurement.

Validity is related to test purpose and content. Writing item specifications and test blueprints are ways to ensure a valid test. Smith & Ragan (1993) state that an instrument has content valid if

1. its individual items are consistent with the objectives they claim to assess;
2. the items for each objective are representative of the range of items that are possible to develop for that objective;
3. objectives upon which the instrument is based are adequately sampled.

Practicality

The third quality of a test and as important as validity and reliability is practicality. Smith, & Ragan (1993, p. 9) call it "the realities of the assessment situation." These are the mundane yet necessary administrative concerns such

as costs to purchase, score, interpret and report the scores; costs related to test security and material needs such as cassette recorders, pencils, answer sheets (Harrison, 1983; Henning, 1987; Pike, 1979).

Methods and Forms of Testing

Tests can be of different types. They can be criterion-referenced or norm-referenced; direct or indirect; discrete or integrative; objective or subjective. Each of these aspects of testing could be used for any of the types of tests discussed above: diagnostic, achievement, proficiency and placement.

Criterion-referenced vs. Norm-referenced Tests

The notion of norm-referencing is based on the normal distribution curve of attributes and abilities. It makes use of rank order; it places individuals on a scale: students are compared to each other. Norm-referenced tests are useful for selection decisions but not useful in determining whether an individual is competent in a particular skill or possesses particular knowledge (Smith & Ragan, 1993).

On the other hand, criterion-referenced tests rank by criteria or goals which are task related. The emphasis is on discriminating tasks to verify the number of objectives passed. These tests are useful in determining whether an individual is competent in a particular skill or possesses particular knowledge. Criterion-referenced tests are written before instruction begins. Students are compared to a degree of mastery and not to cut-off criteria. The mastery level is often set at 80-90% (Henning, 1987).

According to Davies (1990), criterion-referencing has always been a particular use of norm-referencing. Both are two sides of the same phenomenon. Norm-referencing at some point always uses criterion-referencing in order to determine a cut-off, a level that needs to be reached for some purpose like placement, for instance. Any cut-off is a criterion. And criterion-referencing requires norm-referencing in order to establish just what it is learners are capable of, what the best can do in a limited amount of time.

The major difference between criterion-referenced and norm-referenced tests is how student performance is interpreted (Dick & Carey, 1990, p.142;

Hughes, 1989). Student performance can be described in both criterion and norm referenced manners. Norm-referencing compares a candidate's performance to that of other candidates. Criterion-referencing relates the performance to what the candidates can do with the language. In criterion referenced tests, one does not get a normal distribution curve, but rather, a skewed distribution because the test writer expects the students to have achieved mastery (Smith & Ragan, 1993).

Dick and Carey (1990) explain that standardized tests are constructed by selecting items "based not only on the domain but also on item statistical properties designed to automatically create variability in student performance." They question the test companies who report scores as criterion-referenced or mastery level since it indicates achievement in part, but not all, of the instructional goals. Test manuals that describe item selection based on "difficulty and discrimination factors as criteria for selecting items for the test" are, in fact, norm-referenced tests.

Normed or standardized tests must have been previously administered to a large sample of more than one thousand people from the target population (Henning, 1987). Acceptable standards of achievement are determined only after the test has been developed and administered. Standards are found by reference to a mean of other students from the same population. Items of various levels of difficulty are purposely included since the test must discriminate between the low-achieving and high-achieving students.

Standardized tests have both strengths and weaknesses. On the one hand, they are useful to compare students; they provide confidence in the results obtained since estimates of reliability and validity are provided; and research is replicable. On the other hand, standardized tests are valid only for the population on which they have been normed; and since they are developed independent of a course of instruction, it is difficult to match results to instructional objectives.

Direct vs. Indirect

Direct methods of testing require candidates to perform precisely the skill to be measured: for instance, write a composition or speak on a topic. It is limited

to a small sample of tasks (Henning, 1987; Hughes, 1989). It is difficult to infer from these tests whether the students will perform as well on similar tasks in different situations.

Indirect testing methods attempt to measure the abilities underlying the skills we are interested in by testing a larger sample of the abilities that we believe underlie the skill. For example, a random cloze test would be considered an indirect form of assessing writing. But good scores on indirect tests do not accurately predict scores on a direct measure of a skill.

The main problem with indirect testing is that the relationship between performance on these types of tests and the actual performance of the skills in which we are usually more interested tends to be rather weak and uncertain. Indirect measures do not yield the expected high correlations with the construct of productive language skills (speaking and writing). For example, correlations of the TOEFL test with the direct measure of writing compositions did not yield high correlations. Similarly, in 1961, Lado included test items which asked candidates to identify words that rhyme with each other. He was unable to obtain high correlations with the language construct of pronunciation (Hughes, 1989). In spite of this lack of correlation to the actual skill, the results from these tests are more generalizable than results from direct forms of testing.

Discrete vs. Integrative

Discrete point testing assesses only one element at a time, item by item. Items on a particular word or grammar point are but two examples of discrete point testing. Each item is independent of the others so that the order of the items can be changed or one of them can be omitted without influencing performance on the other items. One correct response does not depend on a correct response to an earlier item (Carroll & Hall, 1985).

Integrative tests require the candidate to combine many language elements in the completion of a task: write a composition, make notes from a listening passage, take a dictation, complete a random cloze. Integrative tests have sets of items (Carroll & Hall, 1985; Henning, 1987).

Statistically analyzing the items on integrative tests becomes problematic especially in cases of computer adaptive testing which relies on the building of

an item bank. Furthermore, it was found that there were “no statistically revealing differences” (Farhady, 1979) between discrete point and integrative forms of testing (Henning, 1987; Hughes, 1989).

Objective vs. Subjective

The distinction between objective and subjective testing lies in the methods of scoring. In subjective scoring, judgement (and, therefore, expertise in the content area) on the part of the rater is needed whereas objectively scored tests compare examinees' responses to a scoring key. This means that objectively scored tests can be graded by man or machine since no knowledge or training in the content area is required on the part of the scorer (Hughes, 1989). It is because of this that objectively scored tests are more reliable measures (Henning, 1987).

In summary, direct tests that are integrative in terms of language tasks and approximate "real" communication are also subjective in terms of scoring, thereby raising reliability concerns. Indirect tests, although more objective and most likely, more reliable, require discrete point testing.

Pencil and paper vs. Computer adaptive

Traditionally, tests have been done using pencil and paper and correction undertaken in a variety of ways, including the use of answer key templates. With the introduction of computers, data were entered into a program and machine-corrected. After the development of optical scanners, correcting became automated. A traditional pencil and paper test can be entered in a computer program which displays each item at random or in a pre-determined sequence and then, scores the responses.

Computer adaptive tests are computer based but the items are displayed on screen at random and based on the test taker's response to the previous item. If the test taker misses an item, the next item to be presented will be slightly easier, while correct responses will lead to the presentation of progressively more difficult items (Alderson et al., 1995; Bachman, 1990; Cohen, A, 1994; Dandonoli, 1987).

Underlying a computer adaptive test (CAT) is an algorithm that manages a calibrated bank of items to present questions on screen within a certain level of difficulty based on the respondent's previous answer. The test is "adapted" to the level of the respondent. Students whose response patterns are erratic can be flagged for further testing. Adapting is impossible to do with a paper and pencil test that is corrected by hand or by optical scanner.

Non-computerized testing cannot spot the anomalies in response patterns that indicate that the students might be trying to back place themselves, for instance.

Test Design

The design of the test is based on both testing principles and theory and on the data obtained from the needs assessment questionnaires. The needs assessment data clearly indicate that a priority for both deans and language teachers is a valid and reliable test that efficiently tests large numbers of students. The test needs to be relatively easy to correct and administer since it will most likely be corrected by the Registrar's personnel and not by language experts.

The development team concluded that the most efficient way to test the large number of students was to develop a test using multiple choice items. But the multiple choice question format has drawbacks. One is that cheating non-verbally during a test is possible unless candidates are given different versions of the test (Hughes, 1989). Although it is a factor to be considered in most testing situations, it is not an issue in this context. Of the problems colleges face, cheating of a different sort seems to be more prevalent. Many students back place themselves by deliberately choosing the wrong answer to some of the questions in an effort to lower their placement results. This form of back placement has also been referred to as sandbagging or fudging.

Guessing constitutes another important factor in the design of a multiple choice test. If three choices are offered, students have a 33% chance of choosing the correct response. The inclusion of four-option fixed responses raises the level of difficulty of the multiple choice format. This curbs the effect of guessing by decreasing the probability of getting the right answer by chance

(Heaton, 1975). The test, then, will be comprised of four-option multiple choice items - the correct answer and three distractors. Each of the options will be in alphabetical order and in equivalent length and form.

Pike (1979) reported that Educational Testing Services (ETS), the developers of the TOEFL test (Test of English as a Foreign Language), studied the possibility of changing the format and the content of the TOEFL; it was concluded that many ostensibly reasonable suggestions for change turned out not to be valid or feasible after all (from the foreword by John B. Carroll). In addition, Henning (1987) states that test format must be familiar to students; otherwise, performance will suffer, altering the nature of what is being measured by the test.

Nonetheless, the test design did include some tasks that are less usual: a mini-talk in the listening comprehension section and error analysis items in the written section of the test. The comments received from the teachers and the students in the formative evaluation phase of the development support the desirability of familiarity since the sub test that created the most difficulty was a task that was unlike the others.

The design of the test is indirect and based on discrete point items that can be objectively scored. The test should contain enough material to test economically the skills originally specified in the curriculum and should have enough items to ensure that there are items corresponding to each of the levels in the curriculum. The items should be set on an "incline of difficulty" from easy to more difficult with questions for students of all levels to satisfy the need for variance in placement testing (Carroll & Hall, 1985; Harrison, 1983).

The test content was determined by the standards and objectives set out in the Ministry of Education document entitled *Des Collèges pour le Québec du XXI^e siècle: Formation générale*. Three of the four credited levels of English are included in the document. The lower level course and the remedial course were developed after the Ministry document had been published and after the Ministry realized that three levels of a second language did not suffice given the wide variance in the language proficiency of the students across the province. These additional course descriptions were sent to the colleges as annexes to the original document. The test content is general and within the range of the

students at each of the levels and covers the major objectives of the ESL curriculum.

The analysis of the standards and objectives led to the development of the matrices by language skill. The matrices in Appendix B illustrate the progression in the level of difficulty of each of the courses. The test has to place the students according to these levels. The course descriptions also involve a progression in the level of difficulty of the grammar and in knowledge of vocabulary. It should also be noted that the lower level courses emphasize listening and speaking skills.

The directions written for the test constitute an important part of the validity of the test. Smith & Ragan (1993) state that "if a learner has acquired the skill then nothing in the test should interfere with the learner's demonstrating that skill." The directions should be clear and concise and provide all necessary information such as time limit, word limit and any special needs. There should also be instructions as to whether the students should guess. If possible, the directions should be written in the students' own language using familiar terms (Dick & Carey, 1990, p.134). For these reasons, the directions for the test are written in French with examples in all sub tests with the exception of the mini-talk. The inclusion of a sample mini-talk example would increase the time needed to complete the preliminary version of the test, which already exceeds the bounds of what ESL teachers considered acceptable.

The design of the test focuses on listening comprehension and written English. There are nine sub tests: three listening comprehension sub tests (statements, dialogues and mini-talks), two structure sub tests (grammar and error analysis), two vocabulary and two reading sub tests.

Figure 1 illustrates the design of the test combined with the test tasks. The dotted lines represent the production skills that are not included in the design of the test given the decision to develop an objectively-scored test.

Figure 1. Test design

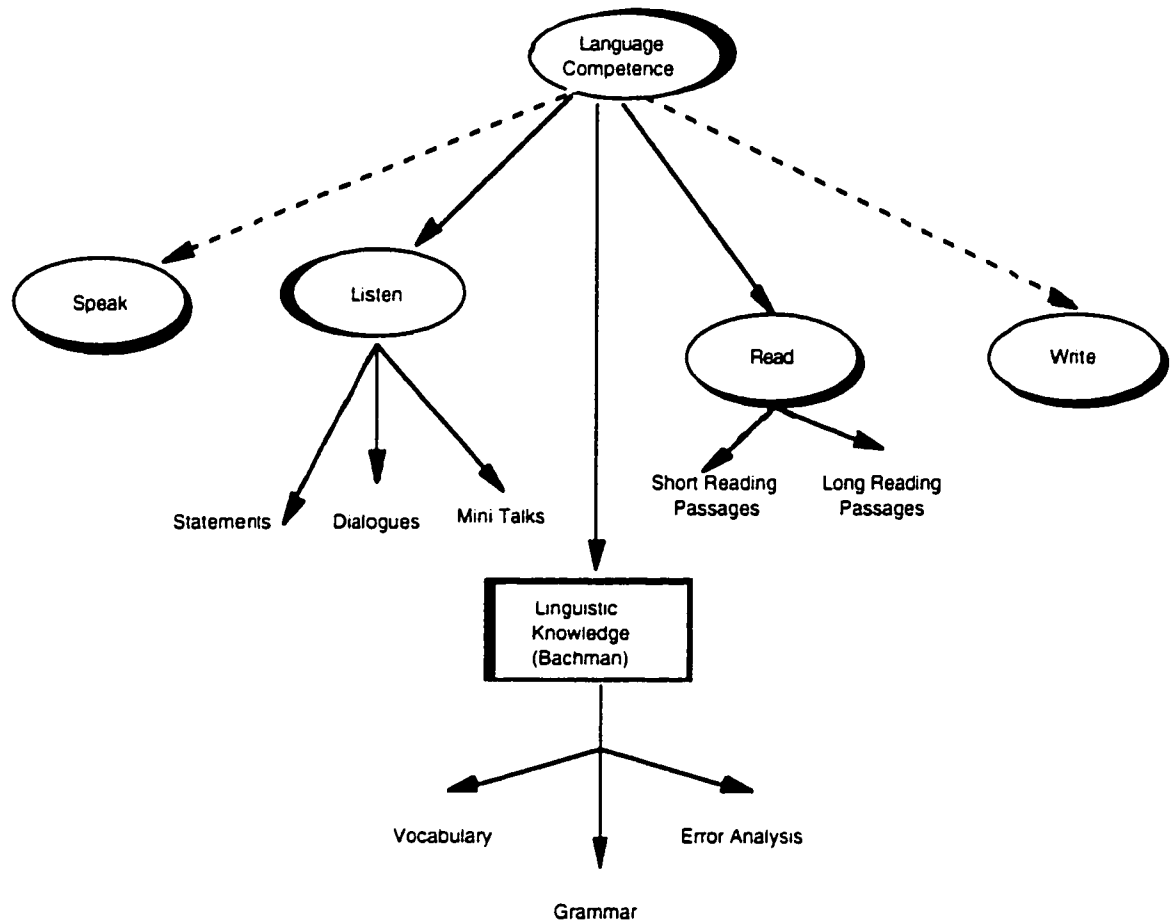


Figure 1. Test design: language skills and related test tasks.

In the listening comprehension component, the statements, dialogues and mini-talks are heard only. Four different voices recorded these items. Only the statements are repeated; the other item types are not. The statements and dialogues each have one question, while the mini-talk has four linked questions. Students must keep pace with the speed of the recording and the time allotted to read the question and the four response options in the test booklet. The number of the item is given on the recording and is followed by a period of silence to allow students to read the question and the four options. Students listen and then have time to inscribe their choice on the answer sheet.

Statements range in length from 7 to 19 words: an average of 15 words. Because they are so short, they are repeated. Dialogues take place between a

man and a woman and very in length from 17 (one item) to 57 words (two items) with an average of 36 words. The question related to the dialogue is written. The mini-talks range in length from 105 to 186 words and have four linked question items. The mini-talk that was used as an anchor item in each of the forms has 105 words.

Students complete the written component of the test at their own pace. No end time is set because the students must have enough time to finish so that data can be collected on all the items. The vocabulary items require two different tasks: fill in the blanks and choosing equivalent forms. They test basic to advanced vocabulary. The error analysis sub test requires students to choose the correct sentence from four options. The grammar section includes items on prepositions, verb tenses and verb forms (from the simplest to the more complex structures). The short reading selections have an average of 40 words; the longer texts range from 142 to 294 words.

In all, 244 items were written for the test: 78 for the listening component and 166 for the written. The total items were then organized into three tests with anchor items in each of the sub tests on an incline of difficulty. Each sub test starts with an easier item, progresses to more difficult ones and ends with an easy item.

The items had to be divided into three separate tests since students could not be asked to do all of the items. The anchor items, the same item repeated in the three forms of the test, were used to compare the data from different samples during the eventual large scale field testing. They provided the basis for the eventual application of item response theory. However, for the purposes of the preliminary field testing, all the items, including the anchor items are analyzed using classical item analysis (Roid & Haladyna, 1982; Alderson et al., 1995).

Table 17 details the design of the test with the items separated into three forms of the test. Each form included different items in addition to the anchor items. In the listening comprehension component of the test, students did a total of 36 items. In the written component of the test, students completed 74 or 75 items depending on the form. Form 3 had a total of 110 items; the other two forms had 111, the grammar section having an unequal number of items.

Table 17

Design of the placement test

Listening Comprehension

Sub Test	Task	Form 1	Form 2	Form 3	Anchor items	n=
1	Statements	8	8	8	6	30
2	Dialogues	9	9	9	5	32
3	Mini-talk 1				4	
3	Mini-talk 2	4	4	4		16

Written Comprehension

Sub Test	Task	Form 1	Form 2	Form 3	Anchor items	n=
4	Vocabulary 1	9	9	9	8	35
5	Vocabulary 2	4	4	4	4	16
6	Grammar	18	18	17	8	61
7	Error Analysis	8	8	8	3	27
8	Short Reading	3	3	3	2	11
9	Long Reading 1				4	
9	Long Reading 2	4	4	4		16

FORMATIVE EVALUATION PHASE

In this phase of the development, the preliminary version of the test was piloted under conditions approximating those of the final version. At two sites, students responded to a formative evaluation questionnaire. This chapter deals with the literature on formative evaluation, sample selection, the methodology and the data analysis of the field test.

Formative Evaluation Literature Review

The main purpose of the formative evaluation stage of development according to a variety of authors is to obtain feedback on instructional materials in order to improve them. Authors deal mainly with the formative evaluation of instructor-delivered materials, but the same principles can be applied to test development. Authors insist that questions be asked at each stage of the test development from design reviews, expert reviews, and field trials leading to a continual programme of test improvement (Carroll & Hall, 1985; Dick & Carey, 1990; Smith & Ragan, 1993).

The various phases of formative evaluation are

- one to one interviews
- small group data collection 8-20
- field trial with approximately 30 participants (Dick & Carey, 1990).

The one-to-one interviews help to revise materials in the early stages of development, whereas the small group data collection is normally done when the materials are ready for pilot testing in a field trial. Field trials need to be conducted in a manner "that closely resembles that intended for the ultimate use of the instructional materials." (Dick & Carey, 1990).

Alderson et al. (1995) define pretesting as all the trials of a test that take place before it is launched or becomes operational. Most of the pretesting takes place during the main trials but these should be preceded by less formal pretesting, which we call pilot testing, with a few colleagues or by running a trial on about a hundred students.

In addition, Alderson et al. (1995) reason that we cannot know what a test is doing until it has been tried out. The data from pretesting reveal how difficult and how clear an item is. The data also let the developer know if the items work. In placement testing, an item works if it succeeds in distinguishing students at different levels; the more proficient students can answer it correctly while the weaker ones cannot.

According to Dick and Carey (1990), instructors and test designers should ensure that

1. test directions are clear, simple and easy to follow;
2. each test item is clear and conveys the intended information or stimulus;
3. conditions under which responses are made are realistic;
4. the response methods are clear;
5. appropriate space, time and equipment are available to respond appropriately.

The literature on instructional design stresses the need to evaluate formatively all aspects of the materials before a final product is published. The best source of information to judge the effectiveness of an instrument, whether a survey, a test or instructional materials, is the intended audience of the specific instrument (Carroll & Hall, 1985; Dick & Carey, 1990; Smith & Ragan, 1993; Alderson et al., 1995). The same principles underlying the formative evaluation phase of the needs assessment questionnaire can be applied to the formative evaluation of the items and the test itself.

Methodology

Both the survey and the preliminary version of the test were formatively evaluated with participants from each of the groups, that is teachers, deans and students. The formative evaluation of the test was conducted with some ESL teachers and with students to verify the following aspects of the test:

- test design, soundtrack and instructions
- practicality and administrative concerns
- analysis of the items: item difficulty and level of discrimination.

Test items and directions must undergo formative evaluation before being used to assess student performance (Dick & Carey, 1990). In addition to the data needed to determine which of the items were working to discriminate between levels, the test itself had to be globally evaluated. Preparation of the items for large scale field testing included two preliminary phases, the first less formal and structured than the second.

Items were first pre-tested at Collège De Maisonneuve in Montreal with a small group of students using an audio language laboratory that analyzes student responses to multiple choice questions. The small informal testing was conducted with two groups of advanced students, who were expected to have little, if any, difficulty with the items. Interestingly, one item proved difficult for some of the students. The item required students to understand time-telling and from that understanding calculate a number of minutes. The difficulty for some students lay in the non-linguistic sub objective: mathematics. Without preliminary testing, this difficulty would have gone unnoticed and the flawed item would have been included in the field testing. The item was subsequently re-written.

A second phase followed with a larger sample (n= 299). This pilot testing provided the information that allowed the authors to edit or discard bad items (Harrison, 1983). It would provide the fine tuning necessary before undertaking the large scale testing scheduled for the second year of the development project.

An unforeseen third phase became necessary after the data of the field test revealed that an insufficient number of students had responded to form 2. Since the items needed to be validated before inclusion in the large scale field test versions, more data on these items were required in order to make informed decisions.

Sample

The sample included 299 students at the five levels of language proficiency from the various regions in Quebec and from colleges in the public and private sectors selected on a proportionate and representative basis. The students involved in the study were attending a French post secondary college and were

completing the first of two compulsory English second language courses at one of the five levels offered. The students were either in their first or second year college.

A larger sample would have provided more precise data but timing affected the size of the available sample. The items were not ready for testing until very close to the end of the winter semester, too close to the end of semester to ensure large scale participation. In addition, most colleges normally give the general English course in the fall semester and not in the winter semester. This fact alone eliminated a lot of potential test sites. Ideally, testing should have been postponed to the beginning of the following semester, but the tight deadlines did not allow for such a change in the timetable.

The *Annuaire des cégeps 1995-1996* was the source for the data on regional divisions and the total population for each college that is part of the *Fédération des cégeps*. Table 18 details the total population of those colleges in each of the regions but does not include some of the smaller colleges.

Table 18

Student population by region

Region	Public	Private	Total	% Public	% Private	Total
Montreal area	63,399	9,276	72,675	42.84%	6.27%	49.11%
Québec area	24,482	2,490	26,972	16.54%	1.68%	18.23%
Estrie	7,321	400	7,721	4.95%	0.27%	5.22%
Centre du Québec	10,358	1,750	12,108	7.00%	1.18%	8.18%
Cote-Nord	1,912		1,912	1.29%	0.00%	1.29%
Nord-Ouest*	2,830		2,830	1.91%	0.00%	1.91%
Outaouais*	4,307		4,307	2.91%	0.00%	2.91%
Bas-Saint-Laurent--	9,056		9,056	6.12%	0.00%	6.12%
Gaspésie						
Saguenay-Lac-Saint-Jean	10,396		10,396	7.03%	0.00%	7.03%
TOTAL	134,061	13,916	147,977	90.60%	9.40%	100%

*This region was divided into two because, although geographically connected, contact with English is completely different. The Outaouais region is

near Ottawa, Ontario - a bilingual capital city while Nord-Ouest is in an exclusively francophone area.

Representative sampling was necessary since it has often been said that English differs according to the region. In order for the sample to be truly representative, it was essential for the sampling to be proportionate to the regions. However, given the number of placement tests used and the lack of correlation studies between them, it was necessary to control this variable. Only those colleges who use TCALS 100 were initially selected to test the preliminary items. TCALS was chosen as a control variable simply because it allowed for a larger number of test takers. Since the experimental testing would eventually be conducted with a sample of 2000, it was essential that sample selection variables allow for the largest number of potential test sites to obtain as large a pool of students as possible.

Table 19 details the composition of the sample. There were six test sites involved from Quebec City and four other regions and the sample was drawn from private and public colleges.

Table 19

Sample by college and by region

The following colleges participated in the preliminary testing of the items.

Collège de Rimouski	Bas-Saint-Laurent- Gaspésie	Public
Collège de Sherbrooke	Etrie	Public
Collège de Limoilou	Québec	Public
Collège Lionel-Groulx	Montreal	Public
Collège Laflèche	Centre du Québec	Private
Collège de Chicoutimi	Saguenay-Lac-Saint-Jean	Public

Pilot testing, initially scheduled for April, had to be postponed until the beginning of May 1996. The imposed timetable and the limited participation from volunteer item writers did not allow for sufficient time to develop the large number of items required.

Pilot testing was conducted by the on-site course teacher following detailed instructions from the researchers on the procedures to follow during the testing.

The researchers were unable to conduct the testing because the students were tested in many colleges across the province and because of time and budget constraints.

Each test site received sufficient copies of the test booklet, computerized answer sheets and ethics letters. They also received an audio cassette for the recorded parts of the test as well as a return address sticker. Procedural instructions were included. Appendix C contains copies of the documents that were sent to the test sites excluding the actual tests. Research ethics procedures were respected in all cases. A letter addressed to the students describing the research was presented to them prior to the actual pilot test with assurances that the data would be kept confidential.

In preparation for the testing sessions, teachers had to arrange for a suitable audio tape cassette player. Ninety minutes had to be set aside for the testing, time that would include receiving instructions, completing the identification section of the computerized answer sheet and signing the letter of agreement to participate.

Students were each given a test booklet for the test form they were to do and a computerized answer sheet. It was not possible to distribute the three forms of the test randomly within one class group because of the recorded listening comprehension component. After the first three recorded listening comprehension sub tests, students completed the written part of the test at their own speed and handed in the materials when they finished. The test site teachers were asked to complete a questionnaire on the face validity of the preliminary version while supervising the test. They also listed the names of the students and each student's start and end time as well as any particularities that arose during the testing session that could possibly affect performance on the test.

The materials were then returned to the research team for data compilation and analysis. The multiple choice items were scored using an optical scanner at the Data Processing Laboratory at l'Université de Montréal.

Two test sites in two different regions of Quebec agreed to conduct the formative evaluations (n = 118). Students were asked to respond to a fifteen-

item formative evaluation questionnaire after completing the placement test items. The questionnaire, Appendix D, written in French to facilitate comprehension, was completed after the test booklet was handed in. Students responded to each statement on a five point Likert scale ranging from *Tout à fait en désaccord* to *Tout à fait en accord*.

Results of the Pilot Testing

The pilot testing provided important information in two key areas: how the students perceived the test and how well the items were working. The questions for the formative evaluation phase were limited to those parts of the test development that had raised concerns for the researchers such as the listening comprehension sections of the test since these sections require a certain minimum language proficiency.

The evaluation statements were numbered from 151 in order to clearly separate the responses to the formative evaluation questionnaire from the responses to the test items which were numbered from 1 to 110. The responses on the five point scale are detailed in Appendix D and summarized here. The student responses on the five-point Likert scale are expressed in percentage. For the purpose of this discussion, the two points on the Likert scale with the highest percentage endorsement are added and reported.

The quantitative and the qualitative data of the formative evaluation questionnaire indicate that the test was generally well received by the students. In general, students found the test acceptable in terms of the level of difficulty (63.56%). The clarity of the written instructions, the usefulness of the examples and the soundtrack ranked high on the Likert scale. Students did not have as much difficulty completing the written sections because each had an associated example and were not timed. The written instructions were found to be clear (83.05%), as well as the voices of the males and the females on the audio tape (72.03% and 68.64% respectively). The soundtrack had a 78.81% satisfaction rate and the questionnaire was clear and easy to read and follow (83.05%).

This level of satisfaction was not the case for the mini-talk section of the listening comprehension component. Although 72.89% of the students found the examples useful, most students would have preferred an example for the

mini-talk section. Students disagreed with the statement that an example was not needed in the mini-talk sub test (72.03%). Only 27.96% agreed with the statement. The qualitative comments included in Appendix E bear this out as well as the feedback from the teachers who said the students were not prepared for the four linked items that were based on the mini-talks.

Some responses can be explained by the students' level of English. The data for question 158 dealing with time allotted to read the item reflected the overall composition of the sample. Only 24.58% completely agreed that sufficient time was allotted to read the question and the four possible answers. The need for more time to read questions reflects the linguistic proficiency of the students. In addition, five seconds was not sufficient to fill in the response sheet for 30.5% of the students. A longer lapse of time must be allotted to read the item and to fill in the answer on the answer sheet for the large scale testing.

Question 157 on the speed with which the statements, dialogues and mini-talks were delivered obtained an almost even distribution across four of the five points on the Likert scale ranging from 20.34% to 26.27%. These questions would necessarily produce these data given the range of language ability of the sample.

The data on the items for placement testing provide information on how difficult the test is and how well it distinguishes the better students from the poorer ones (Harrison, 1983). Tables 20, 21 and 22 report the descriptive statistics for each of the nine sub tests for each form of the test. The sub test numbers correspond to the list of tasks found in Table 17 which describes the design of the test. Given the range of language ability of the students which this test has to place, the test was expected to generate low means. The mean score describes how difficult the test is (Harrison, 1983).

Table 20

Descriptive statistics for the 9 sub tests in form 1

Sub tests	1	2	3	4	5	6	7	8	9
Number of items	14	14	8	17	8	26	11	5	8
Number of examinees	105	105	105	105	105	105	105	105	105
Mean	5.99	5.75	2.03	7.66	3.16	11.80	4.15	1.41	2.26
Variance	7.02	7.42	1.63	11.31	2.52	30.12	5.65	1.70	3.89
Std. Dev.	2.65	2.73	1.28	3.36	1.59	5.49	2.38	1.31	1.97
Minimum	1	0	0	0	0	0	0	0	0
Maximum	12	13	6	17	8	24	9	5	8
Median	6	5	2	7	3	11	4	1	2

Table 21

Descriptive statistics for the 9 sub tests in form 2

Sub tests	1	2	3	4	5	6	7	8	9
Number of items	14	14	8	17	8	26	11	5	8
Number of examinees	120	120	120	120	120	120	120	120	120
Mean	8.79	8.41	4.06	10.37	3.80	15.61	6.97	2.84	4.44
Variance	6.63	7.13	2.88	9.02	2.30	25.91	3.37	2.10	3.60
Std. Dev.	2.58	2.67	1.70	3.00	1.52	5.10	1.84	1.43	1.90
Minimum	2	1	0	2	1	4	2	0	0
Maximum	14	13	7	17	7	24	11	5	8
Median	9	9	4	10	4	16	7	3	4

Table 22

Descriptive statistics for the 9 sub tests in form 3

Sub tests	1	2	3	4	5	6	7	8	9
Number of items	14	14	8	17	8	25	11	5	8
Number of examinees	74	74	74	74	74	74	74	74	74
Mean	7.39	7.53	3.16	10.08	3.99	14.58	6.65	3.08	4.89
Variance	15.75	14.36	3.81	21.45	2.31	43.03	8.82	1.94	5.20
Std. Dev.	3.97	3.79	1.95	4.63	1.52	6.56	2.97	1.39	2.28
Minimum	1	0	0	2	1	2	0	0	0
Maximum	14	14	8	17	7	25	11	5	8
Median	8	8	3	10	4	14	7	3	5

In order to collect data on each of the items, a classical item analysis (Roid & Haladyna, 1982; Alderson et al., 1995) was conducted. For each item and for each of the four possible responses, the proportion endorsing and the point biserial correlation were reported and analyzed. The proportion endorsing is the percentage of students who chose the correct answer. It describes the facility- difficulty of the item. The difficulty of an item, called the 'facility index', is the percentage of students who got the right answer (Harrison, 1983).

The biserial, a value ranging from -1 to + 1, is the correlation between the item and the total score on the particular sub test. The biserial indicates whether the item discriminates. If low scorers get an item wrong and high scorers get it right, the item is said to discriminate. One expects high scorers to get easy items correct. If all the items in a test discriminate, then the test divides the students reliably. This is justification for accepting or rejecting items at the review stage (Carroll, & Hall, 1985). If the value of a distractor is greater than .30, it is an effective distractor. One expects negative-value distractors and positive-value answer. This means that the best students got the item right.

The data obtained for each item provided the basis for decisions concerning the fate of each item: acceptance, rejection or revision. Usually, only items that have values of > .30 are kept but sometimes, items that have lower values are kept for other reasons, such as content and extreme difficulty or facility. An item is not rejected on statistical grounds alone. An item that tests a necessary

language objective can be kept or modified (Carroll, & Hall, 1985). A distractor can also be kept, revised or rejected if it has attracted at least 5% of the low-scoring students. An item is rejected if the statistics are so poor as to make re-writing useless or impossible. Revisions included re-writing the stem and, in a few cases, the four options.

Tables 23, 24 and 25 illustrate how each of the items was analyzed using sample items which were eventually discarded. Some items could not be revised, and others tested an aspect of English that even advanced students have not yet mastered.

Table 23

Sample Item 1

	Prop. Endorsing	Biserial
Before going into the States, you have to go through the _____.		
a. boundaries	0.076	0.133
b. customs	0.238	0.252*
c. frontiers	0.476	-0.319
d. lines	0.181	0.317

The asterisk indicates the correct answer, B. However, distractors A and D have positive biserial values meaning that these options attracted students who had a high global score on the test. (7.6% and 18.1% respectively). It is essential that distractors have a negative correlation, that these options be attractive only to those students who score low on the test. Clearly, the vocabulary in this item does not discriminate among the levels and is best eliminated from the test.

Table 24

Sample Item 2

	Prop. Endorsing	Biserial
If you really intend to hand in all your work on time, you should not postpone _____ your assignment.		
a. to do	0.486	0.301
b. doing	0.286	0.199*
c. do	0.086	-0.235
d. did	0.086	-0.324

The asterisk indicates the correct answer, B. However, distractor A attracted many students (48.6%) who had a high global score on the test. In fact, a larger percent of the students chose the incorrect option than the correct answer. The use of gerunds does not seem to discriminate among the levels since even advanced students had difficulty choosing the right answer. The item is best eliminated from the test.

Table 25

Sample Item 3

	Prop. Endorsing	Biserial
a. How many homework do you have this week?	0.243	-0.359
b. How many homeworks do you have this week?	0.486	0.030
c. How much homework do you have this week?	0.176	0.732*
d. How much homeworks do you have this week?	0.068	-0.217

The asterisk indicates the correct answer, C. Only the very advanced students got this item correct as seen by the high biserial value (0.732). Fewer than the expected 25% which can be accounted for by chance got this item right. Distractor B attracted 48.6% of the students who had a high global score on the test. Non-countable nouns clearly have not been mastered even by

advanced students. This item is best eliminated from the test since it cannot successfully be re-written.

Twenty-six items were discarded. These items tested advanced vocabulary and verb forms, prepositional phrases and "faux amis" - words that are similar in French and English but have different meanings. Examples of "faux amis" are words like "apartment", "sensible", and "comedian". Items that were proven to be highly discriminating but extremely difficult were also eliminated since the curriculum does not require the test to distinguish the very advanced student from the bilingual one. Whenever possible, difficult items were simplified. Of the four items which were found to be too difficult, one was re-written since it was based on a dialogue and it was less complicated to revise the question than to re-write and record a new dialogue.

In the listening comprehension sub tests, major revisions were made to the listening prompts as well as to the distractors. In all, 138 items had modifications made to the distractors. All the distractors with a positive biserial value, indicating that some high scoring students had chosen the option, were re-written so as not to be attractive to the higher scoring students. In addition, 16 new items had to be written for the large scale field test version.

Other reasons for revisions were formatting problems (3), typographical errors (3), punctuation errors (3). The distractors in 28 items had to be alphabetized in order to conform to the distractor criteria specification. Six items had more than one correct answer. Two items were found to be testing two objectives thereby making the item invalid. In some cases, distractors were revised because they attracted too many or too few responses (77 distractors and 63 distractors, respectively).

There were two items with perfect biserial values of +1.00. The worst item of the test had positive biserial values for the correct answer and each of the three distractors. Clearly, it was an item beyond the ability of the target audience of the test.

C O N C L U S I O N

The needs assessment and the formative evaluation phases of instructional design were instrumental in providing the data for informed decision-making in the production of a valid and reliable placement test in English for Quebec's college students. It is precisely during these two phases that those for whom the instrument is intended become involved. These two phases emphasize the value of the input of the test users and test takers by making them, in some way, the experts who can pinpoint the flaws and weaknesses of the instrument.

The questionnaires implicitly acknowledged the expertise of the teachers, students and the deans. These two phases of the development gave each of the groups a voice in the production of the final product. It recognized the fact that both students and teachers are the best source of information about student placement.

The models of test construction do not include the needs assessment and formative evaluation phases that are such a fundamental part of the instructional design process. It is precisely these aspects of instructional design that allowed the team to prepare the items and, therefore, the test for large scale field testing within constraining deadlines. In the end, these phases actually saved time and prevented costly mistakes.

Concluding Remarks on the Importance of Needs Assessment

The needs assessment provided the data to compare the opinions of the deans and teachers on several issues, most notably, the willingness to standardize the placement test and the cut-off scores. The similarity of their views, and the deans' greater acceptance of standardization, would not have emerged without the issue being addressed in the needs assessment. The data on this issue was the basis for the recommendation by the test development team at a Provincial Coordination meeting eighteen months later to standardize the test and the cut-off scores.

An excellent indicator of the value of needs assessment was demonstrated by the response rate of the deans, teachers and department coordinators to the questionnaires and the promptness of the deans' responses. Another good

indicator of its value was shown in the readiness with which teachers agreed to participate in the focus group discussions which were held in the evenings, after their teaching day. Their involvement provided necessary input for a precise and useful questionnaire.

Without a needs assessment some important and essential information about the current situation of placement testing in Quebec would not have been obtained. With it, the extent of problems such as back placement was able to be described precisely. Since sixty-nine percent of the colleges surveyed have no policy regarding this form of "cheating", the needs assessment report might sensitize teachers and deans to the desirability of adopting a policy concerning back placement.

Expectations of the teachers regarding the design of the test became clear. Without the needs assessment, other test designs could have been developed from the analysis of the objectives and standards. The test team might have opted for a direct test of students' ability to speak English, for example, if they had relied solely on the analysis of the curriculum. The needs assessment data informed the test developers of the deans and teachers opinions about who should be responsible for correcting the test. The data also pin-pointed some inconsistency between teacher expectation regarding the content of the test and the practical concern of correcting the test.

The needs assessment permitted the examination of practical issues in placement testing and possible solutions which would encourage deans to either continue using a test to place students or start using a placement test and not use other means to place students. In addition, it allowed the test development team to explore the feasibility of developing a computer adaptive test from the item bank.

Concluding Remarks on the Importance of Formative Evaluation

Pre-testing test items with a target group sample is a phase of development shared by test design and instructional design models. The major advantage of the instructional design model lies in the formative evaluation phase which not only assesses student performance on the items but has the added feature of obtaining data on questions related to test qualities.

The formative evaluation phase of instructional design gave the students, the eventual test-takers, a voice in the development process. By answering the formative evaluation questionnaire, they were able to express their opinions about key features of the test, features that, if uncorrected, could have threatened the outcome of the large scale field testing. Based on the data obtained, corrections were made to the instructions and the response time - both critical aspects of a test.

Despite the known benefits of these two stages, time constraints were argued as a major reason against undertaking the formative evaluation. The results of the questionnaire were so pertinent to the development that, in the subsequent field testing with 2000 students, the formative evaluation questionnaire was administered without hesitation. The second set of data revealed that the changes to the response time and the instructions were satisfactory.

Without both the needs assessment and the formative evaluation phases, it would have been impossible to produce a preliminary version of a placement test in such a short time, and with such limited resources. With both, it was possible to obtain an impressive level of participation from the teachers, department coordinators, deans and especially the students.

E P I L O G U E

During the second year of the placement test development, items were revised or discarded. Then followed the planning and organization of the large scale testing at 14 test sites across Quebec with 2000 students. Having learned from the previous field trials of the necessity to set the dates for testing early in the semester, a decision was made in early August 1996 to conduct testing in the first week of November, after midterm exams and many weeks before the end of semester frenzy.

As events unfolded, November 4 turned out to be the week that college students voted a general strike against proposed government changes in education funding. Some of the test sites were on strike for three weeks. This delayed the testing at some sites, the compilation of the data and the

preparation of the final version of the test scheduled for early spring. Consequently, colleges were only able to use the new test in May 1997.

The instructional design approach of consulting the milieu continued throughout the second year. There was a second formative evaluation of the test and an informal consultation of the students for the selection of the French narrator to read the French instructions.

Deans and department coordinators were again consulted regarding the final design of the test. The outcome of the last consultation and the number of "discriminating" items allowed the test development team to design two different forms of the test: a standard 86-item test and two 65-item pencil and paper adapted tests. Colleges are free to choose the test that best suits their needs. Each test places students at all the levels.

Students who do the adapted version of the test must first answer a self-evaluation question and, depending on their response, are given the appropriate test, Adapted 1 or Adapted 2 (the easier or the more difficult version, respectively). The self-assessment question was selected as a result of a concurrent questionnaire that was administered after the preliminary version of the test during each of the field tests. A number of self-assessment questions were pilot tested with the same sample at the same time as the test items were tested. One specific question had a high correlation with the scores on the preliminary version of the test.

In May 1997, the three forms of the *Test de classement d'anglais langue seconde au collégial* (TCALS II) were presented to the English second language teachers, the Ministry of Education and representatives of the colleges. A few colleges used the new placement test to place the students for the fall semester. Other colleges plan to use the test within a year. Monitoring of the test remains to be done including follow up on test use and cut-off scores and possible recommendations for future revisions.

REFERENCES

- Alderson, J. C., Clapham, C & Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.
- Alderson, J. C. & Krahnake, K. J. & Stansfield, C. W. (1987). *Reviews of English Language Proficiency Tests*. Washington, D.C.: Teachers of English to Speakers of Other Languages.
- Alderson, J. C. (1990). Learner-Centred Testing through Computers: Institutional Issues in Individual Assessment. In De Jong, John, H. A. L. & Stevenson, D. K. Individualizing the Assessment of Language Abilities. England: Multilingual Matters Ltd.
- Bachman, Lyle F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, L. F. & Palmer, A. S. (1984). Some comments of the terminology of language testing. In Rivera, Charlene. Communicative Competence Approaches to Language Proficiency Assessment: Research and Application. Avon, England: Multilingual Matters Ltd.
- Berdie, D. R. , Anderson J. F. & Niebuhr, M. A. (1986). *Questionnaires: Design and Use*. N.J.: The Scarecrow Press, Inc.
- Carroll, B. J. & Hall, P. J. (1985). *Make Your Own Language Tests: A Practical Guide to Writing Language Performance Tests*. NY: Pergamon Press Inc.
- Brindley, Geoff. 1986. *The Assessment of Second Language Proficiency: Issues and Approaches*. Australia: N.C.R.C.
- Carroll, John B. (1987). Psychometric Theory and Language Testing. In Grotjahn, R., Klein-Braley, C., & Stevenson, D. K. Taking Their Measure: The Validity and Validation of Language Tests. Bochum: Studienverlag Dr. N. Brockmeyer.
- Choi, Inn-Chull. (1992). *An Application of Item Response Theory to Language Testing*. New York: Peter Lang Publishing Inc.
- Cohen, Andrew D. (1994). *Assessing Language Ability in the Classroom*. Boston, Mass: Heinle & Heinle Publishers.
- Dandonoli, P. (1987). ACTFL's Current Research in Proficiency Testing. In Byrnes, H. & Canale, M. Defining and Developing Proficiency: Guidelines, Implementations, and Concepts. Illinois: National Textbook Company.
- Davies, Alan. (1990). *Principles of Language Testing*. Cambridge, Mass: Basil Blackwell, Inc.

- Dick, W. & Carey, L. (1990). *The Systematic Design of Instruction. Third Edition.* NY: Harper Collins Publisher.
- Formation générale des collèges pour le Québec du XXI e siècle.* (1993). Québec: Gouvernement du Québec. Ministère de l'Éducation. Direction générale de l'enseignement collégial.
- Fournier, P. (1995). *Rapport d'étape(anglais langue seconde): Comité aviseur pour le suivi de la mise en oeuvre de la formation générale renouvelée et des cours de mise à niveau.* Québec: Ministère de l'Éducation.
- Fournier, P. (1992). *Pour un test incontestable: Rapport de recherche sur les tests de classement en anglais (langue seconde) au collégial.* Québec: Direction générale de l'enseignement collégial. Ministère de l'Enseignement supérieur et de la science.
- Gradman, H. L. & Hanania, E. (1990). Discrete Focus vs. Global Tests: Performance on Selected Verb Structures. In De Jong, John, H. A. L. & Stevenson, D. K. Individualizing the Assessment of Language Abilities. Avon, England: Multilingual Matters Ltd.
- Harley, B., Cummins, J., Swain, M., & Allen, P. (1990). The nature of language proficiency. In Harley, B. & Allen, P. , Cummins, J. & Swain, M. The Development of Second Language Proficiency. New York: Cambridge University Press.
- Harris, David, P. (1969). *Testing English as a Second Language.* New York: McGraw-Hill Book Company.
- Harrison, Andrew. (1983). *A Language Testing Handbook.* In Roger H. Flavell Essential Language Teaching Series. London: The Macmillan Press Limited.
- Heaton, J. B. 1975. *Writing English Language Tests.* London: Longman Group Limited.
- Henning, Grant. (1991). Validating an Item Bank in a Computer-Assisted or Computer-Adaptive Test: Using Item Response Theory for the Process of Validating CATS. In Dunkel, Patricia. Computer-Assisted Language Learning and Testing: Research Issues and Practice. New York: Newbury House.
- Henning, Grant. (1987). *A Guide to Language Testing: Development, Evaluation, Research.* Cambridge, Mass.: Newbury House Publishers, Inc.
- Hughes, Arthur. (1989). *Testing for Language Teachers.* Cambridge: Cambridge University Press.

- Jenks, Frederick. (1987). Michigan Test of English Language Proficiency. In Alderson, J. C. & Krahnake, K. J. & Stansfield, C. W. Reviews of English Language Proficiency Tests. Washington, D.C.: Teachers of English to Speakers of Other Languages.
- Kaya-Carton, E., Carton, A. S., Dandonoli, P. (1991). Developing a Computer-Adaptive Test of French Reading Proficiency. In Dunkel, Patricia. Computer-Assisted Language Learning and Testing: Research Issues and Practice. New York: Newbury House.
- Kubiszyn, T. & Borich G. (1984). *Educational Testing and Measurement*. Illinois: Scott, Foresman and Company.
- Laurier, M. (1996). Un test de classement qui se modèle sur l'apprenant. Proceedings of the Third Annual Colloquium of the University of Ottawa's Centre for Research on Language Teaching and Learning, Canada, p. 48-51.
- Madsen, Harold S. (1991). Computer-Adaptive Testing of Listening and Reading Comprehension: The Brigham Young University Approach. In Dunkel, Patricia. Computer-Assisted Language Learning and Testing: Research Issues and Practice. New York: Newbury House.
- Oller, John W. Jr. (1978). How Important is Language Proficiency to IQ and Other Educational Tests? In Oller, John W. Jr. & Perkins, Kyle. Language in Education: testing the tests. MASS: Newbury House Publishers, Inc.
- Oller, John W. Jr. & Perkins, Kyle. (1978). Language Proficiency as a Source of Variance in Self-Reported Affective Variables. In Oller, John W. Jr. & Perkins, Kyle. Language in Education: testing the tests. MASS: Newbury House Publishers, Inc.
- Oxford, Rebecca. (1987). Comprehensive English Language Test. In Alderson, J. C. & Krahnake, K. J. & Stansfield, C. W. Reviews of English Language Proficiency Tests. Washington, D.C.: Teachers of English to Speakers of Other Languages.
- Pearo, C. & Rock, C. (1988). *Who's Using What in ESL in the CEGEP: A survey of materials used in the teaching of English second language in the CEGEPs*. Montreal: Collège André-Laurendeau and La Coordination provinciale d'anglais, langue seconde.
- Pike, Lewis, W. (1979). *An Evaluation of Alternative Item Formats for Testing English as a Foreign Language*. TOEFL Research Reports.
- Roid, Gale H. & Haladyna, Thomas M. (1982). *A Technology for Test-Item Writing*. Toronto: Academic Press, Inc.

- Rossett, Allison. (1987). *Training Needs Assessment*. Arwady, J. W. Techniques in Training and Performance Development Series. Englewood Cliffs, New Jersey: Educational Technology Publications.
- Rothwell, W.J. & Kazanas, H.C. (1992). *Mastering the Instructional Design Process*. San Francisco: Jossey-Bass, Inc.
- Smith, P. L. & Ragan, T. J. (1993). *Instructional Design*. New York: Macmillan Publishing Company.
- Underhill, Nic. (1992). *Testing Spoken Language: A handbook of oral testing techniques*. Swan, M. Cambridge Handbooks for Language Teachers. New York: Cambridge University Press.
- Wall, D., Clapham, C. & Alderson, J. C. (1994). Evaluating a Placement Test. Language Testing, 11(3) 321-344 Davies, A. & Upshur, J. (Eds.)
- Weir, Cyril. (1993). *Understanding and Developing Language Tests*. New York: Prentice Hall, Inc.
- Westaway, G., Alderson, J.C., & Clapham, C.M. (1990). Directions in Testing for Specific Purposes. In De Jong, John, H. A. L. & Stevenson, D. K. Individualizing the Assessment of Language Abilities. Avon, England: Multilingual Matters Ltd.

APPENDIX A:
NEEDS ASSESSMENT QUESTIONNAIRES

October 16, 1995

Dear Coordinator,

As you can see, we are well into the first phase of the placement test project. We are counting on all our colleagues to provide us with the information we need to proceed with our work. Enclosed in this package is:

- copies of Teacher Questionnaire and cover letter;
- one Coordinator Questionnaire;
- one Level Profiles worksheet;
- a few pre-addressed mailing labels.

We are enlisting your help in distributing the enclosed questionnaires to all the ESL teachers in your department. If necessary, could you make additional photocopies. Teachers should answer on their own; at this time we are not looking for a consensus from the department members.

Please encourage teachers to return their completed questionnaires to Cégep André-Laurendeau by November 3, if not sooner. With your support we hope to obtain a high rate of response.

The Department Coordinator questionnaire deals primarily with information that you have or can get. Some of the questions are repeated on the questionnaire that has been sent to your Dean. However, we ask you to provide the information based on your own knowledge and resources.

In addition, we ask you to please find volunteers in your department who would be willing to fill out the Level Profiles worksheet. These responses will serve to construct a more accurate picture of ESL students across Quebec.

Thank you for your collaboration in this very important phase of the placement test project. If you have any questions, do not hesitate to call us.

Lydia Froio (514) 254-7131 ext 4553
Chuck Pearo (514) 364-3320 ext 563

October 16, 1995

Dear ESL colleague,

The *Direction générale de l'enseignement collégial* has accepted a two-year project to develop and design a new placement test in English as a second language. Lydia Froio and Chuck Pearo will be working on the test along with Michel Laurier, a specialist in testing and evaluation, from the University of Montreal. In order to address your placement testing concerns we have developed three questionnaires: one for ESL teachers, one for department coordinators, and another for administrators.

We realize how busy everyone is. However, it should not take you more than fifteen to thirty minutes to fill out the questionnaire. Your insights and expertise are vital to the success of this project. We appreciate the time and thought you devote to answering the questions, and assure you that no questionnaire will be eliminated.

The survey is divided into two main parts. The first deals with your opinions on a new placement test, and the second, evaluates the test you use at present. Here are a few suggestions to keep in mind when answering questions:

- answer all questions;
- use the many areas designated for comments. If you require more space, please add sheets;
- answer questions as they are worded and not in relation to administrative or practical considerations;
- answer questions based on Block A courses (*cours de formation générale commune*), and not Block B (*formation générale propre*);
- help us by writing legibly.

Please mail the completed forms to Chuck Pearo, at Cégep André-Laurendeau, by November 3. We have included several return address stickers for your convenience. A report will be sent to your college as soon as we have compiled the results of the survey.

Have a productive and successful semester. Thank you for your participation!

Lydia Froio (514) 254-7131 est. 4553
Chuck Pearo (514) 364-3320 ext. 563
Michel Laurier

Les données quantitatives du questionnaire destiné à la Direction des études

Les données suivantes proviennent des 48 collèges publiques et privés qui ont répondu à ce questionnaire. Les commentaires et les données qualitatives seront disponibles dans le rapport qui sera publié en juin 1996. Pour chacune des choix possibles, on trouvera le nombre de réponses et le pourcentage entre parenthèses.

Le test de classement en anglais langue seconde

1. Quel service a la responsabilité d'administrer le test de classement en anglais langue seconde?

- | | |
|---|---|
| <input type="checkbox"/> ₁ Le registrariat
33 (68,8%) | <input type="checkbox"/> ₂ Le département
7 (14,6%) |
| <input type="checkbox"/> ₃ Un autre service
5 (10,4%) | <input type="checkbox"/> ₄ Registrariat et département
3 (6,3%) |

2. Quel service corrige le test?

- | | | |
|--|--|--|
| <input type="checkbox"/> ₁ Le registrariat
24 (50%) | <input type="checkbox"/> ₂ Le département
10 (20,8%) | |
| <input type="checkbox"/> ₃ Un autre service
12 (25%) | <input type="checkbox"/> ₄ Registrariat + département
1 (2,1%) | <input type="checkbox"/> ₃₅ Département + autre
1 (2,1%) |

3. Comment le test de classement est-il corrigé?

- | | | |
|--|---|---|
| <input type="checkbox"/> ₁ Manuellement
26 (54,2%) | <input type="checkbox"/> ₂ Par lecteur optique
15 (31,3%) | <input type="checkbox"/> ₃ Par ordinateur
7 (14,6%) |
|--|---|---|

4. Combien d'élèves par année passent le test?

Moyenne	1141
Mediane	1000
Écart type	782,40

5. Combien d'élèves passent le test lors d'une même séance de groupe?

Entre _____ (minimum) et _____ (maximum)

Minimum		Maximum	
Moyenne	112,75	Moyenne	320,21
Mediane	50	Mediane	170
Écart type	145,53	Écart type	398,61

6. Combien de locaux différents sont utilisés lors d'une même séance de groupe?

Entre _____ (minimum) et _____ (maximum)

Minimum		Maximum	
Moyenne	2,83	Moyenne	7,83
Mediane	1	Mediane	5
Ecart type	3,82	Ecart type	12,29

7. Combien de séances de groupe y a-t-il dans une année? _____

Moyenne	9,43
Mediane	5
Ecart type	10,39

8. Combien d'élèves passent le test en dehors des séances de groupe (administration individuelle)? _____

Moyenne	32,6
Mediane	12
Ecart type	47,34

9. À quel moment de leurs études collégiales la plupart des élèves passent-ils le test?

₁ Avant le 1er semestre
41 (85,4%)

₂ Pendant le 1er semestre
2 (4,2%)

₃ Pendant le 2^e semestre
1 (2,1%)

₄ Autre _____
1 (2,1%)

₅ Avant et pendant le 1er semestre
3 (6,3%)

10. Dans votre collège, est-il possible pour les élèves mal classés de changer de niveau?

₁ Oui
45 (93,8%)

₀ Non
3 (6,3%)

11. Quel délai allouez-vous pour les changements de niveaux?

₁ Une semaine
23 (47,9%)

₂ Deux semaines
16 (33,3%)

₃ Autre (précisez)
7 (14,6%)

₄ N.A.
2 (4,2%)

12. Des mesures sont-elles prévues, dans votre établissement, pour les cas de tricherie au test de classement?

₁ Oui
10 (21,3%)

₀ Non
37 (78,7%)

13. Est-ce que vous considérez que le test est facile à administrer?

<input type="checkbox"/> ₁ Oui 40 (83,3%)	<input type="checkbox"/> ₀ Non 6 (12,5%)	<input type="checkbox"/> ₂ Incertain 2 (4,2%)
---	--	---

14. Pensez-vous qu'il est nécessaire de classer les élèves en anglais langue seconde par niveau de compétence?

<input type="checkbox"/> ₁ Oui 43 (89,6%)	<input type="checkbox"/> ₀ Non 2 (4,2%)	<input type="checkbox"/> ₂ Incertain 3 (6,3%)
---	---	---

Le nouveau test de classement

15. Souhaiteriez-vous voir:

a) un nouveau test de classement?

<input type="checkbox"/> ₁ Oui 34 (70,8%)	<input type="checkbox"/> ₀ Non 6 (12,5%)	<input type="checkbox"/> ₂ Incertain 8 (16,7%)
---	--	--

b) le même test de classement utilisé dans tous les cégeps?

<input type="checkbox"/> ₁ Oui 40 (85,1%)	<input type="checkbox"/> ₀ Non 3 (6,4%)	<input type="checkbox"/> ₂ Incertain 4 (8,5%)
---	---	---

c) les mêmes scores de césure (seuils de classement) utilisés dans tous les cégeps?

<input type="checkbox"/> ₁ Oui 37 (78,7%)	<input type="checkbox"/> ₀ Non 5 (10,6%)	<input type="checkbox"/> ₂ Incertain 5 (10,6%)
---	--	--

d) un test de classement administré à l'aide de l'ordinateur?

<input type="checkbox"/> ₁ Oui 21 (44,7%)	<input type="checkbox"/> ₀ Non 13 (27,7%)	<input type="checkbox"/> ₂ Incertain 13 (27,7%)
---	---	---

16. Préférez-vous un test:

<input type="checkbox"/> ₁ papier/crayon (conventionnel)? 24 (52,2%)	<input type="checkbox"/> ₂ informatisé? 18 (39,1%)
--	--

<input type="checkbox"/> ₃ Les deux 2 (4,3%)	<input type="checkbox"/> ₄ Ni l'un ni l'autre 2 (4,3%)
--	--

17. Avez-vous des réticences à l'égard d'un test de classement administré à l'aide de l'ordinateur?

<input type="checkbox"/> ₁ Oui 15 (31,9%)	<input type="checkbox"/> ₀ Non 22 (46,8%)	<input type="checkbox"/> ₂ Incertain 10 (21,3%)
---	---	---

18. Où les élèves devraient-ils passer le test de classement?

- ₁ Au cégep ₂ À l'école secondaire
 37 (80,4%) 9 (19,6%)

19. Combien de temps devrait durer le test? _____ minutes

Moyenne	50,7
Mediane	45
Écart type	19,92

20. Quel service devrait avoir la responsabilité:

a) d'administrer le test?

- ₁ Le registrariat ₂ Le département
 29 (63%) 7 (15,2%)
- ₃ Un autre service ₄ Les deux
 5 (10,9%) 3 (6,5%)
- ₅ L'un ou l'autre ₆ Département + autre service
 1 (2,2%) 1 (2,2%)

b) de corriger le test?

- ₁ Le registrariat ₂ Le département ₃ Un autre service
 19 (41,3%) 12 (26,1%) 11 (23,9%)
- ₄ Les deux ₅ Regist. + autre service ₆ L'un ou l'autre
 1 (2,2%) 1 (2,2%) 1 (2,2%)
- ₇ Départ. + autre service
 1 (2,2%)

21. Dans votre collège, combien y a-t-il de laboratoires d'ordinateurs à la disposition des élèves? _____

Moyenne	12,98
Mediane	6,5
Écart type	28,23

22. Dans votre collège, combien y a-t-il de postes d'ordinateurs à la disposition des élèves?

a) Macintosh

Moyenne	27,28
Mediane	15
Écart type	45,28

b) IBM / compatible

Moyenne	146,55
Mediane	115
Écart type	99,03

23. Combien d'entre eux pourraient être utilisés par les élèves pour un test de classement qui durerait trente minutes? _____

Moyenne	112,63
Mediane	78
Écart type	84,05

24. Combien d'entre eux sont reliés à un réseau? _____

Moyenne	129,57
Mediane	86
Écart type	105,29

Quantitative Data from 238 Teacher Questionnaires

238 teachers representing 49 public and private colleges responded to this questionnaire. The first number indicates the number of responses with the percentage in parentheses.

A New Test

1. Do we need a new placement test?

- | | | |
|---|--|---|
| <input type="checkbox"/> ₁ Yes | <input type="checkbox"/> ₀ No | <input type="checkbox"/> ₂ Undecided |
| 182 (77.1%) | 5 (2.1%) | 49 (20.8%) |

2. Approximately how long should a placement test take? _____ minutes

Minutes	
Mean	54.6
Median	50
Mode	60
SD	23.67

3. Should all colleges use the same placement test?

- | | | |
|---|--|---|
| <input type="checkbox"/> ₁ Yes | <input type="checkbox"/> ₀ No | <input type="checkbox"/> ₂ Undecided |
| 192 (81%) | 21 (8.9%) | 24 (10.1%) |

4. Should cut-off scores be the same for all colleges?

- | | | |
|--|--|---|
| <input type="checkbox"/> ₁ Yes | <input type="checkbox"/> ₀ No | <input type="checkbox"/> ₂ Undecided |
| 154 (65.5%) | 47 (20%) | 33 (14%) |
| <input type="checkbox"/> ₂ It depends | | |
| 1 (.4%) | | |

5. Who should be responsible for administering a placement test?

- | | |
|--|--|
| <input type="checkbox"/> ₁ Registrar's Office | <input type="checkbox"/> ₂ Language Department |
| 111 (46.8%) | 91 (38.4%) |
| <input type="checkbox"/> ₃ Other | <input type="checkbox"/> ₄ Registrar + Language Dept. |
| 6 (2.5%) | 22 (9.3%) |
| <input type="checkbox"/> ₅ Registrar + Other | <input type="checkbox"/> ₈ It depends |
| 6 (2.5%) | 1 (.4%) |

6. Who should correct a placement test?
- | | |
|--|--|
| <input type="checkbox"/> ₁ Registrar's Office | <input type="checkbox"/> ₂ Language Department |
| 105 (44.7%) | 90 (38.3%) |
| <input type="checkbox"/> ₃ Other | <input type="checkbox"/> ₄ Registrar + Language Dept. |
| 15 (6.4%) | 12 (5.1%) |
| <input type="checkbox"/> ₅ Registrar + Other | <input type="checkbox"/> ₈ It depends |
| 12 (5.1%) | 1 (.4%) |
7. Would you be interested in using computers for placement test-taking?
- | | | |
|---|--|---|
| <input type="checkbox"/> ₁ Yes | <input type="checkbox"/> ₀ No | <input type="checkbox"/> ₂ Undecided |
| 133 (57.6%) | 38 (16.5%) | 60 (26%) |
8. What type of test would you prefer?
- | | |
|--|---|
| <input type="checkbox"/> ₁ A pen and paper test | <input type="checkbox"/> ₂ A computerized test |
| 94 (41.8%) | 109 (48.4%) |
| <input type="checkbox"/> ₃ Both types | |
| 22 (9.8%) | |
9. What should a new placement test include?

Circle the number which corresponds to the level of importance you place on each item listed.

	Very important	Important	Somewhat important	Not important at all
	1	2	3	4
<u>Language Skills</u>				
a) Reading component	156	53	20	4
b) Composition component	133	62	30	3
c) Listening component	168	48	15	3
d) Speaking component	124	58	33	11
e) Grammar items	124	75	22	3
<u>Testing-tasks</u>				
f) Audio segments	146	53	15	5
g) Video segments	35	54	59	58
h) Graphic-related questions	29	56	67	43

10. Are you interested in writing items for the new placement test?
- | | | |
|---|--|---|
| <input type="checkbox"/> ₁ Yes | <input type="checkbox"/> ₀ No | <input type="checkbox"/> ₂ Maybe |
| 55 (24.3%) | 165 (73%) | 5 (2.2%) |
11. Would you be willing to ask some students to volunteer for test-taking?
- | | |
|---|--|
| <input type="checkbox"/> ₁ Yes | <input type="checkbox"/> ₀ No |
| 203 (85.7%) | 34 (14.3%) |
12. Would you be willing to read and comment on a working draft of the test?
- | | | |
|---|--|---|
| <input type="checkbox"/> ₁ Yes | <input type="checkbox"/> ₀ No | <input type="checkbox"/> ₂ Maybe |
| 187 (79.9%) | 46 (19.7%) | 1 (0.4%) |

The Present Test

13. Which placement test are you using?
- | | | |
|---|--|--|
| <input type="checkbox"/> ₁ TCALS (50 questions) | <input type="checkbox"/> ₄ Michigan | <input type="checkbox"/> ₆ In-house |
| 12 (12.8%) | 24 (11.6%) | 28 (13.5%) |
| <input type="checkbox"/> ₂ TCALS (100 questions) | <input type="checkbox"/> ₅ CELT | <input type="checkbox"/> ₇ Other (specify)_____ |
| 105 (50.7%) | 15 (7.2%) | 6 (2.9%) |
| <input type="checkbox"/> ₃ TCALS (150 questions) | | <input type="checkbox"/> ₈ None |
| 13 (6.3%) | | 4 (1.9%) |
14. Are you satisfied with the way your placement test classifies students?
- | | |
|--|---|
| <input type="checkbox"/> ₁ Very satisfied | <input type="checkbox"/> ₃ Dissatisfied |
| 9 (4%) | 73 (32.4%) |
| <input type="checkbox"/> ₂ Satisfied | <input type="checkbox"/> ₄ Very dissatisfied |
| 99 (44%) | 11 (4.9%) |
| <input type="checkbox"/> ₈ Can't answer | |
| 33 (14.7%) | |

15. In your experience, what are the reasons for misclassification?

Check one or more reasons relevant to your situation.				Rank
a) <input type="checkbox"/>	The test does not measure the four language skill areas.	168	(76.0%)	2
b) <input type="checkbox"/>	The test content does not correspond to the course content.	88	(39.8%)	3
c) <input type="checkbox"/>	The test is too short.	4	(1.8%)	9
d) <input type="checkbox"/>	The test has not been fully validated.	14	(6.3%)	7
e) <input type="checkbox"/>	The test is scored by different people.	5	(2.3%)	8
f) <input type="checkbox"/>	Students taking the test come from diverse backgrounds.	45	(20.4%)	4
g) <input type="checkbox"/>	Other means are used to place students. (e.g. high school marks)	15	(6.8%)	6
h) <input type="checkbox"/>	Clerical errors occur.	43	(19.5%)	5
i) <input type="checkbox"/>	Students cheat in order to be placed at a lower level.	172	(77.8%)	1
j) <input type="checkbox"/>	Students cheat in order to be placed at a higher level.	15	(6.8%)	6

16. At the beginning of this semester, approximately 8_a (#) students out of 88_b (#) students in my groups were misclassified. Of these, 4_c (#) were changed.

These figures represent the mean. In other words, 10.04% of the students are misclassified. Of the students mis-classified, 55.43% change level.

17. Is the placement test you use appropriate to today's students in terms of....?
(NA means the item is not applicable to the test you use.)

a) Reading component	<input type="checkbox"/> ₁ Yes	<input type="checkbox"/> ₀ No	<input type="checkbox"/> ₂ Don't know	<input type="checkbox"/> ₈ NA
	54	72	44	32
b) Composition component	<input type="checkbox"/> ₁ Yes	<input type="checkbox"/> ₀ No	<input type="checkbox"/> ₂ Don't know	<input type="checkbox"/> ₈ NA
	20	82	34	70
c) Listening component	<input type="checkbox"/> ₁ Yes	<input type="checkbox"/> ₀ No	<input type="checkbox"/> ₂ Don't know	<input type="checkbox"/> ₈ NA
	100	43	41	22
d) Speaking component	<input type="checkbox"/> ₁ Yes	<input type="checkbox"/> ₀ No	<input type="checkbox"/> ₂ Don't know	<input type="checkbox"/> ₈ NA
	5	85	34	80
e) Grammar items	<input type="checkbox"/> ₁ Yes	<input type="checkbox"/> ₀ No	<input type="checkbox"/> ₂ Don't know	<input type="checkbox"/> ₈ NA
	117	33	44	7
f) Written test	<input type="checkbox"/> ₁ Yes	<input type="checkbox"/> ₀ No	<input type="checkbox"/> ₂ Don't know	<input type="checkbox"/> ₈ NA
	23	62	49	64
g) Questions with graphics	<input type="checkbox"/> ₁ Yes	<input type="checkbox"/> ₀ No	<input type="checkbox"/> ₂ Don't know	<input type="checkbox"/> ₈ NA
	6	49	50	95

h) Quality of the audio	<input type="checkbox"/> ₁ Yes	<input type="checkbox"/> ₀ No	<input type="checkbox"/> ₂ Don't know	<input type="checkbox"/> ₈ NA
	74	44	55	29
i) Layout and design	<input type="checkbox"/> ₁ Yes	<input type="checkbox"/> ₀ No	<input type="checkbox"/> ₂ Don't know	<input type="checkbox"/> ₈ NA
	52	52	70	22

18. Have you read the placement test you use in your college?

<input type="checkbox"/> ₁ Yes	<input type="checkbox"/> ₀ No
167 (73.9 %)	58 (25.7%)

19. Have you ever taken the placement test you use in your college?

<input type="checkbox"/> ₁ Yes	<input type="checkbox"/> ₀ No
89 (39.4%)	135 (59.7%)

20. How familiar are you with the content of the placement test?

<input type="checkbox"/> ₁ Very familiar	<input type="checkbox"/> ₃ Somewhat familiar
60 (26.3%)	71 (31.1%)
<input type="checkbox"/> ₂ Familiar	<input type="checkbox"/> ₄ Not familiar at all
58 (25.4%)	39 (17.1%)

AND finally

A mini portrait of the ESL teachers in Quebec who responded to the Teacher Survey Questionnaire.

In all, I have taught ESL for **13.3 years**; I've been teaching at the college level for **8.3 years**. (50% of survey respondents have 1 to 3 years experience at the college level.)

Quantitative Data from 45 Department Coordinator Questionnaires

Forty-five department coordinators from public and private colleges responded to this questionnaire.

The first number indicates the number of responses with the percentage in parentheses.

1. How many ESL teachers are there in your department? **7.73 teachers (SD 4.64)**

2. Which placement test are you using?

- | | | |
|--|--|--|
| <input type="checkbox"/> 1 TCALS (50 questions)
2 (4.4%) | <input type="checkbox"/> 4 Michigan
5 (11.1%) | <input type="checkbox"/> 6 In-house
4 (8.9%) |
| <input type="checkbox"/> 2 TCALS (100 questions)
22 (48.9%) | <input type="checkbox"/> 5 CELT
3 (6.7%) | <input type="checkbox"/> 7 Other (specify) _____
3 (6.7%) |
| <input type="checkbox"/> 3 TCALS (150 questions)
2 (4.4%) | | <input type="checkbox"/> 8 None
4 (8.9%) |

3. Do you use

- 1** the complete test
36 (90%)
- 2** only certain parts of the test?
4 (10%)

4. Have you modified the existing test in any way?

- 1** Yes
8 (20%)
- 0** No
32 (80%)

5. Do you calculate the scores as recommended in the test manual?

- 1** Yes
22 (59.5%)
- 0** No
9 (24.3%)
- 2** Don't know
4 (10.8%)
- 3** NA
2 (5.4%)

6. What are the cut-off scores for each course listed? CF Annex

7. What was the failure rate for each level in 1994-1995?

	Average failure rate
Remedial	27.46%
104 (A0)	26.76%
101 (A1)	21.86%
102 (A2)	20.67%
103 (A3)	22.79%

8. When do the majority of your students do the placement test?

₁ Before the 1st semester ₂ During the 1st semester
30 (73.2%) **4 (9.8%)**

₃ During the 2nd semester ₄ Other _____
4 (9.8%) **3 (7.3%)**

9. **1214,9_a** (#) students are tested annually. Of these, **421,1_b** (#) students are tested at the same time in **9,3_c** (#) different rooms.

10. How many students are registered in an English course (*Bloc A: formation générale commune*) this semester at each of the levels?

Some colleges do not offer or have not yet given the remedial course or the 103.

	Average number of students	
Remedial	58.18	(12.8%)
104 (A0)	274.03	(40.5%)
101 (A1)	171.41	(25.1%)
102 (A2)	116.46	(15.7%)
103 (A3)	47.06	(6.1%)

f) Of these, approximately how many were misclassified? **37.4 (5.63%)**

11. Which of the following reasons for misclassification occurs more frequently in your college?

Rank the following reasons for misclassification in terms of their relative frequency during this semester. Assign each reason a rank from 1 (the most frequent) to 4 (the least frequent). Use a different number for each of the reasons.

	Mean Rank
a) The test does not measure what it should.	1 <input type="checkbox"/>
b) Clerical errors	4 <input type="checkbox"/>
c) Cheating	2 <input type="checkbox"/>
d) Diversity of the population	3 <input type="checkbox"/>

12. Does your department have a policy with regard to cheating on the placement test?

- | | |
|---|--|
| <input type="checkbox"/> Yes ₁
13 (31%) | <input type="checkbox"/> No ₀
29 (69%) |
|---|--|

13. Who is responsible for administering the placement test?

- | | |
|--|--|
| <input type="checkbox"/> ₁ Registrar's Office
28 (68,3%) | <input type="checkbox"/> ₂ Language Department
6 (14,6%) |
| <input type="checkbox"/> ₃ Other
5 (12,2%) | <input type="checkbox"/> ₄ Registrar + Department
2 (4,9%) |

14. Who corrects the placement test?

- | | |
|--|--|
| <input type="checkbox"/> ₁ Registrar's Office
28 (65,9%) | <input type="checkbox"/> ₂ Language Department
6 (7,3%) |
| <input type="checkbox"/> ₃ Other
10 (24,4%) | <input type="checkbox"/> ₄ Registrar + Department
1 (2,4%) |

15. How are most of the placement test copies corrected?

- | | | |
|--|--|---|
| <input type="checkbox"/> ₁ Manually
20 (48,8%) | <input type="checkbox"/> ₂ Optical Scanner
12 (29,3%) | <input type="checkbox"/> ₃ Computer
7 (17,1%) |
| <input type="checkbox"/> ₄ Other
1 (2,4%) | <input type="checkbox"/> ₅ Scanner + Computer
1 (2,4%) | |

16. How many computer labs are there in your college? CF Annex

17. How many student computer stations are there in your college?

	Mean	SD	Total
Macintosh	313.6	35.6	1011
IBM / compatible	139.1	108.1	5006

APPENDIX B:
LANGUAGE SKILL MATRICES

Listening Matrix

Performance Criteria		# min	Type of Text		Adapted		Authentic				
Level	Context		Simple	Average	Every-day	Social	Cult.	Lit.	Acad.		
001	<ul style="list-style-type: none"> • interprétation correcte du message • reconnaissance du sens général d'un texte simple 		X								
104	<ul style="list-style-type: none"> • reconnaissance du sens général et des idées essentielles 		X	X							
101	<ul style="list-style-type: none"> • reconnaissance explicite du sens général et des idées essentielles • reconnaissance d'une suite logique des éléments 	2		X							
102	<ul style="list-style-type: none"> • reformulation des idées essentielles d'un message audio d'une durée de 5 minutes 	5							X	X	X
103	<ul style="list-style-type: none"> • analyse critique et étude comparative • reformulation et justification du texte (thèmes, symboles, contenu) et des détails pertinents 								X	X	X

Reading Matrix

Performance Criteria		# wds	Type of Text		Adapted		Authentic			
Level			Context	Simple	Average	Every-day	Social	Cult.	Lit.	Acad.
001	<ul style="list-style-type: none"> • interprétation correcte du message • reconnaissance du sens général d'un texte simple 		<ul style="list-style-type: none"> • sujets concrets, simples, pragmatiques • la vie courante • message à l'impératif 	X						
104	<ul style="list-style-type: none"> • reconnaissance du sens général et des idées principales 	500	<ul style="list-style-type: none"> • intérêt général • la vie courante • vocab. d'usage courant • phrases simples • récit linéaire 	X	X					
101	<ul style="list-style-type: none"> • reconnaissance explicite du sens général et des idées principales • reconnaissance d'une suite logique des éléments 	env 750	<ul style="list-style-type: none"> • contexte social habituel • intérêt général • articles/extraits • vocab. d'usage courant • complexité moyenne 		X					
102	<ul style="list-style-type: none"> • reformulation des idées principales et secondaires 	min 750	<ul style="list-style-type: none"> • socioculturel ou littéraire • articles/extraits • difficulté moyenne 				X	X	X	X
103	<ul style="list-style-type: none"> • analyse critique et étude comparative • reformulation et justification du texte (thèmes, symboles, contenu) et des détails pertinents 		<ul style="list-style-type: none"> • socioculturel/littéraire • version originale • production littéraire ou culturelle • romans, nouvelles, essais, théâtre 					X	X	X

Speaking Matrix

Performance Criteria		# min.	Context	Accuracy
Level				
001	<ul style="list-style-type: none"> répondre à des questions simples d'usage courant formuler des questions simples décrire des événements, des faits, des personnes et des objets participer à un dialogue encadré reformulation intelligible du sens général questions, dialogues, échanges d'idées 		<ul style="list-style-type: none"> situations simples de la vie courante sujets concrets, simples, et pragmatiques descriptions sommaires, exactes 	<ul style="list-style-type: none"> application satisfaisante du code grammatical de façon à ne pas nuire à la compréhension pron., intonation et débit convenables
104	<ul style="list-style-type: none"> reformulation intelligible du sens général questions, dialogues, échanges d'idées 	< 5 min.	<ul style="list-style-type: none"> sujets familiers 	<ul style="list-style-type: none"> application satisfaisante du code grammatical de façon à ne pas nuire à la compréhension
101	<ul style="list-style-type: none"> description intelligible, cohérent d'un événement sous forme d'un exposé improvisé dialogue improvisé question intelligibles, pertinentes 3 temps de verbes minimum 3 modals (de base) minimum variété de "question words" 	15 phrases	<ul style="list-style-type: none"> mises en situation qui reflètent un contexte social habituel 	<ul style="list-style-type: none"> formulation de questions grammaticalement correctes dans la majorité des cas
102	<ul style="list-style-type: none"> exposé clair et cohérent comprenant des propositions adverbiales et relatives verbes à la voix passive 	10 quest.	<ul style="list-style-type: none"> socioculturel ou littéraire 	<ul style="list-style-type: none"> respect du code grammatical
103	<ul style="list-style-type: none"> présentation claire, cohérente et structurée comprenant des éléments de comparaison pertinents et justifiés 	5 à 8 min.	<ul style="list-style-type: none"> à partir de productions culturelles et littéraires en version originale 	<ul style="list-style-type: none"> respect des conventions de l'anglais standard
		7 à 10 min.		

Writing Matrix

Performance Criteria		# wds	Context	Accuracy
Level				
001	<ul style="list-style-type: none"> forme interrogative phrases courtes et complètes 	75	<ul style="list-style-type: none"> situations simples de la vie courante 	<ul style="list-style-type: none"> application satisfaisante du code grammatical de façon à ne pas nuire à la compréhension
104	<ul style="list-style-type: none"> texte clair et cohérent phrases courtes et complètes 2 temps de verbes punctuation de base 	150	<ul style="list-style-type: none"> sujets familiaux, concrets et pragmatiques 	<ul style="list-style-type: none"> application satisfaisante du code grammatical dans une forme compréhensible
101	<ul style="list-style-type: none"> texte clair et cohérent 3 temps de verbes 3 phrases composées punctuation de base 	200	<ul style="list-style-type: none"> sujets familiaux contexte social habituel 	<ul style="list-style-type: none"> application satisfaisante du code grammatical dans une forme qui ne nuit pas à la compréhension
102	<ul style="list-style-type: none"> texte clair, structuré et cohérent 3 idées distinctes, liées logiquement entre elles phrases composées/complexes 3 temps de verbes minimum voix passive dans au moins 2 phrases 	400	<ul style="list-style-type: none"> socioculturel ou littéraire 	<ul style="list-style-type: none"> aucune erreur nuisant à la compréhension 1 faute aux 20 mots
103	<ul style="list-style-type: none"> analyse structurée, cohérente, claire et révisée phrases simples, complexes, composées et complexes-composées 	600	<ul style="list-style-type: none"> à partir de productions culturelles et littéraires en version originale 	<ul style="list-style-type: none"> 1 faute aux 30 mots

APPENDIX C:
TEST SITE DOCUMENTS

May 3, 1996

Dear Guardian Angel,

We appreciate your help in this experimental stage of the placement test development. This will help us to collect data on preliminary items in order to devise a final version.

Included in this package are the following materials:

- _____ Student Test Books
- _____ Computerized Answer Sheets
- _____ "Une note à l'élève"
- A cassette with the recorded version of the test (Compréhension auditive)
- A return address sticker

What you will need:

- A tape recorder
- Some sharpened pencils

Here are some directions on how to proceed with the test:

- For the recorded version, do not stop or rewind the tape until the end (*Fin de la partie compréhension auditive*)
- All tests should be done individually and without reference material.
- Please allow the students a reasonable time to finish.
- When they finish, retrieve all materials

Give us the following information for each student:

- name

- score on the TCALS 100
- starting and finishing time for each student

Mail all material back to us (use the enclosed sticker)

If you notice anything unusual about students' reactions please write it down for us.

Thank you for accepting and going that extra mile at a very hectic period in the term.

Sincerely,

Lydia Froio

Charles Pearo

Test de classement anglais langue seconde

ANNEX

Please verify the form of the test you are giving before announcing these corrections to the students. Hopefully there aren't any others.

CORRECTIONS TO FORM 1

PLEASE ANNOUNCE THESE CORRECTIONS BEFORE STARTING OF THE TEST.

13. The name used on the tape is "Keith" not "John".

17. What does the man think?
 - c. That wine costs more in liquor stores.

CORRECTIONS TO FORM 2

PLEASE ANNOUNCE THESE CORRECTIONS BEFORE STARTING OF THE TEST.

13. The name used on the tape is "Keith" not "John".

CORRECTIONS TO FORM 3

PLEASE ANNOUNCE THESE CORRECTIONS BEFORE STARTING OF THE TEST.

13. The name used on the tape is "Keith" not "John".

NOTE À L'ÉTUDIANT(E)

Ce test est la version expérimentale d'un outil de classement que nous sommes en train d'élaborer pour l'ensemble des cégeps de la province. Vos résultats ne seront pas utilisés pour modifier votre classement actuel ou pour établir votre note finale. Par contre, vos réponses sont essentielles pour mettre au point un test de classement juste et équitable. Toutes les données recueillies dans le cadre de la présente expérimentation, seront traitées de façon tout à fait confidentielle. Nous vous demandons de répondre à ce test le plus sérieusement possible sans toutefois passer trop de temps sur chaque item. Si vous acceptez de faire l'épreuve, nous vous prions de signer la présente feuille afin de confirmer que vous avez pris connaissance du but de cette expérimentation. Nous vous remercions beaucoup de votre collaboration.

Signé à _____ le _____

Signature

APPENDIX D:
FORMATIVE EVALUATION
QUESTIONNAIRES AND QUANTITATIVE AND
QUALITATIVE DATA

QUESTIONNAIRE D'ÉVALUATION
DU TEST DE CLASSEMENT
ANGLAIS LANGUE SECONDE

D'abord nous vous remercions d'avoir accepté de compléter la version expérimentale du test de classement anglais langue seconde. Nous apprécierons vos commentaires sur certains aspects du test de classement que vous venez de terminer afin de pouvoir en corriger les lacunes.

Lisez les questions et indiquez votre accord à chaque énoncé en utilisant l'échelle suivante:

Tout à fait en désaccord	Plutôt en désaccord	Plus ou moins d'accord	Plutôt en accord	Tout à fait en accord
1	2	3	4	5

Noircissez sur votre feuille de réponses le numéro qui correspond à votre choix.

SVP

Portez bien attention aux numéros de la question à laquelle vous répondez. Les questions commencent à numéro 151.

151.	Les instructions écrites dans la partie <i>Compréhension auditive</i> étaient claires.	1	2	3	4	5	
		5.02	2.54	9.32	22.03	61.02	
152.	Les exemples dans la partie <i>Compréhension auditive</i> étaient utiles.	1	2	3	4	5	
		5.93	5.08	16.10	34.75	38.14	
153.	Je n'avais pas besoin d'exemple dans la section Mini-exposé de la partie <i>Compréhension auditive</i> .	1	2	3	4	5	
		15.25	24.58	32.20	16.1	11.86	
154.	La qualité du son de la bande sonore était bonne.	1	2	3	4	5	
		7.63	6.78	6.78	39.83	38.98	
155.	La voix des femmes de la bande sonore était claire.	1	2	3	4	5	
		6.78	11.86	12.71	39.83	28.81	

156.	La voix des hommes de la bande sonore était claire.	1	2	3	4	5					
		5.08	10.17	12.71	42.37	29.66					
157.	Les personnes parlaient à une vitesse convenable.	1	2	3	4	5					
		7.63	20.34	23.73	26.27	22.03					
158.	J'avais assez de temps pour lire les choix de réponses (20 secondes) dans la partie <i>Compréhension auditive</i> .	1	2	3	4	5					
		5.93	16.95	24.58	27.97	24.58					
159.	J'avais assez de temps pour inscrire la réponse (5 secondes) dans la partie <i>Compréhension auditive</i> .	1	2	3	4	5					
		7.63	22.03	30.51	26.27	12.71					
160.	La longueur du test est convenable.	1	2	3	4	5					
		11.02	14.41	34.75	31.36	8.47					
161.	Le niveau de difficulté du test est convenable.	1	2	3	4	5					
		7.63	16.95	31.36	32.20	11.02					
162.	La mise en page du test est claire.	1	2	3	4	5					
		5.93	1.69	8.47	29.66	53.39					
163.	En faisant le test, je me sentais à l'aise.	1	2	3	4	5					
		8.47	8.47	21.19	33.90	27.12					
164.	Le contenu et les thèmes de la partie <i>Compréhension auditive</i> du test étaient intéressants.	1	2	3	4	5					
		5.93	10.17	36.44	33.90	12.71					
165.	Le contenu et les thèmes de la partie <i>Compréhension écrite</i> du test étaient intéressants.	1	2	3	4	5					
		5.08	8.47	31.36	37.29	16.95					

SVP, COMMENTEZ

Formative Evaluation Phase: Qualitative data from students who did the preliminary test

The students were invited to make their comments in English or French. They are transcribed here verbatim.

Length of the test	<ul style="list-style-type: none">• 5 times "trop long" and "trop de questions"• Cependant, il est un peu trop long.• A little bit long.
General	<ul style="list-style-type: none">• Peu intéressant• Faute d'orthographe p. 9
Instructions and examples	<ul style="list-style-type: none">• Toujours donner un exemple avant de commencer une partie.• At the end the directions weren't really explained.• The instructions of the mini-exposé were not clear enough.• Un exemple aurait été le bienvenue, car il nous aurait fait comprendre la bonne technique à adopter lors de l'écoute car on ne s'attend pas à ce que c'est en réalité. Le fait qu'on dise les chiffres des 4 réponses n'est pas une très bonne idée non plus.
Response time	<ul style="list-style-type: none">• On avait pas le temps de répondre.• Also, 20 seconds to read the answers shouldn't be mandatory. For some questions, less than 5 was enough while for others, 20 seconds was on the border.
Bell	<ul style="list-style-type: none">• La sonnette est très fatigante.• Le "ding" avant chaque question est fatigant.
Speed	<ul style="list-style-type: none">• Ça parlait vite un peu;• Les discussions allaient trop vite
Repetitions	<ul style="list-style-type: none">• Il faudrait répéter les énoncés. (probably meant the dialogues.)

Level of difficulty	<ul style="list-style-type: none"> • Très difficile • Il y a quelques mots que je ne comprenais pas dans les choix de réponses. • The vocabulary was hard. • The vocabulary is kind of hard; this is more literary vocabulary. • Less vocabulary • The degree of difficulty of the test was too difficult in the writing comprehension because the vocabulary was too unusual. • Good diversity in difficulty • A little bit difficult • Les mini-exposés étaient plutôt difficile. • But the mini-exposé seemed a bit too fast. • Some answers are ambiguous in the listening comprehension section. • The questions in some areas of the test (mostly in the written comprehension section) were easy to understand but the answer choice was not always as clear as the question. Example: "how many times ("essais" in French) does it take to finish?" is as good as "how much time (hours) does it take to finish?"
Other	<ul style="list-style-type: none"> • On perd l'attention. • In general, the test is well done. • Le questionnaire est bien fait. • Maybe more reading comprehension.

These comments are transcribed verbatim from the teachers' comment page.

Test design

- Too many questions
 - Way too long; reading long text at the end, the students are too tired; perhaps put it at the beginning.
 - Some of the easier questions could come at the beginning of the test. The 104's were discouraged . Most of the students left before I started the test as they had heard from the other groups that it was impossible to complete.
 - This took over an hour to do. Could the placement test be limited to 40 minutes or so?
-

Soundtrack

- Response time needs better articulation; speaks too fast.
 - Response time 1-13 too long; difficulty understanding woman's voice (#2-3).
 - Form 2 - the only response time problem came in the mini-exposé section.
 - Very clear
-

Instructions

- Students felt it was too slow.
 - Section 3 = students expected questions to be asked. They did not fully understand the instructions.
 - Mini-exposé section = students did not seem to absorb the fact that there were four questions to answer after the exposé. They presumed that the format was similar to the previous sections. Students felt they should have had the time to read the questions before listening to the exposés. Mini-exposé No. 2 was not announced properly - no warning given, and again, no time to read the question before hearing the text.
 - Despite the instructions for section 3, some students (who happened to be very bright individuals) didn't realize that four questions (#29-32) were to be answered based on the same excerpt. I think that students should be warned that the text they'll hear in mini talk 2 is a poem.
 - La partie des mini-exposés était mal expliquée; faudrait donner du temps pour lire toutes les questions-réponses avant la lecture et le spécifier; de plus inclure une petite partie écrite serait bien (100 mots environ) pour mieux voir le niveau dans lequel inscrire les gens.
 - Use a francophone for the instructions and question numbers. Students are so easily distracted. No insult to the speaker intended, but his accent did provoke a few giggles.
-

Items	<ul style="list-style-type: none"> • Form 3= text too long, even for an anglo; students advanced laughed at #23; #75= what does the briefcase have to do with going? #99= no a) choice; #66= I have a problem with the underlined in the following "A few years ago, I _____ in Baie Comeau <u>for four years</u>." • Form 2 = There were 2 questions #23. Students did the second - the first had no corresponding answers. • Listening = lots of idiomatic expressions - too difficult at the beginning of cassette ex. 4a • #7 pronoun error - should be "she" - it's a woman's voice. • #34 p. 8 what» why • There were a lot of "faux-amis" which although good for distinguishing 103's from the rest, do not, it seems to me, help verify a student's general command of the language.
Unforeseen difficulties	<ul style="list-style-type: none"> • Section 3 mini-exposé - no time to read the questions beforehand; • A lot of unknown vocabulary; students felt they had to be very good in English to take the test. • Listening - lots of time to read the statements but not enough time to answer the questions. The speed of the conversation was very difficult for the 104's to understand.
Other comments	<ul style="list-style-type: none"> • Prendre note, une erreur à la question 99. • Students found the test very difficult; they did not feel they would have a high score on it.