AN INVESTIGATION OF THE PRAGMATIC VALIDITY OF THREE CLASSROOM ESL TESTS

Francis J. Bonkowski

A Thesis

in

The TESL Centre

Presented in Partial Fulfillment of the Requirements
For the degree of Master of Arts in Applied Linguistics
at Concordia University
Montréal, Québec, Canada

August 1981

ABSTRACT

AN INVESTIGATION OF THE PRAGMATIC VALIDITY OF THREE CLASSROOM ESL TESTS

Francis J. Bonkowski

The pragmatic validity of a dictation and two modified cloze tests was investigated using correlation and regression analysis. A mixed group of high school ESL learners (n = 52) was divided into two ability levels: low and intermediate. The main concern is whether the three tests are good measures of ESL attainment for learners of differing language ability. In particular, do these tests assess global knowledge of the second language and do they more effectively predict final ESL test performance than either previous ESL achievement or a discrete-point proficiency test? Also are teacher grades in secondary 4 and 5 useful in assessing final ESL achievement?

The hypothesis that procedures like the modified cloze and dictation assess global language knowledge at different levels of ESL ability is not supported. The second hypothesis that these tests more effectively measure ESL attainment than the above-mentioned variables is only partially supported. Different classroom tests are effective for each ability level. The third hypothesis that teacher grades provide useful information in the assessment of ESL performance at different ability levels is supported.

ACKNOWLEDGEMENTS

I wish to express my deepest appreciation to all the kind people who in some way assisted me with this project.

Professors Bruce Barkman and John A. Upshur most generously offered encouragement and technical expertise as well as tough-minded criticism during this endeavour.

A graduate study group formed of Rosalie Banko, Jeffrey Barlow and Valerie Mollard provided intellectual stimulus and moral support to undertake this work.

Aurèle Daoust and Richard L'Abbé of the Collège Ville-Marie were a rich source of inspiration for their interest in educational research.

Professors M. Komoda and E. Brussell of the Psychology Department were most gracious in explaining the use of statistical techniques.

Many thanks are also due to Gary Libben for unlocking the secrets of the SPSS manual, Ann Globensky for numerous bibliographic searches, and Marlene Tash for extensive editing of the final draft.

Finally, heartfelt thanks go to Catherine Faure who so patiently and lovingly edited and typed the final manuscript.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

Classroom Testing

In a high school program of English as a second language one of the primary concerns of the classroom teacher is to establish the current level of achievement of his students. The classroom teacher needs to be aware of how much the students know and of what they can do in the second language. With this knowledge the effective teacher can adapt himself to the language abilities of his students. In order to get accurate information about them, the teacher must use evaluative tools that are not only simple to use, but that also provide consistent results and are truthful in what they measure (Green, 1975). The classroom teacher often devises ad-hoc tests to suit his own purposes (Davies, 1977). These tests do not always produce reliable and valid information; it is therefore difficult to compare an individual student to the group with such tests. The purpose of this study is to investigate the validity of three ad-hoc tests used to discriminate among learners in two ESL groups in the last year of high school.

Two common classroom tools, achievement and diagnostic tests, have their own special purposes. Teacher-made achievement tests are designed to measure how much of the course material presented over a period of time has been mastered. Their linguistic content is derived from a particular curriculum. Diagnostic tests, on the other hand, aim at revealing patterns of error among students. Neither test provides useful quantitative information about what Clark (1972) calls the "current nature or level of student accomplishment" in ESL without respect to any particular school instruction (p. 2). The term "attainment test" as used by Ingram (1968) describes an instrument which assesses "what the student knows and can do irrespective of how he learned it or what his learning ability is" (p. 38). This is quite similar to a proficiency test which measures the student's ability to use the language for real-life purposes (Clark, 1972). The main difference according to Davies (1977) is that a proficiency test implies control of the language for a specific "extra-linguistic purpose" (p. 46). Cohen (1980) suggests that the same classroom or school test used to assess achievement at the end of instruction can be used to assess language proficiency at the beginning. It would seem that an "attainment test" combines aspects of both the constructs of achievement and proficiency. It tests global

knowledge of the linguistic elements of the second language and indicates whether students control language skills.

A current trend in second-language testing is the use of global language tests which measure the student's ability to understand and use the language in context (Valette, 1977). Two examples are the cloze procedure and dictation. Oller (1979) claims that they are integrative tests because they reflect the "normal uses of the language" and they incorporate an element of real time (p. 39). Both tests can be especially useful for the classroom teacher (Cohen, 1980). Since the teacher is often uncertain about which testing technique to use, the cloze procedure and dictation offer him simple and practical ways of testing. It has been claimed that these testing techniques are highly reliable and valid with second language learners (Oller, 1973a). These claims will be discussed in Chapter II.

Specific Problem: Ranking ESL Students

At the Collège Ville-Marie, a co-educational private high school for French-speaking students, students are not streamed in second language classes. As a result ESL teachers are faced with students who vary considerably both in their knowledge of English and their ability

to understand and use the second language. There are also new students who arrive every year from other institutions. Confronted with this mixed grouping of students, teachers who wish to rank students can consult previous ESL school and final achievement grades, but these scores are not always readily available. The Collège encourages its teachers to give "starting point" tests to assess the current level of attainment of the students in ESL. The results permit the teacher to compare one student to another and to rank them within their particular group. With this information the teacher can adapt both the curriculum - where it is possible - and his teaching strategies to the range of learning abilities in the group.

The Ministry of Education in Quebec plans to abolish streaming of ESL students at the secondary level. This will create a situation at the provincial level which the Collège Ville-Marie has already experienced for many years. To prepare for this eventuality the Ministry of Education is developing a set of attainment tests entitled, "Tests relatifs à l'apprentissage scolaire et parascolaire".[1] These objective tests will presumably furnish useful quantitative information about the ESL attainment of French-speaking students in secondary 3 and 4.

It is within this framework that this study investigates the pragmatic validity of three teacher-prepared classroom ESL tests. These tests are based on the cloze procedure and dictation, both of which are potentially useful classroom tools. They seem to provide a range of scores and to discriminate well among students of different abilities, two essential characteristics of good language tests (Allen & Davies, 1977).

Approach

This study, using two groups of ESL high school students of different language abilities, examines the relationship in scores obtained from three teacher-prepared global language tests: two modified cloze tests and a dictation test. Two other kinds of variables are also included in the study: a discrete-point proficiency test and ESL achievement results obtained from the provincial achievement examinations and teacher grades in secondary 4 and 5. The primary aim of the study is to investigate whether the three global language tests administered in three different school terms possess good discriminative power and whether they effectively measure comprehensive ESL skills for students of different ESL abilities in secondary 5. A secondary aim is to ascertain whether the three global language tests are better

predictors of final ESL achievement than either a proficiency test or previous ESL school performance.

A discrete-point proficiency test is used as a variable in the study to test an assertion made by proponents of global language tests. Oller, in particular, has claimed that the cloze procedure and dictation produce global language tests that "provide more accurate information concerning language proficiency" and even level of achievement than language tests based on the discrete-point approach (1979, p. 9). ESL achievement results - including both provincial examination scores and teacher grades in secondary 4 and 5 - are used as variables to examine their usefulness in predicting ESL scores from one year to another. Lother (1967) has shown that the scores of school language examinations and final standardized language examinations taken in one year correlate highly with the scores of similar tests taken the following year for native speakers of English.

A third aim of this study is to examine whether teacher grades add useful information in the assessment of final ESL achievement. Guilford (1965) has stated that school marks as assigned by teachers are sometimes considered poor measures of achievement. The Quebec Ministry of Education includes both teacher grades and

provincial examination results in equal degree in computing final ESL achievement grades. This study investigates whether teacher grades add useful information in the assessment of ESL achievement when they are added to final provincial examination results.

Research Questions

In order to evaluate three teacher-prepared global language tests - two modified cloze tests and a dictation test - used to determine the level of ESL attainment of high school students of differing ESL abilities, this study examined the relationship among several measures of ESL achievement. There are three main research questions:

1. Do scores on the two modified cloze tests and the dictation test correlate highly enough to indicate that they are measuring similar linguistic skills?

2. Do the scores on the two modified cloze tests and the dictation test account for variance in final achievement examination scores that is not predicted by either the scores on a proficiency test or previous achievement scores?

3. Do teacher grades add significant variance to the scores of the secondary 5 final ESL grade?

Assumptions and Limitations

Many assumptions are inherent in a preliminary investigation of this kind undertaken within the classroom. They arise from the nature of the sample, from the kind of tests used, and from the availability of a criterion measure.

The students were assumed to belong to two different ability groups, low and intermediate, on the basis of previous teacher grades in secondary 4. The students are considered quite representative of French-speaking high school students in Montreal.

Two of the instruments (the modified cloze tests) are regarded as legitimate members of the family of global language tests even though they are, in fact, tests of recognition and not production. The data they provide can be discussed in the light of previous research with the cloze procedure and dictation. Another instrument, a "proficiency" test developed in a graduate testing course, was never statistically validated against an external criterion. It is assumed, however, that the test possesses good content validity since the language domain from which the items were chosen was felt to be typical of the level of ESL needed at CEGEP or junior college.

The criterion measure used for achievement, the final provincial achievement examination administered by the Ministry of Education, is assumed to have good content validity.

In a study investigating the statistical relationship among test scores and teacher grades in which correlation and multiple regression are the main techniques used, the larger the sample the more stable the results. Correlation coefficients are particularly susceptible to sampling fluctuations when the samples are small, and coefficients that are quite high may occur due to chance alone (Tate, 1968). Also there is considerable unreliability in using multiple regression equations with small samples (Kerlinger & Pedhazur, 1975). It would therefore be imprudent to generalize about the results of this preliminary study to a larger student population without further investigation in other schools.

There is a weakness in the research design that should be noted. Although the purpose of the study is to investigate the usefulness of three classroom tests as measures of student attainment in ESL at the beginning of the year, the tests were given in three different terms of the language course. Both the modified cloze test and the "proficiency" test were given in the third term of study. All four tests under study should have

been given in the first term at approximately the same time. This would have eliminated the possibility of higher correlations between the scores of the two tests and the scores of the criterion measure due to the short interval between test administrations.

FOOTNOTES

[1]The tests are currently being developed at the Direction générale du développement pédagogique by a committee headed by D. Lussier-Chasles, Spring, 1981.

CHAPTER II

REVIEW OF THE LITERATURE

Cloze and Dictation Procedures

Research using both the basic cloze test and dictation has generally shown rather high correlations between the socres of the two tests. From these correlations researchers have concluded that these two types of test measure overall second language proficiency.

Oller and Conrad (1971) have investigated whether the cloze procedure effectively differentiates among the levels of proficiency of non-native and native speakers. They compared the scores on a cloze test (using exact-word answers) with the results of the UCLA ESL placement examination whose subtests included vocabulary, grammar, reading comprehension and dictation. Although the cloze test did distinguish well between beginning and intermediate ESL learners, it did not discriminate among learners of higher proficiency. The highest product-moment correlations were obtained between the cloze and both the dictation subtest (.82) and the reading comprehension subtest (.80) of the placement examination. Using questionable reasoning,[1] the researchers claimed

that the cloze test and the two above-mentioned subtests are "all measures of integrative skills" (p. 192).

Oller (1972a) reported a study comparing three cloze tests of different levels of difficulty and the revised form of the UCLA ESL placement examination consisting of the four previously-mentioned subtests. He found that, of all the subtests, dictation correlated highest with the easy cloze (.76), the medium cloze (.84), and the difficult cloze (.85). Having assumed that the cloze tests were valid measures, the researcher noted that the dictation provided a "comprehensive sampling of the integrative skills involved in the understanding of a language" (p. 352).

In another study Oller (1972b) investigated the effects of different scoring methods and levels of difficulty on the results of the cloze procedure. The revised form of the UCLA ESL placement examination was used again in comparison with three different cloze tests. Regardless of the scoring method used, the combined score of the cloze tests correlated highest with the dictation subtest of the ESL placement examination. The dictation results correlated with the cloze scores at .71 for the exact-word scoring method and at .80 for the acceptable-word scoring method. It was claimed that the cloze procedure tends to "correlate with tests that require high-

level integrative skills" like dictation (p. 157).

Both the cloze procedure and dictation were compared with the TOEFL (Irvine, Atai & Oller, 1974). A cloze test and two dictations, one easy and one difficult, were given to 159 foreign students. The cloze passage was scored by both the exact-word and acceptable-word methods. The acceptable cloze correlated slightly more highly (.75) with the dictation total score than the exact word one (.69). As Table 1 shows, the listening comprehension subtest of the TOEFL correlated at .76 with both cloze tests and at .69 with the dictation total score.

Table 1

Pearson Product-Moment Correlation Matrix for TOEFL Part-Scores of Listening Comprehension, the Cloze Test Scored by Exact and Acceptable Word Methods, and Two Dictations

	1	2	3	4	5	6
1. Listening Comprehension	X	.76	.69	.76	.70	.67
2. Acceptable Cloze		X	.75	.94	.76	.72
3. Dictation total			X	.69	.96	.93
4. Exact cloze				X	.70	.69
5. Easy cloze					X	.85
6. Hard dictation						X

Note. From Irvine, Atai, and Oller, 1974, p. 249.

The lowest correlation was between the listening comprehension subtest and the hard dictation which was based on a passage from Scientific American. The findings supported the claims Oller had made in previous research. The authors concluded that cloze tests, dictations and listening comprehension tasks were integrative in nature because they tap the learners' underlying language competence. They "account for more variance whenever testing methods are compared than do discrete sampling approaches" (p. 250).

Krzyzanowski (1976) also investigated the validity of the cloze procedure. He gave a cloze test, a dictation, a structure test and a vocabulary test to 93 Polish high school students. The sample consisted of two groups at the intermediate ESL level and a third group at the advanced ESL level. Krzyzanowski found that the cloze test "discriminated sufficiently between the test subjects at all levels of proficiency" (p. 32), although it discriminated better at the intermediate level. Cloze scores, based on the acceptable-word method, were correlated with the scores on the other tests. The results are shown in the following table:

Table 2

Coefficients of Correlation Between Cloze and Other Tests

Group	Cloze Reliability	Vocabulary	Structure	Dictation
1 (Interm)	.86	.89	.88	.86
2 (Interm)	.87	.90	.79	.75
3 (Adv)	.72	.65	.49	.78

Note. From Krzyzanowski, 1976, p. 34.

Noting that the scores of the cloze test correlated best with scores of the vocabulary test for groups 1 and 2, Krzyzanowski stated that "the assumption that a cloze test correlates the best with dictation (Oller & Conrad, 1971) does not apply" for the two intermediate groups (p. 33). He also noted that the high correlations in the second language measures might be due to the kind of general second language instruction that the students in the study received. The four tests under investigation might be tapping various components of a second language skill that represent language proficiency.

In a study examining the effects of text difficulty, scoring method and deletion rate on the validity of the

cloze procedure, Alderson (1979) used the cloze procedure and dictation in comparison with Ingram's ELBA test. He gave 12 different cloze tests to 12 groups of 30 students. Changing any of the three variables described above had drastic effects on the correlations of the cloze tests with the other measures. It was found, for example, that an easy cloze test, corrected by the semantically acceptable method and deleted at every 8th word, correlated at .45 with an easy dictation. The same text, corrected in the same way but with every 12th word deleted, correlated at .91 with the easy dictation. Of all the scoring methods used in the study, the acceptable-word cloze correlated best with dictation. But even with this scoring method, the range of correlation with dictation was between .38 and .91. Alderson also found that the cloze tests related more to the subtests of grammar and vocabulary than to the subtest of reading comprehension. The cloze tests did not relate more to dictation than to the ELBA test. The author could only conclude that the cloze procedure is very sensitive to the deletion of individual words. The study also showed that not every method of creating a cloze procedure automatically produces valid tests as the previously mentioned studies had claimed.

It has been suggested that the cloze procedure and dictation can be effective in measuring achievement (Cohen, 1980). Valette (1977) noted that a dictation test consisting of unfamiliar material could serve as an achievement test when students are compared across different programs.

The cloze procedure has been the object of interest in several studies investigating its relationship to academic achievement, and particularly language achievement. In his clozentropy study, Darnell (1970) compared the cloze scores and the results of the TOEFL with the grade-point average of the students in the sample. He found no significant correlations at all.

Davies (1975) correlated a speeded modified cloze test, which is described on page 24, with final ESL achievement results and a tutor's assessment of second-language proficiency. He reported correlations ranging between .40 and .70, but no other details were mentioned. In a study of the multiple-choice cloze technique, Jonz (1976) found a correlation of .71 between his MC cloze test and the class placement of 33 foreign students grouped into five different course levels.

Hanzeli (1977) gave 107 college students learning French at three different course levels a cloze test of 80

items. Every 7th word was deleted from the passage selected. In the study he compared the scores of the cloze test with two academic measures of achievement: (a) general grade-point average at the university, and (b) final course grades in French. Correcting the cloze test for acceptable answers, Hanzeli found the following correlations with class grades in French:

Table 3

Correlations Between Acceptable Cloze and Course Grades in French

Class	N	r^a		
301	8	.45	$p <$ .131	NS
203	24	.69	$p <$ .001	
202	24	.65	$p <$ .001	
201	16	-.15	$p <$ .293	NS
103	35	.45	$p <$ .008	

Note. From Hanzeli, 1977, p. 872.

[a]Pearson product-moment correlations.

Hanzeli noted that the small sample size in group 301 and the lack of seriousness in group 201 probably explained the insignificant findings. Yet on the whole he was confident that "cloze scores predict course grades" quite well (p. 871). He called for more research into the matter.

Both the cloze procedure and dictation have been studied as measures of language proficiency - for native speakers of English and second language learners - in relation to standardized intelligence and general language achievement tests. Stump (1978) and Streiff (1978) investigated to what extent these two global language measures as well as standardized tests are dependent on underlying language proficiency.

Stump (1978) studied the relationship of scores on the cloze procedure, several dictation tests, the Iowa Test of Basic Skills (ITBS) and the Lorge-Thorndike Intelligence test administered to approximately 100 English speaking children. Using a principal components analysis, Stump stated that one single factor accounted for 54 percent of the total variance in all the above-mentioned tests. He concluded that cloze and dictation tasks accurately predict scores on standardized intelligence and achievement tests.

Streiff (1978) also hypothesized that cloze scores "predict, or explain in a statistical sense, a great deal of the variance in performance on school tests" (p. 71). In previous research she had found that, for a group of 2nd and 3rd grade Alaskan Eskimo children, there were high correlations between the cloze procedure and the Metropolitan Achievement reading subtest (.67) and the

total test (.81). In this study she studied the performance of 47 Hopi bilingual children on a written cloze test, an oral cloze test, and the California Achievement test and compared it with the performance of English speakers. Using a principal components analysis she found that a single factor contributed an average of 79 percent of the variance to each test of the Hopi students. Streiff concluded that both the California Achievement test and the two kinds of cloze test were dependent on an underlying language proficiency factor.

In summary, the research studies discussed in this chapter examined the cloze procedure and dictation in relation to different external criterion measures for students at all academic levels. There is contradictory evidence about the validity of these two testing techniques. There are some indications though that the high correlations between the scores of the cloze procedure and dictation may, on the one hand, be due to some underlying language proficiency. Or, on the other hand, they may be attributable to various language components learned during instruction. It would also seem that both techniques are promising tools in predicting academic success in a second language.

Modified Cloze Procedures

There has been considerable interest in different modifications brought to the basic cloze procedure. One modification consists of supplying the student with alternate responses, either one or more distractors from which to choose. Assuming that the traditional cloze procedure was too difficult for English speaking 1st grade students, Gallant (cited in Jongsma, 1971) used a multiple-choice format in which two distractors were provided. The researcher, though, never stated how the distractors were selected.

Porter (1976) suggested that the multiple-choice cloze procedure was a useful test of reading comprehension in a foreign language learning situation. He reasoned that beginning and intermediate students could probably recognize the appropriate word in a cloze passage if they were given a number of alternatives. By eliminating the need to produce language, the modified cloze procedure would provide a simple and straightforward means of testing comprehension. Although the selection of alternative words is not a simple task, Porter thought that the test writer could use his intuition in writing distractors of varying difficulty according to the linguistic attainment of the students. For instance, distractors in tests for beginners would be

both grammatically and semantically unacceptable; whereas distractors devised for advanced students would be all grammatically suitable when knowledge of register is called upon.

Jonz (1976) deviated from the regular cloze procedure and experimented with the multiple-choice format. He did this for two reasons. First, the modified cloze procedure afforded the test scorer greater objectivity, and second, it allowed students to compare their own responses with a limited range of choices. He created his choice of distractors from the highest frequency unacceptable responses supplied by non-native speakers taking a conventional cloze test in English.

Ozete (1977) developed a modified cloze test for beginning and intermediate students known as the reading-input test. It was intended to determine the students' ability to comprehend reading materials at different levels of difficulty. Only one distractor, which was chosen from the vocabulary of the text, was given along with the correct response. The researcher claimed that the modified cloze procedure didn't interfere with the reading process, that it tapped reader familiarity with elements of the passage, and that it isolated the reading process from the writing process as Porter had reasoned (p. 566).

In order to establish the reading level of 5th grade English students, Piluski and Piluski (1977) experimented with a multiple-choice cloze test known as the maze technique. The students were provided two distractors, one that was syntactically acceptable but semantically inappropriate, and a second one which was neither semantically nor syntactically appropriate. One of their conclusions was that researchers should be careful in selecting distractors semantically closer to the omitted word.

Hinofotis and Snow (1980) studied the multiple-choice cloze as an evaluative tool in measuring second language proficiency. Its main advantage is that it simplified the scoring procedure of the basic cloze test. The distractors in their modified version were taken from the most frequent incorrect responses of non-native speakers who took an open-ended English cloze test. Results showed that the modified cloze was easier than the open-ended version. Also scores on the modified cloze correlated less well with an external criterion than the scores of the conventional cloze test scored for either exact or acceptable responses.

A different kind of modified procedure was proposed by Davies (1975). He deleted only certain categories of words, particularly function words, and he left the first letter of the correct response as a clue.[2] Davies'

rationale for these changes to the conventional cloze was that it seemed practical (p. 124). Oller (1979) has criticized this fill-in-the-blank format because it departs from the theory of pragmatic testing. Hinofotis (1976) termed it a discrete-point test because of its selective sampling of function and content words. In spite of these criticisms, Davies' suggestion that the first letter of the correct response be given as a clue is used as the basis for the modified cloze in this study.

These are some of the modifications that have recently been made to the basic cloze procedure in order to improve upon it.

FOOTNOTES

[1]By grouping together the scores of ESL students of widely-varying language abilities, the researcher obtained high correlations. These might be the result of the amount of instruction rather than of some underlying language proficiency.

[2]Davies highly recommended that ESL students be provided the first letter in order to increase motivation. Personal communication, Concordia University, 1977.

CHAPTER III

PROCEDURE

Description of the Sample

The secondary 5 students in the study attended the Collège Ville-Marie in Montreal. The Collège is a small, private co-educational high school for French-speaking students in grades 7 to 11. It seeks to provide quality education that is an alternative to public schooling. Although it charges a moderate tuition fee, its doors are open to any family. Unlike other private institutions, the Collège draws its clientele from all socio-economic classes and from all levels of academic ability. Its student body is, therefore, very similar to that of a public school.

Students in secondary 4 and 5 at the Collège are grouped together first, on the basis of mathematical skills, and second, according to the courses they have selected. All students in the study had previously studied ESL for four years at the Collège or at another French language school in the Montreal area. All the students followed the same ESL course in secondary 5 - a course taught by this researcher. Students with a false beginner's knowledge of English attended the same class as those with more advanced

ESL skills. The students were assigned to two groups referred to as Group 1 (low level) and Group 2 (intermediate level) using teacher grades from secondary 4 as the basis for group assignment. Group 1 contained 30 students (18 boys and 12 girls), and Group 2 contained 22 students (18 boys and 4 girls).

Variables

Three teacher-made classroom tests, two modified cloze tests and a dictation test, as well as an experimental proficiency test were administered to the students during three terms of the 1977-78 academic year. This was done in order to obtain information about their current level of ESL attainment. In addition, the students' three ESL achievement grades earned in secondary 4 were used as data: (1) the year's school mark as assigned by the classroom ESL teacher, (2) scores on two final provincial ESL achievement examinations (Exam 412 for Group 1 and Exam 422 for Group 2), and (3) a final grade which combined in equal degree scores 1 and 2. These measures constitute the seven independent variables in this correlational study. Scores of the two groups on the final provincial ESL examination obtained at the end of secondary 5 represent the dependent variable for the second research question. This question, formulated in Chapter I,

investigates the correlation of scores of these seven variables with later ESL academic performance of the students at different levels of ESL ability. The four independent variables for the third research question are the three achievement scores in secondary 4 and the scores on the dictation test. The dependent variable is the final ESL grade obtained at the end of instruction in secondary 5.

The three research questions are restated here:

1. Do the scores on two modified cloze tests and a dictation test correlate highly enough to indicate that they are measuring similar linguistic skills?

2. Do the scores on two modified cloze tests and a dictation test account for variance in final examination scores that is not predicted by either previous achievement scores or the scores on a discrete-point test of proficiency?

3. Do teacher grades add significant variance to the scores of the secondary 5 final ESL grade?

Multiple-choice cloze test. The modified cloze test of 50 items was devised and administered to the students at the beginning of the first term of instruction. A modified cloze procedure providing four alternate responses was used for two reasons: (1) to provide greater ease and objectivity in scoring, and (2) to reduce the

frustration of the ESL students taking the test.

A passage of approximately 350 words in length was chosen from Scope Magazine (1976). The magazine contains reading material that secondary 5 students are expected to handle with ease or at least with no great difficulty. Entitled "A Battle with Sharks" and written by H.S. Mazet, the story was an adaptation written at the 2nd grade reading level for native speakers according to Fry's estimate of readability. The Fry readability formula was used because it is recommended for reading material at the primary grade level (Fry, 1968).

Fifty words were left at the beginning of the passage as an introduction. Then every 5th word, in most cases, was deleted to produce a total of 50 blanks. There were four distractors accompanying the correct answer and they were placed on separate pages at the end of the story. In most cases the distractors belonged to the same grammatical category as the correct word. The following sentence is an example from the test:

4. _____, Bent lost his balance and 5. _____ into the sea.

4. a) Lately
   b) Soon
   c) Finally
   d) Now
   e) Suddenly

5. a) falls
   b) fell
   c) felt
   d) fallen
   e) feels

The distractors in question 4 are syntactically, but not semantically, acceptable. Question 5 tests knowledge of verb tenses as well as word meaning (See Appendix 5 for a description of the item analysis of this test).

Dictation test. This test was administered to the students during the second term of instruction. The text used was 150 words in length and was chosen from the classroom textbook the students were using, Book 5 of the New Horizons in English series (1974). The students had not previously studied the dictation text. It was in the form of a letter from one friend to another. The subject was a car accident.

According to Fry's readability estimate, it was equivalent to a 5th grade reading level for native speakers. The letter begins:

It was a real shock/ to get your latest letter./
You're the last person/ in the world/ I'd ever
expect/ to be involved/ in a three-car collision./

The dictation was read three times according to the procedure outlined in Oller (1979).[1] The whole text was read once at normal conversational speed. Then it was read a second time, with punctuation included, in phrases of between three and seven words. Finally, the whole text was read a final time. One point was subtracted for each

error which consisted of an incorrect word (Your for You're), missing or added words, and faulty word order. Also certain flagrant spelling mistakes (tree for three) were considered as errors.

Modified cloze test. A modified version of the cloze procedure based on Davies' suggestion (1975) was prepared and administered to the students in the third term of instruction. The conventional cloze procedure was changed for the same two reasons mentioned for the other cloze test. The passage for this test was approximately 325 words in length and was selected from the SRA Reading Laboratory IIa. Entitled "Lost in the Woods", the story was judged at a 2nd grade readability level by Fry's estimate and a 3rd grade level according to SRA. Every 5th word was deleted following an initial lead-in of approximately 30 words. The first letter of each correct response was left in the text at the beginning of the blank as follows:

One camper could not f_____ his way back to t_____.

There were 35 content words (like "find" and "town") and only 15 structure words.

The students were instructed to fill in the blanks with words that completed the meaning of each sentence. Students had to write their answers on an accompanying answer sheet. The fact that the blanks in the cloze passage were not numbered presented a little difficulty since, later in the test, it was not always easy to associate the blank in the passage with the number on the answer sheet.

Proficiency test. This test was a revised version of an instrument prepared by this researcher and two other teachers in a graduate testing course at Concordia University (Bonkowski, Evans & McKrae, 1977). It was administered to the sample during the third term in the same period as the modified cloze test. It was specifically aimed at French-speaking high school students and designed to test their second language proficiency for entry into college (CEGEP). The original instrument had five parts: grammar, vocabulary, reading comprehension, a modified cloze passage (described previously), and a dictation test (different from the one in this study). The first three sections were used in this test and a fourth section was added based on material covered in class. Three distractors were provided for the 60 questions in the first three sections.

The 25 items on the grammar section were chosen from material presented in the curriculum of the basic English grammar, English Sentence Structure by Robert Krohn (1972). Most of the 20 items in the vocabulary section were drawn from the AA category of the Thorndike-Lorge word count which includes approximately 1000 words. The source of the texts for the reading comprehension subsection containing 15 questions was the translation section of the French A level exams (London Board).

The test was tried out with a group of 80 secondary students at the College in the spring of 1977. After item analysis techniques were employed, several vocabulary and grammar items were rewritten. Two of the three reading passages on which ten questions were based were rejected because they were too difficult. Two new reading passages were chosen from the SRA Reading Laboratory IIa with a 5th and 6th grade reading level.

The fourth section contained ten fill-in-the-blank questions and five true-false questions based on a reading assignment in Scope Magazine given the previous week. It dealt with the history of American entertainment in the late nineteenth and early twentieth century. Since the students could not refer to the text, the 15 questions required not only understanding of the material,

but also a good memory in order to recall it (See Appendices 1-4 for the full texts of the three classroom tests and the proficiency test). In total, the revised test contained 75 questions.

Achievement scores. The three ESL achievement scores in secondary 4 were obtained in the following way:

1. The "teacher grade" represents a student's class performance on quizzes and tests in ESL over the four terms of the academic year. This grade is moderated in terms of the student's results on the final ESL achievement examination (Ministry of Education, 1974).

2. The score on the provincial final ESL examination taken at the end of the year is reported as a converted score derived from a formula applied at the Ministry of Education. This conversion does not change the rank of a student within his school group.

3. The final ESL grade is an average of the converted final examination score and the moderated teacher grade. A student's final report card indicates the classroom teacher grade and the final grade, not the converted score on the final examination.

The provincial final ESL examination in secondary 5 serves as the validating criterion in the second research question. It was prepared by an experienced language

specialist who followed a detailed table of specifications and a specific vocabulary list not exceeding 2500 words. It was then carefully revised by a committee of qualified readers and rewritten according to specifications. The examination is divided into five sections and includes 50 multiple-choice questions each containing four distractors. The five sections include:

1. Vocabulary. 10 items testing knowledge of antonyms, synonyms, and word meaning in context. The following is a sample item:

If Ray is sick, we must postpone the trip. (Synonym)

a) cancel
b) hasten
c) delay
d) oppose
e) plan

2. Grammar. 15 items testing application of grammatical knowledge of verb tenses, pronouns, adjectives, gerunds and infinitives.

3. Word order. 5 items testing the order of elements in a sentence such as the following:

His mother asked/ why/ had/ he/ run/ away.

4. Reading comprehension. 10 questions in all, five questions following each of two 300-word texts testing literal understanding as described in Barrett's Taxonomy (1968).

5. Global comprehension. 5 items to complete by choosing the correct phrase:

Although Celia always shops at expensive stores, ...

a) she spends too much money
b) she can't really afford it
c) she always looks well dressed
d) she likes expensive clothing
e) she dresses beautifully

In addition, five comic strips are included. The students must interpret each one by choosing the best short descriptive paragraph for it.

One can safely assume that the provincial examination has a high degree of content validity since it adequately samples the linguistic knowledge and reading skills that are expected of secondary 5 students.

The final grade obtained at the end of secondary 5 - a composite of the school mark and the results on the final achievement examination - serves as the external criterion for the third research question in the study.

Statistical Procedures

A correlational approach is used to determine the pragmatic validity of the three teacher-made classroom tests. Pragmatic validity can be defined in terms of the estimates obtained by correlating test scores with the

scores obtained from a criterion measure (Ingram, 1977). We are specifically concerned with predictive validity since the criterion measure chosen is the final achievement examination prepared at the Quebec Ministry of Education and administered at the end of instruction. If a strong statistical relationship exists among the scores of the three classroom tests as well as between the scores of the three tests and the scores of the achievement examination, then the classroom tests are indeed good predictors of ESL achievement.

Correlation. The first aspect of the study is the correlation of the scores of the two groups on the three classroom tests. The second aspect is the correlation of scores of seven variables: the three classroom tests, the proficiency test, and the three achievement grades in secondary 4 with the scores of the final achievement examination in secondary 5. The correlation coefficients that result provide an index of the tendency for scores on one measure to covary with scores on the other measure (Oller, 1979). They indicate the direction and magnitude of the relationship among the measuring instruments (Tate, 1968).

In discussing test validity Ebel (1972) stated:

> Scores on good tests intended to predict educational achievement correlated with subsequent good measures of achievement to the extent reflected in coefficients that average about .50 and range from about .30 to .70, depending on the nature of the achievement, the quality of the measures of promise and of attainment, the interval between the measures, and many other factors. pp. 302-303.

Within the realm of language tests that measure the same skill or skills, Harris (1969) noted that "we should expect a fairly high correlation - one in the .70s or .80s perhaps" (p. 20). The higher the correlations between test scores and the criterion measure the better we can predict the criterion from the test. Ingram (1977) provides a valuable guide to the value of correlation coefficients:

.90 to 1.00	very high correlation; very strong relationship
.70 to .90	high correlation; marked relationship
.40 to .70	moderate correlation; substantial relationship

.20 to .40	low correlation; a definite relationship but a small one
.20 and less	slight correlation; relationship so small as to be negligible (p. 25)

Regression. The technique of multiple regression is applied to the scores of students at both ability levels for all variables in order to investigate the second research question. This statistical procedure allows the researcher to compare a number of variables at once with a criterion measure (Kerlinger & Pedhazur, 1975). The amount of variance that the scores of each variable contribute to the final examination scores can be determined. The independent variables are entered into the regression equation in different order depending on the research question. The squared multiple correlation coefficients that result, known as R^2, indicate the extent to which the variance in the criterion variable is accounted for as each of the independent variables is added in linear combination (Horst, 1966). As a result, multiple regression indicates which variable or combination of variables is the best predictor of achievement.

For the third research question, the same technique is also applied to selected variables in the study and the final grade in secondary 5 as the criterion measure. It indicates whether teacher grades add significant information in the assessment of achievement for students at different levels of ability in the second language.

FOOTNOTES

[1]Oller recommended this approach in a seminar on testing at Concordia University, Summer, 1977.

CHAPTER IV

ANALYSIS OF THE DATA

This study investigates one aspect of the pragmatic validity of second language tests - in particular, the predictive validity of three teacher-prepared classroom ESL tests used to discriminate among French-speaking high school students in secondary 5 (grade 11). The ESL students, who made up two groups at different levels of language ability, were so grouped according to the judgement of the secondary 4 classroom teacher. The three classroom tests, two modified cloze tests and a dictation test, were examined for their effectiveness in measuring ESL attainment.

The first two research questions presented in Chapter I touch upon the measurement characteristics of the three global language tests used with two different ESL groups. The third research question concerns the usefulness of teacher grades in assessing the ESL academic performance of the same two groups. These groups represent students at the intermediate and low level of ability in ESL. The three research questions, restated here, examine the effect of ability level on classroom

test results and achievement grades. First, do the three teacher-prepared classroom tests adequately measure the comprehensive ESL skills of secondary 5 students? Second, are the scores on the classroom tests better predictors of final ESL achievement, as measured by the provincial 522 ESL examination, than previous achievement grades or the scores on a proficiency test? Third, do teacher grades provide significant information in the assessment of ESL achievement when they are added to provincial examination results?

Summary Statistics

The mean, standard deviation, and range of scores were calculated for the three classroom tests, the proficiency test, and ESL achievement grades in secondary 4 and 5. Reliability estimates of the three classroom tests and the proficiency test were calculated using the Formula Kuder-Richardson 21 (Ary, Jacobs & Razavich, 1979, p. 214).[1]

The data for students of low ESL ability are presented in Table 4 and for intermediate level students in Table 5. Results of the two cloze tests, taken six months apart, show that Group 1 scored 42 percent and 41 percent correct respectively whereas Group 2 scored above 70 percent correct on both tests. However, standard deviations,

Table 4

Summary Statistics of Classroom Tests, Proficiency Test and Secondary 4 & 5 ESL Achievement Grades for Group 1 (n = 30)

Independent Variables	No.[a]	Mean	% Correct	S.D.	Range	K.R.[b]
Multiple Choice Cloze	50	21.2	42	6.2	7-37	.70
Modified Cloze	50	20.5	41	6.9	5-39	.76
Dictation	150	105.6	70	19.5	50-135	.89
Proficiency Test	75	43.0	57	11.0	24-63	.86
Independent Variables	Achievement Secondary 4					
Final Grade	99	67.7		7.4	53-81	
Teacher Grade	99	69.4		8.0	54-89	
Provincial Exam 412	99	65.8		10.1	43-82	
Dependent Variables	Achievement Secondary 5					
Final Grade	99	54.6		8.6	36-70	
Provincial Exam 522	99	55.0		11.3	29-72	

[a]Indicates maximum score for each variable.

[b]K.R. 21 formula assumes item independence in the test. The reliability estimate for the dictation is being overestimated.

Table 5

Summary Statistics of Classroom Tests, Proficiency Test and Secondary 4 & 5 ESL Achievement Grades for Group 2 (n = 22)

Independent Variables	No.[a]	Mean	% Correct	S.D.	Range	K.R.[b]
Multiple Choice Cloze	50	36.8	74	6.0	25-46	.75
Modified Cloze	50	35.8	72	5.9	24-47	.70
Dictation	150	132.1	88	16.3	90-147	.94
Proficiency Test	75	53.3	71	10.0	37-71	.86
Independent Variables	**Achievement Secondary 4**					
Final Grade	99	75.0		7.6	60-89	
Teacher Grade	99	76.8		8.2	63-90	
Provincial Exam 422	99	74.0[c]		10.6	53-90	
Dependent Variables	**Achievement Secondary 5**					
Final Grade	99	76.9		11.0	56-95	
Provincial Exam 522	99	77.7		10.5	54-92	

[a] Indicates maximum score for each variable.

[b] K.R. 21 formula assumes item independence in the test. The reliability estimate for the dictation is being overestimated.

[c] Based on 19 scores.

ranging between 5.9 and 6.9, and reliability estimates, ranging between .70 and .76, of the cloze tests were very similar for both groups. In sum, the two cloze tests effectively discriminated among students in each group.

Both groups performed best on the dictation test. Low ability students scored 70 percent correct while intermediate students had an average score of 88 percent. Scores were negatively skewed for both groups; that is, the median score was well above the mean score. Yet scores on the dictation test showed good variability, SD = 19.5 for Group 1 and 16.3 for Group 2. Reliability estimates, though probably overestimated by K-R 21, are very good for Group 1 (.89) and Group 2 (.94). In spite of the relative ease of this task, the dictation test did produce a good spread of scores for each group.

The experimental proficiency test had better reliability (.86) than the two cloze tests for each group. Standard deviations for the two groups were about the same (11 and 10). Low ability students had a mean of 57 percent on this test compared to approximately 41 percent on the two cloze tests. On the other hand, intermediate students scored slightly lower than they did on the two cloze tests (71 percent compared to 74 and 72 percent).

Data for final ESL achievement over two years indicate that low level students performed more poorly in secondary 5 while intermediate students performed similarly in both secondary 4 and 5. Low ability students took the provincial 412 ESL examination in secondary 4. A year later the two groups took the same 522 ESL examination. This partly explains the lower scores for the low ability students. Scores on these final achievement examinations in both years for each group showed good score variability and range, indicating that these measures are adequately discriminating among ESL students of different language ability.

T-Test

The subprogram T-test for a comparison of sample means found in The Statistical Package for the Social Sciences (SPSS), 2nd. Edition, 1975, was used for all four tests and final achievement grades in secondary 4 and 5. The purpose of this statistical analysis is to ascertain whether there is in fact a significant difference in scores between two previously-established groups. The data presented in Table 6 indicate that the two groups are indeed different. The results confirm the assumption made earlier that the students represent groups with different second language ability and corroborate the judgement of the

Table 6.

T-Test of Test Scores and Secondary 4 & 5 Achievement Grades for Groups 1 and 2 (n = 30+22)

Variables	t-value	No. Students	df	p
Multiple Choice Cloze	8.99	52	50	.001
Modified Cloze	8.39	52	50	.001
Dictation	5.18	52	50	.001
Proficiency Test	3.46	52	50	.001
Achievement Secondary 4				
Final Grade	3.45	50	48	.001
Teacher Grade	2.99	47	45	.005
Provincial Exam 412/422	2.68	47	45	.011
Achievement Secondary 5				
Final Grade	8.18	52	50	.001
Provincial Exam 522	7.37	52	50	.001

secondary 4 classroom teacher. The highest t-values (8.99 and 8.39) appear for the scores on the two cloze tests; these values show the greatest difference between the two groups. On the other hand, the lowest t-value (2.68) is found for the scores of the two groups on the two different examinations taken in secondary 4 (412 and 422). This is not surprising since the two examinations did not have the same level of difficulty.

Classroom Test Correlation Coefficients

With regards to the first research question, Pearson product-moment correlations were calculated using the scores attained on the three classroom tests for each group. The coefficients found in Tables 7 and 8 reveal the degree to which the tests are measuring similar language skills. Table 7 shows that there is only a significant relationship (.44) between the scores on the two cloze tests for students of low ability in ESL. Only 19 percent of the variance is common in both tests. Students with limited knowledge and skills in the second language do not perform similarly on the three global language tests. Scores for the intermediate students shown in Table 8 indicate relatively high correlation coefficients, especially for the two cloze tests (.86). More than 74 percent of the variance in scores on the modified cloze test

Table 7

Pearson Product-Moment Correlation Coefficients of Classroom Tests for Group 1 (n = 30)

	Multiple Choice Cloze	Modified Cloze
Modified Cloze	.444*	X
Dictation	.103 NS[a]	.238 NS[a]

[a]Not significant, $p > .05$.

*$p < .007$.

Table 8

Pearson Product-Moment Correlation Coefficients of Classroom Tests for Group 2 (n = 22)

	Multiple Choice Cloze	Modified Cloze
Modified Cloze	.855*	X
Dictation	.702*	.779*

*$p < .001$.

is accounted for by scores on the multiple-choice cloze test. The dictation test correlates less well with the two cloze tests, but the correlation coefficients (.70 and .78) are similar to correlations presented in other research reports (Irvine, Atai, & Oller, 1974; Krzyżanowski, 1976). On the whole, those who did well on one test did well on the other, and conversely, those who performed poorly on one test did relatively as poorly on the other tests. Based on these results it would seem that the three global language tests do not measure linguistic behavior in the same way for students of different language ability.

Pragmatic Validity

With respect to the second research question, scores on the three classroom tests were compared to those on the proficiency test and previous achievement in secondary 4 in order to determine the best predictors of ESL achievement in secondary 5.

Correlation. Pearson product-moment correlations were calculated for all measures using as the validating criterion the provincial 522 ESL examination, referred to as Exam 522. The correlation coefficients are shown in Table 9 for each group.

Table 9

Pearson Product-Moment Correlation Coefficients of Predictor Variables With the Provincial Exam 522 for Groups 1 & 2

Predictor Variables	Group 1	p	Group 2	p
Multiple Choice Cloze	.278	.068	.835	.001
Modified Cloze	.659	.001	.851	.001
Dictation	.424	.010	.825	.001
Proficiency Test	.427	.009	.805	.001
Final Grade 4	.575	.001	.855	.001
Teacher Grade 4	.269	.082	.663	.001
Provincial Exam 4	.631	.001	.779	.001

For the low ability students in Group 1 the scores on the modified cloze test correlated at .66 and the secondary 4 provincial examination scores correlated at .63 with final examination scores. The multiple-choice cloze test scores (.28) and teacher grades (.27) correlated least well. Among the three global language tests the modified cloze test is the best predictor of later ESL success.

The intermediate ability students in Group 2 produced correlations that are consistently higher than those for Group 1. The final grade in secondary 4 had the

highest correlation with examination scores (.86). It was followed very closely by correlations of the modified cloze test (.85), the multiple-choice cloze test (.84), and the dictation test (.83). Teacher grades in secondary 4 had the lowest correlation (.66) with examination scores. This was also the case for Group 1 (.27). Both the final achievement grade in secondary 4 and the scores on the three classroom tests are good predictors of ESL success for the intermediate students.

Regression. A second statistical technique is employed to provide further data for the second research question. Use of the subprogram, Regression, found in the SPSS, 2nd Edition, 1975, enables one to ascertain the amount of variance that the scores of each of the predictor variables, either grouped together in blocks or considered separately, account for in the scores on the criterion measure. The utility of regression analysis is that it reveals relationships that are otherwise not apparent through simple correlation. It is quite possible that relatively high correlations add nothing to the explanation of meaningful variance whereas seemingly low correlations could show important relationships (Borg, 1963; Hays, 1963). Two separate analyses, ordered-stepwise regression and multiple-stepwise regression, are performed

for each group (1) to determine whether the scores on the three classroom tests account for more variance in examination scores than do the other predictor variables, and (2) to determine which variables account for most of the variance in the criterion.

The results of the ordered-stepwise regression analysis for Group 1, in which blocks of predictor variables are entered into the equation in different order, are shown in Table 10. Six of the seven predictor variables are significant and account for 61 percent of the variance in the criterion measure, Exam 522. In the first order of entry, scores on the provincial ESL 412 exam, the proficiency test, and teacher grades in secondary 4 account for 42 percent of the variance. Exam scores in secondary 4 account for much more variance (40 percent) than the other two variables. Scores on the three classroom tests add 19 percent more variance to criterion scores.

In the second order of entry, the three classroom tests account for 52 percent of the variance whereas only an additional nine percent comes from the other three variables. Scores on the modified cloze test account for most of the variance (43 percent) among the three global language tests; multiple-choice cloze test scores are hardly significant. Scores on the modified cloze test account for slightly more variance (43 percent) than

Table 10

Ordered-Stepwise Regression of Predictor Variables With the Provincial Exam 522 for Group 1 (n = 30)

Predictor Variables	Multiple R	R^2	R^2 Change	Step Entered	p
1st Order: Proficiency Test and Achievement Secondary 4[a]					
Provincial Examination 412	.6317	.3991	.3991	1	.001
Proficiency Test	.6415	.4116	.0124	2	.001
Teacher Grade 4	.6444	.4152	.0036	3	.004
Modified Cloze	.7118	.5066	.0913	4	.002
Dictation	.7585	.5753	.0686	5	.001
Multiple Choice Cloze	.7789	.6068	.0314	6	.002
2nd Order: Classroom Tests[a]					
Modified Cloze	.6570	.4316	.4316	1	.001
Dictation	.7197	.5179	.0863	2	.001
Multiple Choice Cloze	.7208	.5196	.0017	3	.001
Provincial Examination 412	.7740	.5991	.0794	4	.001
Teacher Grade 4	.7788	.6065	.0074	5	.001
Proficiency Test	.7789	.6068	.0002	6	.002

[a]Final Grade 4 was not entered into the equation because of insufficient F level.

provincial exam scores in secondary 4 (40 percent). The three classroom tests, therefore, as a linear combination account for 10 percent more of the variance in the criterion measure than a combination of the other variables.

A multiple-stepwise regression analysis was then performed with the seven variables entered in succession according to their statistical importance in explaining variance in the criterion measure, Exam 522.

Table 11

Multiple-Stepwise Regression of Predictor Variables With the Provincial Exam 522 for Group 1 (n = 30)

Predictor Variables[a]	Multiple R	R^2	R^2 Change	Step Entered	p
Modified Cloze	.6570	.4316	.4316	1	.001
Dictation	.7197	.5179	.0863	2	.001
Provincial Examination 412	.7582	.5748	.0569	3	.001
Multiple Choice Cloze	.7740	.5991	.0242	4	.001
Teacher Grade 4	.7788	.6065	.0074	5	.001
Proficiency Test	.7789	.6068	.0002	6	.002

[a]Final Grade 4 was not entered into the equation because of insufficient F level.

The modified cloze test accounts for most of the variance (43 percent) in the criterion measure. It is perhaps not surprising since this measure was administered only three months before the provincial examination. It is noteworthy that the dictation test is the second best predictor of the criterion, even though it had the same correlation coefficient as the proficiency test (.42), which was administered at the same time as the modified cloze test. The provincial exam in secondary 4 is also adding some variance (six percent) to the criterion.

The same two kinds of regression analysis were applied to the scores of the predictor variables for the intermediate students in Group 2. The results of the ordered-stepwise regression analysis are shown in Table 12. The total variance accounted for by six predictor variables is 88 percent, much higher than that for Group 1 (61 percent). In the first order of entry with the proficiency test and previous achievement scores, the final grade in secondary 4 accounts for most of the variance (73 percent) among these variables. The three classroom tests provide an additional 10 percent of the variance to the criterion measure. When the three classroom tests are entered into the regression first, they account for 82 percent of the variance, with only six percent of the

Table 12

Ordered-Stepwise Regression of Predictor Variables With the Provincial Exam 522 for Group 2 (n = 22)

Predictor Variables	Multiple R	R^2	R^2 Change*	Step Entered
1st Order: Proficiency Test and Achievement Secondary 4[a]				
Final Grade 4	.8569	.7343	.7343	1
Proficiency Test	.8849	.7831	.0487	2
Teacher Grade 4	.8853	.7837	.0006	3
Multiple Choice Cloze	.9225	.8511	.0674	4
Dictation	.9382	.8802	.0290	5
Modified Cloze	.9397	.8830	.0028	6
2nd Order: Classroom Tests[b]				
Dictation	.8666	.7510	.7510	1
Multiple Choice Cloze	.9043	.8179	.0668	2
Modified Cloze	.9062	.8212	.0032	3
Provincial Exam 4	.9356	.8754	.0542	4
Proficiency Test	.9395	.8827	.0073	5
Teacher Grade 4	.9397	.8830	.0002	6

*$p < .001$.

[a]Provincial Exam 4 was not entered into the equation.

[b]Final Grade 4 was not entered into the equation because of insufficient F level.

variance explained by the other variables. Most of this variance (75 percent) is provided by scores on the dictation test, whereas scores on the two cloze tests add little information (seven percent). It should be noted that the Multiple R of the scores of the dictation test with the criterion was adjusted upwards to 87 percent from a product-moment correlation of 83 percent. As was the case for Group 1, the three classroom tests as a linear combination account for more variance (four percent) in the criterion measure than a combination of the other variables.

Data of the multiple-stepwise regression analysis for all variables (shown in Table 13) indicate that the dictation test is the single best predictor of ESL achievement for the intermediate students. Scores on the dictation test account for 75 percent of the variance, with another eight percent added when final achievement grades in secondary 4 are entered. The four other measures add very little important information. Scores on the modified cloze test contribute no new information (.0028) to the other variables for the intermediate students, although scores on this test accounted for most of the variance (43 percent) in the criterion measure for low level students.

Table 13

Multiple-Stepwise Regression of Predictor Variables With the Provincial Exam 522 for Group 2 (n = 22)

Predictor Variables[a]	Multiple R	R^2	R^2 Change*	Step Entered
Dictation	.8666	.7510	.7510	1
Final Grade 4	.9113	.8306	.0795	2
Multiple Choice Cloze	.9250	.8557	.0250	3
Provincial Exam 422	.9356	.8754	.0197	4
Proficiency Test	.9382	.8802	.0048	5
Modified Cloze	.9397	.8830	.0028	6

[a]Teacher Grade was not entered into the equation.

*$p < .001$.

Teacher Grades

To provide data for the third research question regarding the usefulness of teacher grades in assessing ESL achievement, scores of the three achievement measures in secondary 4 and one classroom test, the dictation, were correlated with final ESL grades at the end of secondary 5. Scores on the two cloze tests and the proficiency test are not included in the data because they were added in the teacher grade in secondary 5, which is part of the final ESL grade (known as Fgrade 5).

Correlation. The results in Table 14 show that, of the four predictor variables, the final grade in secondary 4 has the highest correlation with the criterion measure (Fgrade 5), .77 percent for group 1 and 89 percent for Group 2. The provincial ESL examination in secondary 4 has the second highest correlation for each group.

Table 14

Pearson Product-Moment Correlation Coefficients of Selected Predictor Variables With the Final Grade 5 for Groups 1 & 2

Predictor Variables	Group 1	p	Group 2	p
Dictation	.366	.023	.772	.001
Final Grade 4	.767	.001	.894	.001
Teacher Grade 4	.470	.006	.756	.001
Provincial Exam 4	.753	.001	.779	.001

If one compares the results presented here with the coefficients for the first measure (Exam 522), the correlations of the three previous achievement scores with the second criterion (Fgrade 5) are much higher for Group 1 and equal to or slightly higher for Group 2 than previous

correlations. On the other hand, correlations of the scores of the dictation test with Fgrade 5 for each group are somewhat lower than previous correlations with Exam 522. It would seem that previous achievement measures are effective predictors of later success in ESL for each group in this order: (1) final grades, and (2) provincial exam scores.

Regression. Ordered-simple regression and multiple-stepwise regression operations were performed in order to determine how much variance is explained in the criterion measure (Fgrade 5) in which teacher grades are included. The data for Group 1, presented in Table 15, show that three variables account for 65 percent of the variance in the criterion measure. Whereas dictation scores contributed

Table 15

Ordered-Simple Regression of Selected Predictor Variables With the Final Grade 5 for Group 1 (n = 30)

Predictor Variables[a]	Multiple R	R^2	R^2 Change	Step Entered	p
Dictation	.3739	.1398	.1398	1	.050
Teacher Grade 4	.5552	.3083	.1684	2	.001
Provincial Examination 412	.8088	.6542	.3459		.001

[a]Final Grade 4 was not entered into the equation because of insufficient F level.

small but meaningful variance (nine percent) to the criterion Exam 522, they contribute very little variance to the criterion Fgrade 5.

The data shown in Table 16 of the multiple-stepwise regression indicate that the major portion of the variance in this criterion measure (59 percent) is accounted for by final grades in secondary 4. This variable and dictation account for slightly more variance (62 percent) in

Table 16

Multiple-Stepwise Regression of Selected Predictor Variables With the Final Grade 5 for Group 1 (n = 30)

Predictor Variables[a]	Multiple R	R^2	R^2 Change	Step Entered	p
Final Grade 4	.7672	.5886	.5886	1	.001
Dictation	.7890	.6225	.0339	2	.001
Provincial Examination 412	.8088	.6542	.0316	3	.001

[a]Teacher Grade 4 was not entered into the equation because of insufficient F level.

Fgrade 5 than do the seven predictor variables account for in Exam 522 (61 percent as shown in Table 11).

The data for Group 2, presented in Tables 17 and 18, indicate that scores on the three predictor variables account for 86 percent of the variance of Fgrade 5.

Table 17

Ordered-Simple Regression of Selected Predictor Variables With the Final Grade 5 for Group 2 (n = 22)

Predictor Variables[a]	Multiple R	R^2	R^2 Change*	Step Entered
Dictation	.8397	.7051	.7051	1
Provincial Exam 422	.8900	.7921	.0869	2
Teacher Grade 4	.9278	.8608	.0687	

[a]Final Grade 4 was not entered into the equation because of insufficient F level.

*$p < .001$.

Table 18

Multiple-Stepwise Regression of Selected Predictor Variables With the Final Grade 5 for Group 2 (n = 22)

Predictor Variables[a]	Multiple R	R^2	R^2 Change*	Step Entered
Final Grade 4	.9043	.8178	.8178	1
Dictation	.9274	.8602	.0423	2
Teacher Grade 4	.9278	.8608	.0006	3

[a]Provincial Exam 422 was not entered into the equation because of insufficient F level.

*$p < .001$.

When dictation scores are entered first, they account for 71 percent of the variance in Fgrade 5, whereas 82 percent is accounted for by the final grade in secondary 5 when this variable is entered first. A Multiple R of .84 for the dictation test was adjusted upwards from a product-moment correlation of .77.

As was the case for Group 1, final grades in secondary 4 and dictation scores account for more variance in Fgrade 5 (86 percent) than these same two variables account for in Exam 522 (83 percent as seen in Table 13). The results for both low and intermediate ability ESL students indicate that teacher grades add small but meaningful information to the assessment of ESL achievement when they are added to examination scores.

FOOTNOTES

[1]Formula Kuder-Richardson 21:

$$r_{xx} = \frac{K\sigma_x^2 - \bar{X}(K - \bar{X})}{\sigma_x^2 (K - 1)}$$

r_{xx} = reliability of the whole test

K = number of items in the test

σ_x^2 = variance of the scores

$\bar{X}$ = mean of the scores

CHAPTER V

INTERPRETATION: DISCUSSION AND CONCLUSIONS

This study investigated the pragmatic validity of three classroom tests, two modified cloze tests and a dictation test, with an approach using both correlation and regression analysis. The sample consisted of a mixed group of ESL learners in high school who had previously been assigned to two separate ability subgroups. The principal concern of the study was whether the classroom tests were good measures of ESL attainment for learners of differing language ability. The two main questions about these tests were, firstly, do they assess global knowledge of the second language, and secondly, do they effectively predict ESL final test achievement for both low and intermediate level high school students? In this investigation of student ESL school and examination performance, another concern was whether teacher grades in secondary 4 and 5 are useful in assessing ESL final achievement.

The first research question investigated whether the three classroom tests, the two modified cloze tests and the dictation test, measure comprehensive language

skills and tap the underlying second-language competence of high school ESL learners. The second research question asked whether the three classroom tests provide more information about predicting ESL test achievement - for students at different levels of ability - than either an experimental proficiency test or previous ESL school and examination performance. The third concern was whether high school teacher grades provide useful information in the final assessment of ESL achievement when they are added to provincial test scores. The independent variables in the study, in addition to the scores of the three classroom tests, included scores on a proficiency test, the teacher ESL class grade in secondary 4, scores on two provincial ESL achievement examinations in secondary 4, and the final achievement grade in secondary 4 which was an average of the previous two measures.

Summary statistics shown in Tables 4 and 5 indicated that the dictation test and the proficiency test had the highest reliability and variability, although it was noted that reliability estimates for the dictation might have been overestimated using K-R 21. The two modified cloze tests showed similar moderate reliability estimates (.70 to .76) and variability (5.9 to 6.9) for both groups, but the intermediate level students scored more than 30

percent better than the low level students on both tests. As shown in Table 6, the T-test of significant difference in scores indicated that the three classroom tests had the highest t-values among all the independent variables, and consequently they effectively discriminated between two different groups.

Classroom Test Correlations

The correlation coefficients presented in Tables 7 and 8 for the two groups showed that the three classroom tests did not measure linguistic knowledge and skills in the same way for low and intermediate level students. Therefore the data do not support the hypothesis that the three classroom tests assess global knowledge of the second language at different levels of ability.

The correlations of test scores for the intermediate students ranged between .70 and .86, indicating a very high relationship in what the tests were measuring. These correlations were similar to those reported in previous studies (Oller & Conrad, 1971; Oller, 1972a, 1972b; Irvine, Atai & Oller, 1974; Krzyzanowski, 1976) which used samples of adult, college or high school students. The three classroom tests seem to be good instruments for measuring the comprehensive language knowledge and skills

of students who have attained a certain competence in the second language.

On the other hand, the correlations of test scores for the less competent students were quite different. The correlation of cloze scores was only moderate (.44). Moreover, the correlations of dictation scores with the scores of the two modified cloze tests were not statistically significant.

Oller (1979) argued that low correlations between test scores is relatively uninformative and does not indicate that the two tests measure different skills. He reasoned that low correlations might be the result of many factors, including the lack of test reliability or validity. In order for two tests to correlate highly they must both possess high separate reliabilities. All three classroom tests had good to high reliabilities for the low level students (.70 to .89) and for the intermediate students (.70 to .94). So the low correlations cannot be attributed to the lack of test reliability. Neither can it be claimed that the classroom tests were not valid measures of ESL attainment. It was shown (see Table 9) that the modified cloze test and the dictation test correlated moderately with later achievement (.66 and .42).

The absence of any marked relationship among the three classroom tests must be explained in other ways. The fact the ESL students were reading at a "frustrational" level as termed by Anderson (cited in Oller, 1979, p. 353) is an important consideration.[1] The data showed no relationship between the fairly developed ability of low level students to process the spoken language (mean score on the dictation test was 70 percent) and their relative inability to cope well with the written language in two modified cloze tests which provided alternate responses or word clues. Below a certain level of competence in the second language, weak ESL students perform differently on tests like the modified cloze procedure and dictation.

Anastasi (1976) stated that a particular test may produce different results when given to groups of individuals who differ in some significant way. The same test may measure different skills or functions with students who are not the same age, who are not at the same educational level, or who possess different abilities. The findings here corroborate this hypothesis: ability level is an important factor to consider in interpreting test results. Although all three classroom tests effectively discriminated between low level and intermediate students, the tests did not measure the language abilities of the students in the same way.

Pragmatic Validity

The correlation coefficients shown in Table 9 indicated which independent variables in the study for each group of students had the highest correlations with ESL test achievement at the end of secondary 5. For the low level students, the modified cloze test (.66), the provincial examination in secondary 4 (.63), and the final achievement grade in secondary 5 (.58) were the best predictors of later test performance. Teacher class grades in secondary 4 and the multiple-choice cloze scores had a weak relationship to final examination scores. Both the dictation test (.42) and the proficiency test (.43) showed a moderate correlation.

For the intermediate students, the final achievement grade in secondary 4 (.86), the modified cloze test (.85), the multiple-choice cloze test (.84), and the dictation test (.83) were equally good predictors of achievement examination performance. The teacher class grade in secondary 4 (.66) showed the weakest correlation. Two variables that were among the best predictors of examination performance for both low level and intermediate students were the modified cloze test and previous final ESL achievement grades.

The second research question examined whether the scores of the three classroom tests account for more variance in ESL examination scores for each group than either scores of the proficiency test and both previous class grades and provincial exam scores. When these last four variables were entered first as a group in the stepwise regression for low level students (see Table 10), they accounted for 42 percent of the variance in the criterion measure. When the classroom tests were entered second as a group, they added 19 percent more variance to the criterion measure to produce a total of 61 percent. When the order of entry was reversed, the classroom tests accounted for 52 percent of variance with only 9 percent added by the other four variables. When all seven variables were entered into a multiple-stepwise regression with the criterion measure (Exam 5), the modified cloze test having the highest correlation (.66) accounted for more variance than any other variable (.43). As shown in Table 11, the dictation test was the second best predictor of the criterion measure adding 9 percemt to the explained variance in scores. Although the provincial exam in secondary 4 correlated well with the criterion (.63), it did not add a great deal of meaningful variance to the criterion scores (6 percent).

For the intermediate students (see Table 12), the total variance accounted for in the ordered-stepwise regression of blocks of variables with the criterion measure (Exam 5) was 88 percent. When the scores on the proficiency test and previous class grades and provincial exam scores were entered first, they accounted for 78 percent of the variance in the criterion measure. Entered second, the three classroom tests provided an additional 10 percent of the variance in criterion scores. With the order of entry of variables reversed, the three classroom tests explained 82 percent of the variance in the criterion measure. The other four variables then added only 6 percent to the explained variance in scores. In the multiple-stepwise regression (see Table 13), the dictation test accounted for more variance (.75) in the criterion measure than any other variable. With a simple correlation of .86, the final achievement grade in secondary 4 was the second best predictor of the criterion measure adding 8 percent to the explained variance. No other variable added meaningful variance to the criterion measure.

The results of the regression analysis partially support the hypothesis that the classroom tests provide more information about a student's level of ESL attainment than either a proficiency test or previous ESL

achievement. For low level students, the modified cloze test and the dictation - in that order - accounted for more variance in achievement exam scores than the other variables. For intermediate students, on the other hand, the dictation test and the multiple-choice cloze test - in that order - accounted for more variance in achievement exam scores. Although the two cloze tests did not provide similar information about the two ability levels, the dictation test accounted for meaningful variance in achievement exam scores for both ability levels.

With respect to the low level students, the closer time interval of the modified cloze test and the dictation with the criterion measure as compared to the other variables could be an explanation for the high correlations. However, the discrete-point proficiency test was administered at the same time as the modified cloze test and it correlated with the criterion measure as well as the dictation test (.43 and .42 in Table 9), yet it added no meaningful information in predicting exam scores. It can be claimed that the modified cloze test and the dictation test measured aspects of language learning that took place during the ESL course. Since the scores of the two classroom tests themselves did not correlate significantly, the tests measured aspects of different skills or knowledge and

provided different information in the assessment of ESL attainment. Both the modified cloze test and the dictation test provided more information about ESL attainment than either the proficiency test or previous classroom and exam ESL scores.

In the case of the intermediate students, the dictation test, a global measure of not only listening comprehension but also perhaps of language knowledge, provided the most information about ESL exam performance. The high correlations of all the variables with the criterion measure (exam scores) demonstrated that the intermediate students might possess a certain underlying linguistic competence that is tapped by different types of classroom and exam measures. The fact that the dictation test is the best predictor of ESL achievement on the provincial examination shows its high predictive validity.

Teacher Grades

As shown in Table 14, scores on the dictation test and previous ESL class grades, provincial exam scores, and combined final ESL grades were correlated with combined final ESL grades in secondary 5 (Fgrade 5). Final ESL grades correlated the highest among the independent variables with this criterion measure (.77 for low level students and .89 for intermediate students). With the

addition of teacher class grades in the second criterion (Fgrade 5), the correlations with previous final achievement grades were increased.

As shown in the regression data of Table 16, final ESL grades in secondary 4 accounted for 59 percent of the variance in criterion scores (Fgrade 5) for low level students. Only an additional six percent of the variance was provided by the dictation test and the provincial exam scores. In the previous discussion of the second research question, it was shown that the modified cloze test and the dictation test had accounted for 52 percent of the variance of exam scores alone (see Table 11). With this second criterion measure (Fgrade 5) which includes teacher class grades, more variance (7 percent) in student ESL performance was accounted for by previous final grades.

For intermediate students, Table 17 showed that the dictation test accounted for 70 percent of the variance in the criterion measure (Fgrade 5). As shown in Table 18, however, the best predictor of the criterion measure was final ESL grades in secondary 4 with a R square of .82. Almost no meaningful variance was added to this criterion by the dictation test when it was entered second, although the dictation test had previously accounted for 75 percent

of the variance in the other criterion measure: exam scores (see Table 13). Again with this second criterion measure (Fgrade 5), more variance (7 percent) in student final ESL performance was accounted for by the previous final grades in secondary 4.

The data support the hypothesis that teacher grades in secondary 4 and 5 add small but useful information to examination scores in the assessment of ESL final achievement for different ability levels. The findings indicate that ESL achievement in secondary 4, including class grades and provincial exam scores, is an excellent predictor of later performance in ESL at the end of secondary 5.

Summary

(1) The hypothesis that procedures like the modified cloze and dictation assess the global language knowledge of high school students of different levels of ESL ability was not supported by the data in this study. Students at the low and intermediate levels of ability in ESL did not perform similarly on the three so-called global language measures under investigation. The main conclusion drawn from the low or insignificant correlations among the classroom tests taken by low level students is that these tests do not measure similar skills for students below a certain level of competence in the second language.

(2) The hypothesis that procedures like the modified cloze and dictation more effectively measure ESL attainment - as measured by the final ESL examination - than a proficiency test or previous ESL grades for high school students of different levels of ESL ability was only partially supported by the data. Different classroom tests were effective for each group. The dictation test was the single most effective measure of ESL attainment for intermediate students, whereas both the modified cloze test and the dictation provided important information in assessing low level students. The multiple-choice cloze test, on the other hand, a time-consuming instrument to construct, did not provide any more information in the assessment of ESL knowledge than the two other classroom tests. The results underline the difficulty in producing good multiple-choice cloze tests. The use of both a modified cloze test and a dictation test is recommended for ranking ESL students of different ability at the beginning of instruction.

(3) Finally, the hypothesis that teacher grades provide useful information in the assessment of ESL performance for students of different ability when the grades are added to examination results was supported by the data. Neither secondary 4 teacher grades nor provincial exam scores alone are as effective as the final combined grade in predicting

ESL success in secondary 5. In fact, final ESL grades in secondary 4 more effectively predict final grade results in secondary 5 than they predict ESL exam scores. This provides support for the practice of the Ministry of Education in reporting final combined grades in ESL rather than provincial ESL exam scores alone.

Suggestions for Future Study

One must be careful in attempting to apply the findings here to different groups of ESL students. A weakness in the research design mentioned in Chapter I prevents one from overgeneralizing to other situations. The results of the study regarding the second research question would be more conclusive if the three classroom tests and the proficiency test had been administered to the students at the beginning of instruction in September. Furthermore, more data are needed and further statistical analysis is required in order to confirm the findings. It would be necessary to reproduce the correlational study on a much larger scale with hundreds of students in several different school commissions. The margin of statistical error would be reduced with larger samples. It is the case that large variations are often found in coefficients of correlation based on small samples. For example, a true

correlation of .90 for a sample of 30 can vary between .80 and .95 at the 95 percent confidence interval (Ebel, 1972). Also there is a considerable fluctuation in the multiple regression data in small samples. A sample size of 40 can yield fluctuations as high as .25 in the regression weights (Kerlinger & Pedhazur, 1975). Thus there is a need for more studies at the secondary level which deal not only with the correlation of global language tests, but also with their correlation with an external criterion that has been validated.

It is recommended that studies include groups of students of varying ability at the three levels of the second cycle, including secondary 3, 4 and 5. It might very well be shown that simple tests such as the dictation and the modified cloze test can be used in place of more complicated tests.

FOOTNOTES

[1]In interpreting cloze scores, Anderson suggested that students scoring below 44 percent fall into a frustrational level of reading and that the passage from which the cloze test is chosen is not appropriate for instructional use (cited in Oller, 1979).

REFERENCES

Aitken, K.G. Using cloze procedure as an overall language proficiency test. TESOL Quarterly, 1977, 11, No. 1, 59-68.

Alderson, J.C. The cloze procedure and proficiency in English as a foreign language. TESOL Quarterly, 1979, 13, No. 2, 219-227.

Allen, J.P.B., & Davies, A. Testing and experimental methods. London: Oxford University Press, 1977.

Allen, H.B., & Campbell, R.N. Teaching English as a second language. New York: McGraw Hill, 1972.

Amos, J.R., Brown, F.L., & Mink, O.G. Statistical concepts: A basic program. New York: Harper & Row, 1965.

Anastasi, A. Psychological testing, 4th edition. New York: MacMillan, 1976.

Anderson, B.F. The psychology experiment. Belmont: Wadsworth, 1966.

Ary, D., Jacobs, L., & Razavich, A. Introduction to research in education. New York: Holt, Rinehart & Winston, 1979.

Banko, R. Predicting proficiency of French-speaking teachers. Unpublished MA thesis, Concordia University, in progress.

Barrett, T.C. Cognitive and affective dimensions of reading comprehension. In Innovations and changes in reading instruction. The National Society for Education, University of Chicago Press, LXVII, Part II, 1968.

Blalock, H.M. Social statistics, 2nd edition, New York: McGraw Hill, 1972.

Bonkowski, F.J., Evans, S., & McKrae, A. Experimental proficiency test. Mimeo, Concordia University, 1977.

Borg, W.R. Educational research: An introduction. New York: David McKay Co., 1963.

Bormuth, J.R. The cloze readability procedure. Elementary Teacher, 1968, 45, 429-436.

Campbell, D.T., & Stanley, J.C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally & Co., 1966.

Carroll, J.B. Fundamental considerations in testing for English language proficiency for foreign students. In H.B. Allen & R.N. Campbell (Ed.), Teaching English as a second language. New York: McGraw Hill, 1972.

Carroll, J.B., Carton, A.S., & Wilds, C.P. An investigation of cloze items in the measurement of achievement in foreign languages (College Entrance Examination Board Research and Development Report). Laboratory for Research in Instruction, Harvard University, 1959.

Clark, J.L. Foreign language testing: Theory and practice. Philadelphia: The Center for Curriculum Development, 1972.

Cohen, A.D. Testing language ability in the classroom. Rowley, Mass.: Newbury House, 1980.

Cranney, A.G. The construction of two types of cloze reading tests for college students. Journal of Reading Behavior, 1973, 5, 60-64.

Crocker, A.C. Statistics for the teacher. Middlesex: Penguin Books, 1971.

Darnell, D.K. Clozentropy: A procedure for testing English language proficiency of foreign students. Speech Monographs, 1970, 37, No. 1, 36-46.

Davies, A. Two tests of speeded reading. In R. Jones & B. Spolsky (Eds.), Testing language proficiency. Virginia: Center for Applied Linguistics, 1975.

Davies, A., & Moller, A. Course in English language testing. Mimeo, Concordia University, 1973.

Davies, A. Language testing symposium. London: Oxford University Press, 1968.

Peterson, J., Paradis, E., & Peter, N. Revalidation of the cloze procedure as a measure of the instructional level for high school students. In Diversity in mature reading (22nd yearbook of the National Reading Conference). North Carolina: National Reading Conference, 1973.

Ebel, R.L. Essentials of educational measurement. Englewood Cliffs, New Jersey: Prentice Hall, 1972.

Farhady, H. The disjunctive fallacy between discrete-point and integrative tests. TESOL Quarterly, 1979, 13, No. 3, 347-357.

Ferguson, G.A. Statistical analysis in psychology and education. New York: McGraw Hill, 1971.

Fry, E. A readability formula that saves time. Journal of Reading, 1968, April, 513-516; 575-576.

Gellman, E.S. Statistics for teachers. New York: Harper & Row, 1973.

Green, J.A. Teacher-made tests. 2nd edition. New York: Harper & Row, 1975.

Greene, F. Modification of the cloze procedure and changes in reading test performance. Journal of Educational Measurement, 1965, 2, 213-217.

Guilford, J.P., & Fruchter, B. Fundamental statistics in psychology and education. New York: McGraw Hill, 1973.

Hanzeli, V.E. The effectiveness of cloze tests in measuring the competence of students of French in an academic setting. French Review, 1977, L, No. 6, 865-874.

Harris, D. Current trends: Integrative tests. Workshop at 5th SPEAQ Convention, 1977.

Harris, D. Testing English as a second language. New York: McGraw Hill, 1969.

Hays, W. Statistics for psychologists. New York: Holt, Rinehart & Winston, 1963.

Heaton, J.B. Writing English language tests. London: Longman, 1975.

Hinofotis, F.B. Cloze as an alternative method of ESL placement and proficiency testing. In J.W. Oller & K. Perkins (Eds.), Research in language testing. Rowley, Mass.: Newbury House, 1979.

Hinofotis, F.B. An investigation of the concurrent validity of cloze testing as a measure of overall proficiency in English as a second language. Unpublished doctoral dissertation, Southern Illinois University at Carbondale, 1976.

Hinofotis, F.B., & Snow, B.G. An alternative cloze testing procedure: Multiple-choice format. In J.W. Oller & K. Perkins (Eds.), Research in language testing. Rowley, Mass.: Newbury House, 1979.

Hisama, K. A new direction in measuring proficiency in English as a second language. April, 1977. ED 150 198.

Horst, P. Psychological measurement and prediction. Belmont, California: Wadsworth, 1966.

Ingram, E. Attainment and diagnostic testing. In A. Davies (Ed.), Language testing symposium. London: Oxford University Press, 1968.

Irvine, P., Atai, P., & Oller, J. Cloze, dictation and the test of English as a second language. Language Learning, 1974, 24, 245-252.

Isaac, S., & Michael, W. Handbook in research and evaluation. San Diego, California: Robert R. Knapp, 1979.

Jones, R., & Spolsky, B. Testing language proficiency. Virginia: Center for Applied Linguistics, 1975.

Jongsma, E.A. The cloze procedure: A survey of research. Occasional Papers in Reading, Indiana University, 1971.

Jonz, J. Improving on the basic egg: The M-C cloze. Language Learning, 1976, 26, No. 2, 255-265.

Kerlinger, F.N., & Pedhazur, E. Multiple regression in behavioral research. New York: Holt, Rinehart & Winston, 1973.

Krohn, R. English sentence structure. Ann Arbor: The University of Michigan Press, 1972.

Krzyzanowski, H. Cloze tests as indicators of general language proficiency. Studia Anglica Posnaniensia, 1976, 7, 29-43.

Kurtz, A., & Mayo, S.T. Statistical methods in education and psychology. New York: Springer-Verlag, 1979.

Lado, R. Language testing: A scientific approach. New York: McGraw Hill, 1964.

Lother, H.E. The effectiveness of selected variables for predicting grade IX and grade X achievement as measured by final marks in these grades. Unpublished M. Ed. thesis, University of Manitoba, 1967.

Mains, E. Dictation as a test of second language proficiency. Unpublished MA thesis, Concordia University, 1980.

Morissette, D. Les examens de rendement scolaire. Québec: Les Presses de l'Université Laval, 1979.

Mellgren, L., & Walker, M. (Eds.). New Horizons in English. Reading, Mass.: Addison-Wesley, 1974.

Nunnally, J.C. Educational measurement and evaluation. New York: McGraw Hill, 1964.

Oller, J.W. Assessing competence in ESL: Reading. In L. Palmer & B. Spolsky (Eds.), Papers on language testing: 1967-1975. Washington, D.C.: TESOL, 1975.

Oller, J.W. Dictation as test of ESL proficiency. In H.B. Allen & R.N. Campbell (Eds.), Teaching English as a second language. New York: McGraw Hill, 1972a.

Oller, J.W. Cloze tests of second language proficiency and what they measure. Language Learning, 1973a, 23, 105-118.

Oller, J.W. Discrete-point tests versus tests of integrative skills. In J.W. Oller & J. Richards (Eds.), Focus on the learner. Rowley, Mass.: Newbury House, 1973b.

Oller, J.W. Language testing. In R. Wardaugh & H.D. Brown (Eds.), Survey of applied linguistics. Ann Arbor: University of Michigan Press, 1977.

Oller, J.W. Language tests at school. London: Longman, 1979.

Oller, J.W. Language testing today. English Teaching Forum, 1976, 14, 22-27.

Oller, J.W. A program for language testing research. Language Learning, 1976, Special Issue 4, 141-165.

Oller, J.W. Scoring methods and difficulty levels for cloze tests of proficiency in English as a second language. Modern Language Journal, 1972, 56, No. 3, 151-158.

Oller, J.W., & Conrad, C. The cloze technique and ESL proficiency. Language Learning, 1971, 21, No. 2, 183-195.

Oller, J.W., & Inal, N. A cloze test of English proficiency. In S. Palmer & B. Spolsky (Eds.), Papers on language testing. Washington, D.C.: TESOL, 1975.

Oller, J.W., & Perkins, K. (Eds.). Language in education: Testing the tests. Rowley, Mass.: Newbury House, 1978.

Oller, J.W., & Perkins, K. (Eds.). Research in language testing. Rowley, Mass.: Newbury House, 1979.

Oller, J.W., & Richards, J. Focus on the learner: Pragmatic perspectives for the language teacher. Rowley, Mass.: Newbury House, 1973.

Oller, J.W., & Streiff, V. Dictation: A test of grammar based expectations. In R. Jones & B. Spolsky (Eds.), Testing language proficiency. Arlington, Va.: Center for Applied Linguistics, 1975.

Ozete, O. The cloze procedure: A modification. Foreign Language Annals, 1977, 10, No. 5, 565-568.

Palmer, L., & Spolsky, B. Papers on language testing: 1967-1974. Washington, D.C.: TESOL, 1975.

Peterson, C.R., & Cartier, F.A. Some theoretical problems and practical solutions in proficiency test validity. In R. Jones & B. Spolsky (Eds.), Testing language proficiency. Virginia: Center for Applied Linguistics, 1975.

Pikulski, J.J., & Pikulski, E.C. Cloze, maze and teacher judgment. The Reading Teacher, 1977, 30, No. 7, 766-770.

Porter, D. Modified cloze procedure: A more valid reading comprehension test. English Language Teaching, 1976, 30, 151-155.

Scholastic Scope Magazine. New York: Scholastic Magazine, 1976.

Siegel, S. Nonparametric statistics for the behavioral sciences. New York: McGraw Hill, 1956.

Spence, J.T. Elementary statistics. Englewood Cliffs, New Jersey: Prentice Hall, 1976.

Spolsky, B. Language testing: The problem of validation. In L. Palmer & B. Spolsky (Eds.), Papers on language testing: 1967-1974.

Spolsky, B., Siguard, B., Sako, M., Walker, E., & Aterburn, C. Preliminary studies in the development of techniques for testing overall 2nd language proficiency. Language Learning, 1968, Special Issue 3, 79-98.

SRA Reading Laboratory, Box IIA. Don Mills, Ontario: Science Research Associates, 1969.

Streiff, V. Relationships among oral and written cloze scores and achievement test scores in a bilingual setting. In J.W. Oller & K. Perkins (Eds.), Language in education: Testing the tests. Rowley, Mass.: Newbury House, 1979.

Stump, T.A. Cloze and dictation tasks as predictors of intelligence and achievement scores. In J.W. Oller & K. Perkins (Eds.), Language in education: Testing the tests. Rowley, Mass.: Newbury House, 1979.

Tate, M.W. Statistics in education and psychology: A first course. New York: Collier-MacMillan, 1968.

Upshur, J. Discussion of "A program for language testing research" by J.W. Oller. In H.D. Brown (Ed.), Papers in second language learning: Proceedings of the 6th annual conference on applied linguistics at the University of Michigan. Language Learning, Special Issue 4, 1976, 167-174.

Wardhaugh, R. Workshop on testing, OSIE Conference, October, 1977.

Wardhaugh, R., & Brown, H.D. (Eds.). Survey of applied linguistics. Ann Arbor: University of Michigan Press, 1977.

Wolfe, G.K., & Williams, C.T. Elements of research: A guide for writers. Chicago: Roosevelt University, 1979.

APPENDICES

Appendix

APPENDIX 1

A BATTLE WITH THE SHARKS

A True Story by Horace S. Mazet

It was December 14, 1948. A Danish freighter was going through calm water about 10 miles from Cuba. Two young members of the crew were on a deck at the back of the ship. Tony La Tona, 13, was from California. Bent Jeppsen, 14, was from Denmark. The boys had become __1__ on shipboard. Now __2__ were relaxing after cleaning __3__ after a meal. __4__, Bent lost his balance and __5__ into the sea. Knowing __6__ Bent was a poor __7__, Tony looked wildly around. __8__ there was no one __9__ sight.

Tony grabbed the __10__ life ring and threw __11__ toward Bent. But the __12__ fell short. So Tony __13__ himself overboard, swam to __14__ life ring, and towed __15__ over to Bent. When Bent __16__ it, the two __17__ looked toward their ship. __18__ had sailed on! No __19__ on board had seen __20__ had happened.

In a __21__ room back in San Francisco __22__ Tony's basketball, football, and __23__. Tony had left them __24__ he had walked out __25__ months before. His mother __26__ not earn enough money __27__ move to a larger __28__ or to an apartment. Tony __29__ the place was too __30__ for his mother, his __31__ sister, and him. So __32__ had shipped out to sea. "__33__ he went to sea", __34__ mother said, "I'm glad he __35__ to swim at the YMCA".

__36__ Tony was telling Bent, "__37__ guess we'll have to __38__ for Cuba". His friend __39__. "Okay". "It's a long swim", Tony __40__. "Take your shoes off. __41__ can swim better without __42__".

Tony told Bent to __43__ in the life ring __44__ paddle with his hands. Tony __45__ onto the ring and __46__. Now and then the __47__ boys changed places. After __48__ hours, some sharks appeared. __49__, one of them gashed __50__ left foot. Bent yelled that he had been bitten. The boys kicked and managed to drive the sharks away for the moment ...

A BATTLE WITH SHARKS, A true story by Horace S. Mazet, Scope Magazine, Vol. 25, No. 5, October 21, 1976.

INSTRUCTIONS: In the following passage, some of the words have been left out. First, read over the entire passage and try to understand what it is about. Then, try to fill in the blanks. It takes exactly one word to fill in each blank. Contractions like can't, or words written with a hyphen like well-being count as single words. Consider the following example:

The __1__ barked furiously, and chased __2__ cat up the tree.

1. a. chicken
   b. squirrel
   c. dog
   d. mouse
   e. bird

2. a. its
   b. the
   c. his
   d. one
   e. none

The correct answers for 1 and 2 are dog, choice c., and the, choice b.

1. a. bad
   b. honest
   c. good
   d. old
   e. two

2. a. they
   b. he
   c. their
   d. friends
   e. them

3. a. around
   b. in
   c. on
   d. up
   e. near

4. a. Lately
   b. Soon
   c. Finally
   d. Now
   e. Suddenly

5. a. falls
   b. fell
   c. felt
   d. fallen
   e. fells

6. a. that
   b. who
   c. where
   d. what
   e. why

7. a. walker
   b. sleeper
   c. swimmer
   d. runner
   e. sport

8. a. Soon
   b. But
   c. And
   d. Later
   e. For

9. a. out
   b. near
   c. on
   d. of
   e. in

10. a. farther
b. near
c. further
d. nearest
e. farthest

11. a. its
b. them
c. him
d. it
e. her

12. a. ring
b. life
c. rope
d. ladder
e. boat

13. a. throws
b. fell
c. threw
d. thrown
e. saw

14. a. this
b. one
c. his
d. the
e. her

15. a. it
b. her
c. them
d. its
e. his

16. a. sat
b. saw
c. pushed
d. threw
e. grabbed

17. a. them
b. boys
c. rings
d. boats
e. fish

18. a. They
b. Her
c. It
d. Their
e. One

19. a. one
b. person
c. none
d. ones
e. friend

20. a. when
b. why
c. where
d. what
e. that

21. a. old
b. hospital
c. house
d. hotel
e. apartment

22. a. while
b. when
c. where
d. was
e. were

23. a. shirts
b. clothes
c. hats
d. socks
e. pants

24. a. that
b. what
c. when
d. while
e. why

25. a. several
b. one
c. few
d. no
e. much

26. a. has
b. is
c. had
d. was
e. did

27. a. to
b. for
c. in
d. on
e. back

28. a. office
b. building
c. house
d. room
e. hotel

29. a. finds
b. felt
c. feels
d. fell
e. feeling

30. a. expensive
b. old
c. small
d. large
e. nice

31. a. two
b. 9 years
c. 10-year-old
d. eleven
e. old

32. a. he
b. his
c. they
d. she
e. her

33. a. Where
b. Because
c. For
d. While
e. Since

34.	a. no	35.	a. started	36.	a. When
	b. her		b. learned		b. Now
	c. one		c. studied		c. Soon
	d. his		d. taught		d. Recently
	e. he		e. followed		e. Lately
37.	a. We	38.	a. swim	39.	a. needed
	b. They		b. walk		b. heard
	c. I		c. start		c. awoke
	d. He		d. look		d. nodded
	e. One		e. run		e. listened
40.	a. saw	41.	a. They	42.	a. me
	b. say		b. You		b. those
	c. saying		c. He		c. this
	d. says		d. It		d. them
	e. said		e. She		e. it
43.	a. grab	44.	a. and	45.	a. climbed
	b. climb		b. to		b. swam
	c. swim		c. not		c. held
	d. sit		d. but		d. sat
	e. put		e. while		e. floated
46.	a. was swimming	47.	a. life	48.	a. some
	b. swim		b. two		b. hundred
	c. is swimming		c. friend		c. few
	d. swims		d. three		d. much
	e. swam		e. good		e. several
49.	a. Before	50.	a. Bent's		
	b. Recently		b. two		
	c. However		c. her		
	d. Suddenly		d. Bent		
	e. Fearfully		e. one		

1.	C	26.	E
2.	A	27.	A
3.	D	28.	D
4.	E	29.	B
5.	B	30.	C
6.	A	31.	C
7.	C	32.	A
8.	B	33.	E
9.	E	34.	D
10.	D	35.	B
11.	D	36.	B
12.	A	37.	C
13.	C	38.	A
14.	D	39.	D
15.	A	40.	E
16.	E	41.	B
17.	B	42.	D
18.	C	43.	D
19.	A	44.	A
20.	D	45.	C
21.	D	46.	E
22.	E	47.	B
23.	B	48.	E
24.	C	49.	D
25.	A	50.	A

APPENDIX 2

MODIFIED CLOZE TEST

READING TEST

Instructions: Fill in the blanks with the word that completes the meaning. You are given the first letter.

Campers who get lost in the woods are in for a hard time. It is always good to know what to do if you lose your way.

One camper could not f____ his way back to t____. Soon it became dark a____ snow began to fall. H____ knew that if he d____ not stay warm he w____ die. So he did a v____ smart thing. He walked r____ a tree all night l____. The snow became higher a____ higher. His path in t____ snow made a wall a____ the wind. The wall o____ snow and the walking k____ him warm. When day c____, he was able to f____ his way back to t____.

You can keep warm b____ walking and by wrapping b____ around you, if you h____ to have them. Don't t____ to find your way b____ at night or you m____ become more lost. If y____ build a fire, people m____ be able to find y____ more easily.

When one b_____ was lost, he found h_____ way to a town b_____ following a power line. R_____ that power lines, fences, r_____ and streams point the w_____ to where people live.

O_____ two men were trying t_____ find their way out o_____ the woods. They walked f_____ a long time. Finally t_____ came back to their o_____ campfire, the place they h_____ started from. They had b_____ going round in a c_____. There is something that a_____ helps to keep you f_____ circling. Look for a t_____ or hill that is f_____ away and walk to i_____. When you reach this p_____, pick another tree or h_____ and head for that o_____. Of course this doesn't h_____ you if you are in a dense forest and cannot see something far away.

1.	find	26.	you
2.	town	27.	boy
3.	and	28.	his
4.	He	29.	by
5.	did	30.	remember
6.	would	31.	roads
7.	very	32.	way
8.	round	33.	once
9.	long	34.	to
10.	and	35.	of
11.	the	36.	for
12.	against	37.	they
13.	of	38.	own
14.	kept	39.	had
15.	came	40.	been
16.	find	41.	circle
17.	town	42.	always
18.	by	43.	from
19.	blankets	44.	tree
20.	happen	45.	far
21.	try	46.	it
22.	back	47.	point
23.	may	48.	hill
24.	you	49.	one
25.	may	50.	help

APPENDIX 3

DICTATION: 150 words

It was a real shock/ to get your latest letter./ You're the last person/ in the world/ I'd ever expect/ to be involved/ in a three-car collision!/ Most of the people I know/ have had an accident/ at some time or another,/ but you're/ such a careful and considerate driver./ None of your friends/ had heard about it/ until I told them./

Anyway, you're alive/ and that's all/ that matters./ What about your car?/ I hope/ you were fully insured,/ since it sounds/ like the damage was quite extensive./ Will you try to have the car repaired/ - or is/ it a total loss?/

You remember Leon,/ don't you?/ He's gone into the used car business./ If you need another car, perhaps he can help you./ He's got some really cheap ones/ on his lot./

Otherwise,/ I have no real news./ Life is the same as always./ Work is boring./ The weather's wet and cold./

APPENDIX 4

PROFICIENCY TEST

Collège Ville-Marie — Secondary V

GRAMMAR

1. "Will you ask John to stay a little later?"
   "Sure, I'll ask him, if he ...

   a) hadn't already left
   b) won't already leave
   c) hasn't already left
   d) will have already left

2. "Frankly, I thought the orange lamp was as bright.... the red one".

   a) like
   b) as
   c) than
   d) from

3. "Would you like more coffee?"
   "Yes, thanks. I have just ... the last of this cup."

   a) drunk
   b) drink
   c) drinking
   d) drank

4. "The students are worried ... passing the final exam in math."

   a) about
   b) for
   c) in
   d) of

5. On the weekend we often ... T.V.

a) watch the
b) are watching the
c) are watching
d) watch

6. "Robert, will you please ... before you leave?"

a) give to me it
b) to me give it
c) it to me give
d) give it to me

7. "Sue and I are completing our fourth year of English. We ... English for four years now."

a) study
b) had been studying
c) have been studying
d) studied

8. "Where have you been?"
"I ... have this bill checked by the manager."

a) had to
b) must
c) had
d) have to

9. "I was home the entire evening."
"Why ... to the party?"

a) you not go
b) didn't you go
c) you didn't go
d) didn't go you

10. "If my sister ... on time, she will meet you in the lounge at 9:15".

a) is arriving
b) will arrive
c) arrives
d) will have arrived

11. When I was a child, I ... baseball after school.

a) used to playing
b) used to play
c) was used to play
d) used playing

12. "I have two hours before my next plane. I will read this book ... waiting."

a) while
b) before
c) during
d) until

13. I had ... my father for some money.

a) to ask
b) to ask to
c) asked to
d) ask

14. If I had the time, I ... a holiday.

a) took
b) will take
c) would take
c) had taken

15. The fire ... by the fireman.

a) put out
b) had put out
c) is putting out
d) was put out

16. Mr. Gorman, the personnel assistant, will consider the possibility ... you for summer work.

a) to hire
b) to hiring
c) for hiring
d) of hiring

17. "What is ... way of getting to the airport?"

    a) the faster
    b) fastest
    c) the fastest
    d) most fast

18. My friends moved south ... the very dry climate there.

    a) because
    b) because of
    c) is because
    d) because is

19. Does the reservation clerk know ...?

    a) when will the plane leave
    b) when will leave the plane
    c) the plane will leave when
    d) when the plane will leave

20. Our community is interested ... reducing pollution.

    a) to
    b) in
    c) for
    d) of

21. "Mrs. Grant has been working too hard lately."
    "Why ... some time off?"

    a) doesn't she take
    b) not she take
    c) doesn't take she
    d) she doesn't take

22. "What happened?"
    "The barn ... by the storm."

    a) blew down
    b) was blown down
    c) was to blown down
    d) was blown to down

23. Could you please tell me ...?

a) it is what time
b) what time it is
c) what is it time
d) what time is it

24. The teacher will punish all of us for ... chalk.

a) throwing
b) to throw
c) throw
d) thrown

25. "I always make typing errors when I am in a rush."
"Why ... do anything right?"

a) I can't ever
b) can't ever I
c) can't I never
d) can't I ever

VOCABULARY

There are two types of question in this test. In the first type you are to select the word which is the closest in meaning to the word underlined in the sentence.

Example: The wind felt moist on his cheek

a) icy
b) damp
c) hot
d) day

Answer: b) is correct.

In the second type you will see a sentence with a space. Choose the word that would best complete the sentence.

Example: The sea is too rough today to go _____.

a) dancing
b) driving
c) skiing
d) sailing

Answer: d) is correct.

26. "Do you think John is going <u>to catch on to</u> the system?"

    a) break
    b) understand
    c) use
    d) enter

27. My friend was <u>keen</u> to start learning French.

    a) eager
    b) forced
    c) unwilling
    d) obliged

28. "The speaker did not mention many facts. So Judy asked her to be more _____."

    a) polite
    b) certain
    c) specific
    d) knowledgeable

29. She saw the line come <u>taut</u> in his fingers.

    a) apart
    b) away
    c) tight
    d) loose

30. Every student has to answer for his own <u>behavior</u>.

    a) situation
    b) conduct
    c) health
    d) insult

31. Your handwriting is good but it is not as _____ as John's.

    a) untidy
    b) narrow
    c) neat
    d) illegible

32. The natives were <u>immune to</u> the poison.

a) allergic to
b) quickly affected by
c) unable to be harmed by
d) susceptible to

33. As soon as I get in the bath the telephone is <u>bound to</u> ring.

a) slow
b) certain
c) ready
d) willing

34. "I am sorry", the operator said, "but the number you _____ no longer exists".

a) directed
b) dialed
c) declared
d) dropped

35. The countryside here is very <u>rugged</u>.

a) flat
b) low
c) pretty
d) rough

36. That child of yours is a real <u>pest</u>.

a) nuisance
b) darling
c) genius
d) angel

37. You should really try to _____ your money.

a) loan
b) budget
c) accommodate
d) evaluate

38. A boxer must be strong and skillful to defeat Ali.

a) powerful
b) young
c) hard-working
d) accomplished

39. There is no known remedy for this disease.

a) torture
b) name
c) symptom
d) treatment

40. A story with a beginning, a middle and an ending has _____.

a) a conclusion
b) excitement
c) a plot
d) interest

41. This is a photograph of Tom's class. Can you pick him out?

a) find
b) shout at
c) follow
d) watch

42. Fruit is very scarce at this time of the year.

a) expensive
b) nutritious
c) good
d) rare

43. I don't like the _____ of this pie.

a) flavor
b) flower
c) fluency
d) flutter

44. The dog <u>seized</u> the boy by the leg.

a) grasped
b) bit
c) kicked
d) scratched

45. I think you would do better to <u>omit</u> this question.

a) answer
b) exclude
c) obey
d) withhold

READING

Directions: This test consists of three reading passages. Each of the three passages is followed by five questions based on the text. Read each passage carefully and answer the questions which follow it. Choose the best answer and write its letter in the space provided.

One night the interpreter came to me saying that a celebrated musician wanted to meet me, and would I go to his box? The beautiful, exotic lady was a member of the Russian Ballet. The interpreter introduced me. The gentleman said that he had enjoyed my performance and was surprised to see how young I was. At these compliments I bowed politely, occasionally taking a furtive glance at his friend.

"You are instinctively a musician and a dancer, monsieur" said he. Feeling that there was no reply to this compliment other than to smile sweetly, I glanced at the interpreter and bowed politely. The musician stood up and extended his hand and I stood up. "Yes", he said, shaking my hand, "you are a true artist". After we left I turned to the interpreter: "Who was the lady with him?"

She is a Russian ballet dancer, Mademoiselle _____. It was a very long and difficult name.

"And what was the gentleman's name?" I asked.
"Debussy", he answered, "the celebrated composer".
"Never heard of him", I remarked.

46. The incident took place at ...

a) a race track
b) a theatre
c) a circus
d) an airport

47. One evening the author was invited to meet ...

a) an interpreter
b) a famous model
c) two members of the Russian ballet
d) a musician

48. The author decided to accept the invitation because ...

a) he was impressed by the name of the famous musician
b) he did not wish to be rude
c) he was interested in the musician's companion
d) he was very inquisitive

49. Debussy wanted to meet the author because he wanted ...

a) to ask his opinion on a ballet
b) to invite him to a concert
c) to compliment him on his performance
d) to teach him Russian

50. The author is ...

a) a female Russian
b) an exotic dancer
c) an elderly musician
d) a young actor

Many false beliefs are linked with sneezing. A sneeze to the right is lucky; to the left, unlucky. The present custom of saying "God bless you" to a sneezer began with the ancient Greeks. They thought a sneeze was a sign of danger. When anyone sneezed, the Greeks cried aloud to Zeus, the most important god, to protect the sneezer. The saying that something is "not to be sneezed at" started in the Middle Ages. People still thought sneezing harmful.

When snuff was introduced to Europe, people who sniffed this scented, powdered tobacco into their nostrils claimed sneezing was healthful. Sneezing was said to refresh the lungs and clear the brain. A pinch of snuff does cause artificial sneezes. But a natural sneeze is a reflex action. It cannot be controlled.

51. According to one belief, sneezing to the left was ...

a) lucky
b) unlucky
c) a prayer to Zeus
d) a controlled act

52. The Greeks thought a sneeze was ...

a) healthful
b) a reflex action
c) lucky
d) none of the above

53. Implied but not stated: the Greeks ...

a) thought Zeus was the most important god
b) believed in only one god
c) worshipped more than one god
d) had no religion

54. Snuff-takers wrongly believed that sneezing ...

a) refreshed the lungs and cleared the brain
b) was a reflex action
c) could not be controlled
d) prevented colds

55. Sneezes caused by snuff-taking are described as ...

a) loud
b) natural
c) artificial
d) uncontrollable

The first list of the body's senses was made by Aristotle, a learned Greek teacher, over two thousand years ago. He named five senses: sight, hearing, taste, smell and touch. In 1826 a British doctor, Sir Charles Bell, added muscle sense. This sense reports all movement to the brain and spinal cord, the control centres to which all senses report.

Recent medical studies have proved that the senses number more than six. Taste is not one sense but four: the tongue identifies sweet, sour, salt and bitter substances. Touch makes the body aware of pressure, pain and temperature. And in the ear, location of hearing, is the sense of balance. Hunger and thirst are also newly identified senses.

Science now divides the senses into two groups. Inner senses are those of muscle, blance, hunger and thirst. All the others are termed outer senses.

56. The sixth sense added was that of ....

a) balance
b) hunger
c) thirst
d) muscle

57. Recent studies have shown that the tongue identifies ...

a) sweet and sour tastes
b) salt and bitter tastes
c) both a and b
d) more than a and b

58. A recently discovered sense of touch is

a) bitterness
b) smoothness
c) pressure
d) balance

59. The ear is connected with the senses of ...

a) pressure and pain
b) temperature and pressure
c) hearing and pressure
d) hearing and balance

60. The two terms that include all known senses are ...

a) brain and spinal cord
b) inner senses and outer senses
c) taste and touch
d) hunger and thirst

COMPLETIONS:

61. After the civil war minstral shows lost their popularity to _____.

62. Early circuses were called by the strange name _____.

63. An invention which brought great changes to the movies was _____.

64. It is claimed that the first music Americans invented was _____.

65. The most popular type of character in the early plays was the _____.

66. Songs created by people who don't write down the words and music are called _____.

67. Early American folk songs included themes about work and _____.

68. Early entertainment viewed black people as intelligent and likeable people.
True or false

69. The "greatest showman of all time" was _____.

70. Ragtime music combined marching music and _____.

71. After television became popular, radio switched its programming primarily to soap operas.
True or false

72. Elvis Presley provided a perfect mix of ragtime and rock and roll styles.
True or false

73. Poor people most enjoyed going to early movies.
True or false

74. Vaudeville shows made people laugh at blacks and _____.

75. The taste of the American public in entertainment has not changed over the years.
True or false

COLLÈGE VILLE-MARIE Secondary V

NOM _____ Classe _____

INTRA - ANGLAIS 578-522
(feuille-réponses)

GRAMMAR

1.	C	6.	D	11.	B	16.	D	21.	A
2.	B	7	C	12.	A	17.	C	22.	B
3.	A	8.	A	13.	A	18.	B	23.	B
4.	A	9.	B	14.	C	19.	D	24.	A
5.	D	10.	C	15.	D	20.	B	25.	D

VOCABULARY

26.	B	30.	B	34.	B	38.	D	42.	D
27.	A	31.	C	35.	D	39.	D	43.	A
28.	C	32.	C	36.	A	40.	C	44.	A
29.	C	33.	B	37.	B	41.	A	45.	B

READING

46.	B	47.	D	48.	C	49.	C	50.	D
51.	B	52.	D	53.	C	54.	A	55.	C
56.	D	57.	C	58.	C	59.	D	60.	B

COMPLETIONS

61. Vaudeville
62. Museums for education
63. sound track
64. ragtime
65. black slave
66 folk songs
67. legends, heroes, fun stories
68. false
69. P.T. Barnum
70. African rhythms
71. false
72. false
73. true
74. immigrants
75. true

APPENDIX 5

ITEM ANALYSIS OF THE MULTIPLE-CHOICE CLOZE TEST

Item analysis techniques described by Davies and Moller (1973) were employed on the results of the cloze test in order to determine the indices of discrimination and difficulty. The responses of the first and third 27.5 percent of the students were used. The data are presented in the following table.

Table 19

Item Discrimination Indices and Item Difficulty Indices of the Multiple-Choice Cloze Test (n = 28)

Discrimination	% of Items[a]
less than .25	.22
.26 - .39	.12
.40 - larger	.66
Difficulty[b]	
.00 - .14	.00
.15 - .29	.10
.30 - .70	.62
.71 - .85	.20
.86 - 1.00	.08

[a]There were 50 items in the test.

[b]Average difficulty = .57.

Of course, the higher the discrimination values the better it is. It is generally held that items which discriminate less than .25 are not acceptable. With respect to level of difficulty, worthwhile items fall within a range of .30 to .70. Oller (1979) stated that some test writers expand that range to include items between .15 and .85.

There were 11 items that did not discriminate well, four of which were very easy and three others which were very difficult. Ten of these items were content words while the other word was a pronoun. Thirty-one items fell within a range of difficulty of .30 to .70 whereas 46 of the 50 items (92 percent) were in the wider range of .15 to .85. The four items that were outside this range were too easy for the students. The average difficulty for all the items was a very acceptable .57. Morrisette (1979) stated that an exam becomes more sensitive to differences between subjects when item difficulty approaches an average level of .50.

It is impressive that a multiple-choice cloze test, based on a fixed rate of deletion and made up of distractors from the same grammatical category, should have such good indices of discrimination and difficulty. The

cloze test could be possibly improved by eliminating the 11 items that did not discriminate well and by lengthening the passage to include other items. Yet research by Cranney (1973) with native speakers showed that revising a very long multiple-choice cloze test by item analysis did not necessarily increase significantly test reliability or validity.