DICTATION AS A PRAGMATIC LANGUAGE TEST

Ⓒ   Elaine J. Mains

A Thesis

in

The Faculty Centre For the Teaching
of English as a Second Language

Presented in Partial Fulfillment of the Requirements
for the degree of Master of Arts in Applied Linguistics at
Concordia University
Montreal, Quebec, Canada

June 1980

ABSTRACT

DICTATION AS A PRAGMATIC LANGUAGE TEST

Elaine J. Mains

A dictation test was researched as a measuring device for testing
overall second language proficiency and for its ability to predict per-
formance after a period of formal instruction. The dichotomy between
the discrete-point approach and the pragmatic/integrative approach to
language testing was outlined and research pertaining to the reliability
and validity of dictation was discussed. The theoretical position
adopted in this research was based on the paradigm of second language
testing posited by John Oller, Jr.

Thirty-nine subjects in the Advanced English Course at the Canadian
Forces Language School were tested with a dictation test and the results
were statistically compared with the Canadian Forces Standardized
Language Proficiency Test. The dictation test did not prove to be a
good predictor of the subjects' overall language proficiency or the
improvement the subjects made over the duration of the Advanced English
Course. The correlation of .75 between the Canadian Forces Standardized
Language Proficiency Test and the Dictation was not strong enough to
conclude that the dictation and the validating criterion were measuring
the same construct such that one could substitute the dictation for the
Canadian Forces Standardized Language Proficiency Test with confidence.
The phenomena of monitoring and avoidance interacting with the influence
attributable to the level of difficulty of the dictated material and its

1

administration may have been responsible for the lower than hypothe-
sized correlations with the validating criterion.

The conclusion drawn from this experiment was that the dictation
test cannot as yet be ruled out as a reliable and valid measure of
second language learners' overall language proficiency or as predictor
of performance until further research clarifies and determines the
effect of these interacting variables on the subjects' performance in
a dictation.

ACKNOWLEDGEMENTS

A thesis of this nature required the support and assistance of a number of different people from a number of different places. Mr. Ronald Mackay, my thesis advisor, provided the on-going and insightful advice as the thesis progressed. The interest and encouragement he has shown in this endeavour sustained me throughout. I would also like to express my gratitude to the faculty members who took the time to discuss the thesis with me whenever the need arose. In particular, Dr. J. Palmer, who read each and every word and provided important feedback during the writing of the thesis deserves special mention.

I am especially grateful to Major Fred Perry, the Commanding Officer of the Canadian Forces Language Standards Control Detachment. He sought and received permission for me to collect the data needed for the thesis at the Language Standards Control Detachment. Major Perry also arranged for the testing of the subjects and enabled me to obtain essential information needed in the thesis. Major Perry, no doubt in characteristic fashion, went beyond what was necessary and took on the extra task of acting as a reader. A special thanks is extended to the staff at the Language Standards Control Detachment who generously stopped their own work to discuss and advise me in mine.

Kevin and Andrea, my children, deserve a loving thanks for coping so well with their student-mother and frequently putting their needs aside to enable me to complete this work. The thesis, however, is dedicated to a very special person, my husband Gordon, who encouraged me from start to finish, who parented often on my behalf and whose cooperation made the thesis possible.

iii

TABLE OF CONTENTS

LIST OF FIGURES

iv

LIST OF TABLES

INTRODUCTION

Language teaching and learning like other facets of the education process are never static; they are subject to change. Theories of language, theories of how people learn languages and empirical research in first and second language learning are in a constant state of development. These forces not only affect methodology but they also filter down into other areas of language learning, particularly that of language testing.

Language tests are administered for many different reasons. Some of the main ones are: diagnosis, placement, determining achievement or progress, measuring aptitude and assessing general language proficiency. The basic purpose of any language test is to provide information to aid in the making of decisions about the learner's progress and in determining a future course of action suitable for the learner. While language tests are only one way of obtaining such information, they can be considered a primary and essential means for doing so. A good language test demonstrates variance among the testees and thus gives a rank order of the performance of a group of learners. Each testee can be presented with the same testing material which can provide a useful quantitative measure. Thus a good language test can give a certain precision of measurement in assessing the learner's language abilities.

A number of factors must be considered when a test writer is developing a language proficiency test. One must begin by defining the

notion of language proficiency and by having some idea how such proficiency is demonstrated behaviorally. Most definitions of language proficiency are purposefully vague such that they admit to a number of possible ways to test for that proficiency. For example, Harris (1969) defines language proficiency as "what an individual is capable of doing now" (p. 4). Brière (1972) defines it as "the degree of competence or the capability in a given language demonstrated by an individual at a given point in time" (p. 322), independent of the kind of instruction, if any, the testee has received in the language. Useful as these definitions may be, they provide little help in determining how best to measure the individual's language proficiency. Brière points out a further difficulty by asking, "since language testing measures behavior, where can we find a model of behavior which we can use as a guide in designing a general proficiency test?" (p. 322).

Testing techniques throughout the history of second language teaching and testing appear to be related to trends in language theories. Spolsky (1979) identifies three major trends in the history of language teaching, the pre-scientific, the psychometric-structuralist, and the integrative-sociolinguistic trends. The pre-scientific period reflected the grammar-translation approach to language teaching and testing. The psychometric-structuralist approach reflected the need for greater objectivity in language testing and the influence of structural linguistics. It is often referred to as the discrete-point approach to testing. The integrative-sociolinguistic approach reflects the communicative uses of language (pragmatics) and the notion of an integrative or "underlying general language competence that might be common

to all of the various modalities of language use and learning" (Oller, 1976, p. 143). The integrative-sociolinguistic approach is linked to the integrative testing approach and more recently to the pragmatic testing approach because the language elements and components are kept intact. and they emphasize the communicative function of language in language testing.

The Discrete-Point Approach

Language testing philosophy in second language learning is presently being influenced by two competing testing approaches, the discrete-point and the integrative approach. The discrete-point approach in language testing reflects structuralist theories of language and the psychometric techniques emphasizing objectivity of scoring procedures that became dominant in the 50's and 60's. As such, it predates the integrative approach. The structuralist approach to language can be described as a purely linguistic approach. Structural linguistics is based on the principles of segmentation and classification and it provided the basis for categorizing or classifying the elements of language. Language has structure, it is systematic, and it can be categorized into various skills, components of skills and aspects of skills. From the structuralist point of view the distinct elements and skills composing the language should be tested separately. The separation or division of language elements and skills led to the notion of the discrete test item. Discrete-point test items are supposed to challenge only one point of the components of the skills and the aspects of skills at a time. For each skill, aspect and component that is isolated a separate sub-test is devised. Within each sub-test the distinct elements within

that component, skill and aspect are tested through a careful selection of separate items. Language proficiency is viewed as the sum of the scores obtained on tests which purport to measure the various skills and components and aspects of skills tested separately.

Lado (1964) is often referred to as the original proponent of the discrete-point testing approach. He based the construction of language tests on structural linguistic theory and on contrastive analysis. By comparing the native language to the target language the differences in structure were highlighted. Those elements and components of the target language which were different from the native language were the areas of the target language that would present the main difficulties to the learner. By contrasting and analyzing the target language with the first language the differences would be pin-pointed and thus the problems that the second language learner faced would be isolated. By testing the learner's control of the problems, Lado believed he was testing control of the target language. Lado's views were soon suggested to be less sound than he originally assumed as the difficulties the learner faced were not necessarily those areas of the target language that were different in structure from the learner's first language (Upshur, 1962). Moreover, the contrastive analysis approach was never fully implemented in language testing for very practical reasons. Many language classes simply had too many different first languages present to enable even the most dedicated to teach or test via the contrastive analysis approach. Although contrastive analysis was quickly discarded, the basic tenets of the discrete-point approach, with its roots in structural linguistics established itself in language testing in such an entrenched

way that it is part of many classroom and standardized language tests today.

. In the discrete-point approach the content of a language test can be specified in terms of a grid or framework representing the skills and components of language. For example, Harris (1969) conceptualizes the components needed to assess language proficiency as illustrated in Figure 1.

COMPONENTS	LANGUAGE SKILLS			
	Listening	Speaking	Reading	Writing
Phonology/Ortho-graphy				
Structure				
Vocabulary				
Rate and general fluency				

Figure 1.   Chart of the language skills and their components (from Testing English as a second language by D. D. Harris. New York: McGraw-Hill, 1969.)

His conceptualization shows the four components of language, phonology/ orthography, structure, vocabulary, as well as rate and general fluency, forming the vertical axis of a two-dimensional framework. The horizontal axis is made up of the four language skills, listening, speaking, reading and writing.

Cooper (1972) in Figure 2 uses a somewhat more complicated concept-ualization.
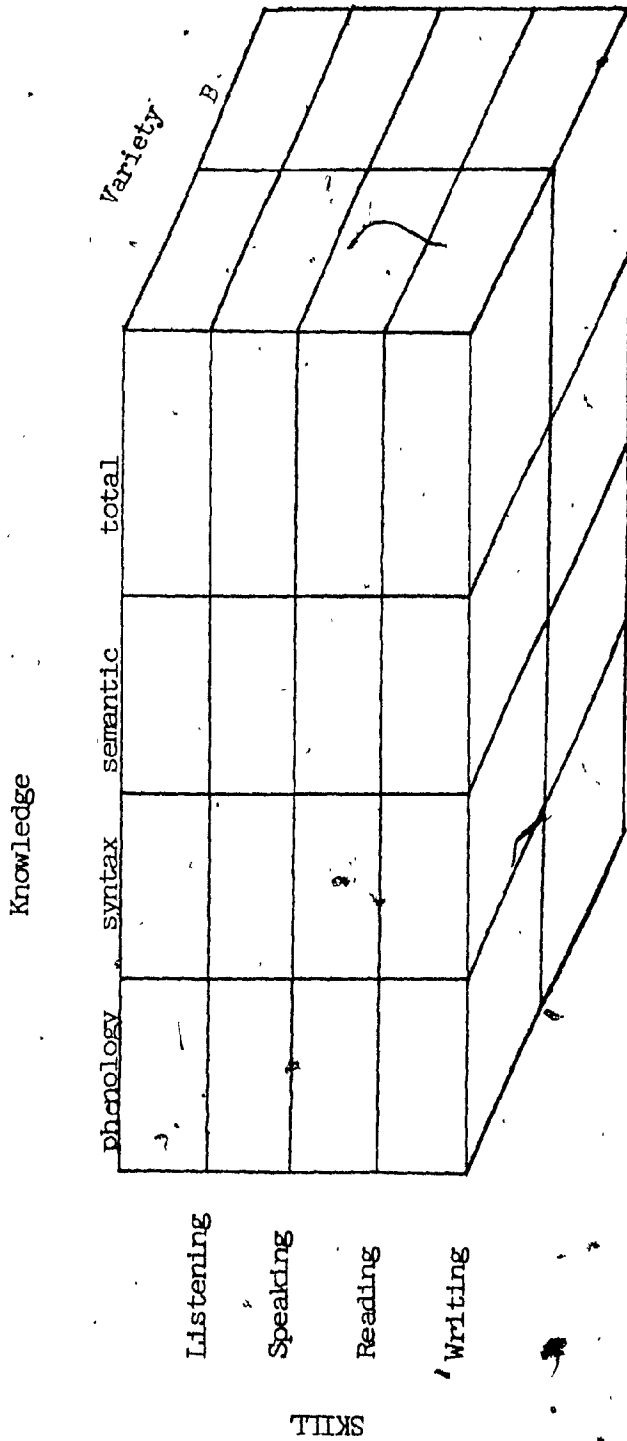
Figure 2. Framework for specifying language test content (from "Testing" by R. L. Cooper. In H. B. Allen & R. N. Campbell (Eds.), Teaching English as a second language: a book of readings. New York: McGraw-Hill, 1972.)

His three dimensional framework contains the four skills (listening, speaking, reading and writing) on the vertical axis applied to four aspects of language knowledge (phonology, syntax, semantics and total); the latter, a combination of the first three aspects, on the horizontal axis. The diagonal axis is made up of two possible varieties of English that the author has identified. Theoretically each of the separate cells could represent a separate sub-test. Thus Harris' framework could conceivably have 16 sub-tests and Cooper's, 32. Both authors are quick to point out however that the number of cells in each framework repre- sents "logical possibilities only" (Cooper, 1972, p. 337). From a practical point of view, 16 or 32 sub-tests would not be feasible. Harris (1969) points out that "in testing, the structures needed in listening and speaking would be regarded as the same as would those of reading and writing" (p. 10). Consequently the number of sub-tests needed would be reduced considerably from the number of possible cells.

Harris (1969) concludes that he would have 6 sub-tests in a language test: listening, speaking, reading, writing, structure and vocabulary. Cooper (1972) suggests that the cells to be assessed would depend on the use to which the test writer makes of the test scores. Thus in the discrete-point approach to language proficiency testing the frameworks provide the basis for the language test content and format.

The Integrative Approach

The competing language testing approach is the integrative approach. The integrative philosophy reflects pragmatic theories of language in which "language proficiency may be a more unitary abstract entity whose components may function more or less similarly in any language based

task" (Stump, 1978, p. 38). Pragmatics is most simply defined as a
language theory that attempts to relate sentences to the contexts of
communication. It emphasizes the way in which people use language in
everyday situations. As Oller (1975) points out, "one of the clear
results of pragmatic investigations thus far is that speech acts are
immensely more complicated than the theories of structural and trans-
formational grammars might have led us to believe" (p. 287). This
statement encapsulates the main criticism leveled against discrete-
point language tests. The approach is too narrowly linguistic because
it does not take into account the communicative function of language.
As well, a proponent of the pragmatic approach would point out that
there is no basis for the assumption that language can be segmented into
separate bits without some loss or distortion of the way in which
language is used. Oller (1979) points out that the way in which we put
language together -- or the organizational constraints of language --
creates an interaction between the elements of the language. This
interaction produces qualities that do not exist in the separate parts.
Oller (1975) also points out that the discrete-point approach fails to
take into account the element of real time. He says "one of the
factors which contributes to the uniqueness and complexity of speech
events in real life is the element of time" (Oller, 1975, p. 288).
Spolsky (1975) points out that the normal redundancy of natural
languages makes no one linguistic item absolutely essential for effective
communication. He says, "it is clear that levels of knowing a language
cannot be characterized in specific linguistic terms, that is, as
mastery of a criterion percentage of items in a grammar and lexicon"

(Spolsky, 1975, p. 77). Thus discrete-point tests which are composed of discrete-point items may not be the most effective measure of the learner's language proficiency.

Language tests which keep the language elements and components intact have been described as integrative language tests. In the past, integrative tests have been defined in opposition to discrete-point tests, that is, if a test is not a discrete-point test, it must be an integrative test. Oller refines the notion of an integrative language test because he prefers to label tests which measure the native-like uses of language as pragmatic tests. He makes the distinction that all pragmatic tests are integrative but not all integrative tests are always pragmatic. The strength of pragmatic tests lies in the fact that they reflect the natural use of language by keeping the language elements intact and incorporate an element of real time. They require the testee to use language in ways which approximate normal human discourse. Oller (1979) states that a pragmatic test is "any procedure or task that causes the learner to relate sequences of linguistic elements via pragmatic mappings to extra-linguistic context" (p. 38). He proposes two naturalness criteria that a valid pragmatic language test must have. First the test must approximate native-like communicative processes — it must contain a time constraint which requires the learner to comprehend verbal sequences rapidly. Oller (1979) puts it this way, "They ... [pragmatic tests] must cause the learner to process ... temporal sequences of elements in the language that conform to normal contextual constraints" (p. 33). The second criterion specifies that a valid language test must incorporate extra-linguistic and linguistic context

such that the learners demonstrate their ability to use language appropriately in normal discourse. Oller (1979) puts it somewhat more succinctly when he says "they require the learner to understand the pragmatic interrelationship of linguistic context and extra-linguistic context" (p. 33). The better known and most frequently used pragmatic tests to date are the written cloze, the dictation and the oral interview test.

Research carried out by proponents of pragmatic tests (Oller, 1979; Stump, 1978; Gunnarson, 1978) suggest that pragmatic tests, because they test language in a more holistic manner measure an underlying language proficiency factor. By keeping the language components intact an interaction between the components creates an important and essential language factor that is not present if the components are separated into discrete items. Thus holistic means that the integrated whole has a reality independent of and greater than the sum of its parts. Scores obtained on a pragmatic language test are the outward manifestation of the learner's internalized linguistic competence. Oller (1979) views global language proficiency as the "psychologically real grammar that the learner has internalized ... an expectancy grammar, that is, a grammatical system that is capable of relating linguistic sequences to extra-linguistic contexts" (p. 24). The individual learning a language is in the process of internalizing a system of grammar that Oller (1979) calls a "pragmatic expectancy grammar" (p. 34). Every natural language has constraints imposed upon it. There are linguistic constraints which shape the form of words, phrases and sentences as well as extra-linguistic constraints which shape the appropriate way that the linguistic aspects of

a language are to be used. In other words, the extra-linguistic aspects of an individual's internalized grammar enables the learner to activate his linguistic knowledge creatively and appropriately. The learner becomes increasingly adept at anticipating what is coming next in discourse situations and the more proficient his expectancy grammar is, the better able he will be to perform in the language.

Dictation as a Pragmatic Language Test

Dictation has long been a part of second language teaching and testing. Its degree of popularity as a testing device has fluctuated according to the language theories popular at any one period of time. It is easy to see why dictation was considered an inadequate test of language ability in the structuralist school. Heaton (1975) sums up the general opinion of the structuralist viewpoint towards dictation by stating, "as a testing device ... dictation ... measures too many different language features to be effective in providing a means of assessing any one skill" (p. 186). The very reason why dictation would be rejected as a language test to a structuralist is the same reason it would be considered to be a valid way of measuring the learner's language proficiency to a pragmatist. Dictation keeps the language components intact and tests language proficiency in a global or wholistic way.

Dictation is now being put forward as a pragmatic language test because it requires the testee to do the kinds of things that native speakers do when processing normal language discourse (Oller, 1979). Administered in such a way as to conform to the two naturalness criteria, dictation can "challenge the student's ability to process

sequences of linguistic elements under constraints of time and to relate strings of the language elements to the broader context of experience" (Gunnarson, 1978, p. 21).

Dictation is primarily an auditory processing task yet it requires the integrative use of a number of language skills and components. The learner must be able to distinguish one word from another in a continuous stream of speech sounds. He must derive meaning from contextually presented linguistic items and he must do so under a time constraint. As Oller puts it "a non-native speaker forms an immediate and wholistic notion of what is being talked about (i.e. meaning) and then analyzes in a 'deliberate, attentive ... sequential' (Neisser, 1967) fashion in order to write down the segmented and classified sequences that he has heard" (Oller and Streiff, 1975, p. 78). This type of processing is referred to as an active "analysis-by-synthesis" (Oller and Streiff, 1975, p. 77) and is considered a highly complex perceptual activity. The listener actively but subconsciously compares what he hears in the stream of speech with his existing knowledge of the language. Oller and Streiff (1975) suggest that the "comparator is no more or no less than a grammar of expectancy" (p. 77). Dictation taps and measures this developing internalized grammar of expectancy which Oller posits as the central component of the learner's language proficiency. The greater the degree of proficiency of the expectancy grammar, the greater the possibility that the learner will accurately interpret the incoming speech sounds. A dictation test requires the testee to demonstrate his ability to comprehend speech by writing it down verbatim. By scoring the dictation in a specific way (see p. 14) a quantitative measure of

the learner's language proficiency is determined. Dictation as a prag-
matic language test activates the internalized expectancy grammar and as
such it provides evidence of the language proficiency of the learner.

Administration of a Dictation Test

The administration of a dictation test is an extremely important
aspect of the dictation procedure if it is to qualify as a pragmatic
language test. The credibility of dictation as a pragmatic language
test probably rests as much on the administration of the dictation as
the scoring of the dictation material. It must be administered so that
it conforms to the two naturalness criteria. Deviations from the two
naturalness criteria would invalidate the results. For example, if a
number of unrelated sentences are dictated to students it is not a
dictation in the sense implied by the two naturalness criteria. The
words, phrases and sentences of a dictation as a pragmatic language
test must be cohesive and coherent. If the dictation is read par-
ticularly slowly so that each word is notably kept separate by the
tester, then it does not qualify as a pragmatic language test. If the
examinee can obtain a repeat of a particular word or phrase then the
test is not a pragmatic language test. Thus when one talks about a
dictation test as a pragmatic language test something very specific is
meant. The presentation of the material in a dictation is crucial in
keeping the dictation within the bounds of a pragmatic language test.
It must simulate what can happen in natural human communication. A
dictation accomplishes this by being read to the examinees in chunks
long enough to challenge the short term memory of the examinee. The
examinee must formulate a broad general idea of what has been said,

then he must determine word boundaries and give graphemic form to the
sounds heard. The entire process, as previously discussed, is con-
sidered a highly complex mental process.

Oller (1979) makes the following suggestions for giving the dic-
tation. It should first be read in its entirety with the students
listening. The dictation is then read in chunks of about seven words
or more depending on the proficiency of the learner. The tester tries
to place the breaks or pauses at natural pausing points. Oller (1979)
says, "breaks should be spaced far enough apart to challenge the limits
of the short term memory of the learner and to force a deeper level of
processing than mere phonetic echoing (p. 273). The pauses between the
dictated phrases are made long enough to enable the testee to write
what he heard. A rough but adequate gauge is to spell the letters of
the words dictated two times through sub-vocally before proceeding.
The passage is then read a third time at normal oral reading speed so
that students have an opportunity to make changes or check their work.
Oller (1979) says that other methods would probably work equally as
well. For example, Valette (1977) suggests that the group of words or
chunks could be repeated once or twice. What is important, however, is
the consistency of presentation. Consistency will help maintain the
same degree of difficulty level for different groups because difficulty
is directly related to the size of the chunks.'

Scoring the Dictation

Prior to the actual scoring, consideration must be given to what
constitutes an error. The following scoring system is based on
suggestions made by Oller (1979). The total score is the number of

words in the selection. Errors in spelling are not counted except those which indicate that the testee had difficulty in the perception of distinct sounds such as "deed" for "did" or which affected the lexical identity of a word such as "their" for "there". Punctuation may or may not be counted depending on the purpose of the dictation test. Oller (1979) says "if the marks are indicated clearly in the oral procedure, putting them in becomes a part of the overall listening comprehension task" (p. 275). A consistent error need only be counted once. One mark is deducted for each incorrect word inserted, for an incorrect word-order, for any missing word and for a word which is distorted phonologically and/or morphologically. Consistency in the scoring method is as important as the method itself.

Selection of the Dictation Material

Oller (1979) suggests that the selection of the dictation material can be left up to the judgment of the experienced teacher. He believes that if dictation meets the naturalness criteria demands, there will be considerable flexibility about the kind of material chosen. He points out that fairly wide ranges of difficulty levels are known to work about equally well on a given population of examinees. Other writers of second language testing texts (Valette, 1977; Heaton, 1975) give no information on how to select materials. It seems intuitively logical, however, that selection of materials should come from natural language sources (textbooks, fiction, newspapers, journals), if the dictation is not to become contrived. At this point in time, selection of the material is made through the common sense judgment of the individual writer making the test.

Empirical Evidence in Support of Dictation

There are a number of research studies which give empirical support for dictation as a pragmatic language test. Valette (1964) correlated the individual total scores of two groups of first year college students learning French as a foreign language. She obtained a correlation of .79 between the dictée and the individual total score of a discrete-point test on one group given only three to four dictées during the semester, that is, minimal practice. She obtained a correlation of .89 with a second group that had been given a dictée at least once a week during the semester. Valette (1964) states that, "for students possessing minimal experience with dictation, the dictée can validly be substituted for the traditional final exam in first year semester French" (p. 434). Thus Valette concluded that dictation was shown to be an effective or valid test procedure which measured the overall proficiency of students who were being taught by an audio-lingual method and, who were not given dictation practice during the term. Oller (1979) states that there is contradictory evidence regarding whether dictation resists practice -- he seems to think it does.

Further support for dictation comes from a research study conducted by Oller in 1971. He intercorrelated data gathered from the administration of the ESLPE 1 at UCLA. The ESLPE 1 is administered to determine the language proficiency of foreign students for placement purposes at UCLA. It is made up of five sub-tests: vocabulary, grammar, composition, phonology and dictation. The sub-test scores were correlated with each of the other sub-tests and with the total score. The data revealed that the dictation score and the total scores of the various sub-tests

correlated at .86. The data also showed that the dictation had a higher correlation with the other parts of the ESLPE 1 than did any other part except for the composition sub-test which correlated at .88.

Oller and Conrad (1971) conducted a pilot study to partially determine the discriminating power of a written cloze test and its validity as a device for measuring ESL skills. The experimenters correlated the scores of the cloze test with the results of the UCLA ESLPE Form 2C. There are five parts in this test -- vocabulary, reading comprehension, grammar, article usage and a dictation. The cloze test correlated with dictation at .82 which was a higher correlation than for the other sub-tests.

In another experimental study, Oller (1972b) correlated three cloze test scores with the part scores of the UCLA, ESLPE 2A Revised. One of the part scores was a dictation test. It correlated at .76, .84 and .85 respectively with the three separate cloze test scores -- a higher correlation than any other part score (vocabulary, reading, grammar) of this test.

Oller (1979) cites another study conducted by S. Johansson in 1972. Johansson's research involved Swedish college students learning English as a foreign language. Johansson's research gives further support for dictation. "His [Johansson's] combined cloze and dictation procedures correlated better with scores on several language tests than any of the other tests correlated with each other" (Oller, 1979, p. 59).

Oller (1972a) conducted a number of intercorrelation studies using four different versions of the ESLPE Form 2. In the four different versions (Form 2A, 2B, 2C, 2D) the dictation correlated at .84, .96,

.96 and .88 respectively with the total score.

Irwin, Atái and Oller (1974) did a correlation study involving a cloze test, a dictation test and the TOEFL. From the correlations obtained they concluded that the listening comprehension sub-test, the cloze test and the dictation test were better predictors of the other parts (in the order given) than any other part score of the TOEFL.

In a paper written by Oller and Streiff (1975) the authors re-evaluated the data of Oller's 1971 research paper. As previously mentioned, the UCLA ESLPE 1 has five parts, vocabulary, grammar, composition, phonology and dictation. In the re-analysis the adjusted dictation score correlated at .94 with the adjusted total (adjusted total means that each of the sub-tests were worth an equal number of points towards the total). Another statistical procedure used was, the inter-correlations of part scores and total scores with self-correlations removed and with equal weightings of part scores. In this analysis the dictation correlated at .85 with the total score.

Stump (1978) intercorrelated the scores of a cloze and dictation test with the scores obtained in two standardized educational tests, the Lorge-Thorndike Intelligence Test and the Iowa Test of Basic Skills (ITBS). Using the statistical procedure of factor analysis, Stump (1978) concluded that "all of the meaningful (non-random) variance of the two standardized tests, as well as the cloze and dictation tasks, can be accounted for by a single unitary factor -- global language proficiency ..." (p. 36).

The consistently high correlation between total scores of discrete-point tests and dictation scores (Valette; 1964; Oller, 1971; Johansson,

1972; Oller, 1972a; Oller and Streiff, 1975) suggest that dictation is a reliable and valid indicator of the learner's language proficiency. At worst, it measured equally as well whatever it is that the discrete-point tests measured -- perhaps a kind of language proficiency. The high correlations between cloze and dictation (Oller and Conrad, 1971; Oller, 1972b; Irwin, Atai and Oller, 1974; Stump, 1978) can be used as evidence to infer that cloze and dictation, despite the different modalities of these tests, share a factor or construct common to both dictation and cloze tests, the learner's underlying language competence in his internalized expectancy grammar. The evidence suggests that not only is dictation as good an indicator of language proficiency as a discrete-point test, but it may in fact be more valid because it appears to tap an underlying competence that discrete-point tests do not get at. Dictation is able to measure the learner's underlying language competence or internalized expectancy grammar because it deals with language in context under constraints of real time.

Specific Problem Area of This Research

The results and conclusions of the research combined with the theoretical discussion presented in the preceding sections suggest that the dictation test is a reliable and valid measure of the learner's developing language proficiency. In order to add further support to dictation as a reliable and valid language test, an experiment was designed to replicate some of the research done previously on dictation. A second aspect of the research experiment was to investigate the degree to which dictation results can be used to predict language proficiency, given the training which a group of language learners received. The following

research hypotheses were tested:

1. The correlation between the scores obtained on a dictation test and the Canadian Forces Standardized Language Proficiency Test will be sufficiently high to suggest that the same construct is being tested by each test.

2. The correlation between the entry dictation scores and the exit Canadian Forces Standardized Language Proficiency Test scores will be sufficiently high to suggest that the entry dictation scores are a strong predictor of the scores obtained on the exit Canadian Forces Standardized Language Proficiency Test.

METHOD

Subjects

The subjects for this research study were the 39 candidates, of all ranks, selected for two separate Advanced English Courses offered at the Canadian Forces Language School, St. Jean. The subjects (2 women, 37 men) are members of the Canadian Armed Forces and have French as their first language. Candidates for the Advanced English Course are selected on the basis of having attained a Level 3 in aural comprehension (see Appendix A for a description of the Rating Scale). Group I was composed of 20 candidates (1 woman, 19 men) whose Advanced English Course began in May 1979 and finished in August 1979. Group II was composed of 19 candidates (1 woman, 18 men) and their course ran from September to December, 1979. The main objective of the Advanced English Course is to improve the reading and writing skills of the candidates. At the beginning and end of each Course the candidates are given a standard-ized English language proficiency test battery (aural comprehension, speaking ability, reading comprehension and writing ability) developed and administered by the Language Standards Control Detachment. During the Advanced English Course a regular dictation was not part of the classroom teaching/learning process, so that there would be no practice effect influence on the subjects' performance.

The Dictation Material

The material selected for the dictation was taken from an article

printed in the Montreal Star on June 30, 1979 (Appendix B). The selection is best characterized as a human interest story. It is made up of 6 sentences with a total of 143 words, which were divided into sequences or chunks ranging from 5 - 13 words. Breaks were placed at natural pause points which accounts for the rather wide range of the number of words per chunk. There was a total of 19 sequences divided in the following way:

Table 1

The Number of Words Per Sequence in the Order Dictated

Sequence Number	Number of Words
1	7
2	7
3	8
4	7
5	8
6	8
7	5
8	6
9	9
10	6
11	7
12	8
13	8
14	15
15	13
16	6
17	7
18	6
19	12

Summarized in another way there was:

Table 2

The Number of Sequences Containing the Same
Number of Words

Number of Sequences	Number of Words
1	15
1	13
1	12
1	9
5	8
5	7
4	6
1	5

Administration and Procedure

Groups I and II wrote the dictation test on the same day they wrote
part of the Canadian Forces Standardized Language Proficiency Test. The
test is completed over a period of two days. Availability constraints
allowed for Group I to be tested at the end of the Advanced English
Course, that is, on August 17, 1979. This will be referred to as Group
I Exit Dictation. Group II was tested at the beginning of the Advanced
English Course on September 6, 1979. This test will be referred to as
Group II Entry Dictation. Group II was also tested at the end of their
Advanced English Course on December 18, 1979. This test will be
referred to as Group II Exit Dictation. The same dictation was
administered on each occasion.

The candidates were aware that the dictation test was being admin-
istered for research purposes. The instructions were given orally in
English. The candidates were told that they would hear the dictation

three times. On the first reading, which was read at normal reading speed, they were instructed to listen only. On the second reading they were told that the dictation would be read in chunks or sequences. The candidates were then told to write the dictation as it was read to them in these predetermined chunks and that enough time would be allowed for them to write each sequence without being rushed (see page 14). A third reading would be given at normal reading speed so that the candidates could make any changes they deemed necessary. The candidates were also told that no other repetitions of the dictation would be given other than what had just been explained, and that punctuation would not be counted. The students were allowed to ask questions about the administration in English or French and all of these questions were answered. It was therefore assumed that all subjects understood the instructions before the dictation was given. The dictation was then administered in the manner described above by the same person (the author of this research paper) on each occasion.

Scoring

The total score was the total number of words in the dictation, that is, 143 possible points. Spelling and punctuation errors were not counted and a consistent error was counted only once. One mark was deducted for a word out of order, for any missing word, for each incorrect word inserted and for a word which was distorted phonologically ("deed" for "did") or morphologically ("whole" for "hole").

The External Validating Criterion

The language proficiency test battery developed by the Language

Standards Control Detachment is used to measure the overall language pro-
ficiency in terms of the individual's language performance. The measure-
ment instrument is best characterized as a norm-referenced language pro-
ficiency measure (see Appendix C for Norms) because the test batteries
are independent of the type of training the candidates have received.
The language proficiency test is composed of four main sub-tests labelled
aural comprehension, speaking ability, reading comprehension and writing
ability. From the raw scores obtained the candidate is given a numerical
rating, ranging from 0-5 in each of the four skills tests. The numerical
rating is based, in principle, on the rating scale devised by the Foreign
Services Institute of the United States State Department. Each of the
six rating levels is based on a functional definition of language pro-
ficiency in each of the four skills (see Appendix A for a description of
each level).

The listening comprehensions test is divided into three sub-tests
and contains a total of 60 items. The questions are of the multiple
choice format. The testee hears a stimulus on the tape and he then
hears the four possible responses or choices. The testee is also able to
read the four choices. In Part 1 the testee hears an incomplete sentence
and he is asked to choose a response which best completes the sentence.
In Part 2 the testee hears a short dialogue and must choose the best
following response out of four possible choices. In Part 3 the testee
hears a statement and he must choose from one of the four possible
choices, the response that means the same as the given statement. There
is a time constraint of 45 minutes for the aural comprehension test.
The raw score is then converted to a numerical rating ranging from 0-5.

The speaking ability test is administered to each candidate individually. The stimulus is tape recorded and the testee responds to questions asked on the tape. As well, the tape recording is accompanied by a visual stimulus (a picture or written stimulus). The teacher makes an on-the-spot assessment using a standard format scoring sheet. The test takes approximately 25 minutes to administer and is divided into five sections. Part I contains 10 separate pictures. The testee hears a response (which is really an answer) and he is expected to give the question that would have to be asked to elicit the response. In Part 2 the testee sees another 10 pictures and he is asked three questions about each picture. He is instructed to answer in complete sentences and to answer in the time allotted. (There is enough time to give an appropriate reply twice, at normal speed.) Part 3 and 4 involve a series of six pictures. The testee is asked to explain each frame, using a complete sentence for each frame. Part 5 involves a subjective evaluation of Part 3 and 4 based on the response in two areas, syntax and pronunciation. The testee can score up to 100 points on the oral test. The raw score is then converted to a numerical proficiency level ranging from 0-5.

The reading comprehension test is made up of 60 multiple choice questions divided into four parts. There is a time allottment of 40 minutes. In Part 1 the testee sees a total of five different pictures. For each picture he must choose out of a possible four, the statement which best corresponds to the picture. In Part 2 the student reads a statement then he must pick out of a possible five choices, the one statement which has the same meaning as the statement given. In Part 3

the testee must read three different passages. Each passage is followed by four multiple choice comprehension questions. The testee must choose the best answer from the four choices given. Part 4 is described by the test writers as a test of global comprehension. It is a multiple choice cloze test. There are 12 words deleted in this section for a possible 12 points. Once again the raw score is converted to a numerical rating based on a range of 0-5.

The writing ability test is composed of 70 items divided into three parts. Once again the multiple choice format is used. This test has a time allottment of 40 minutes. Part 1 is composed of incomplete sentences, that is, one word is deleted in each sentence. The testee must complete each sentence by selecting one choice out of four possible choices. Part 2 is made up of 22 complete sentences. Four words in each sentence are underlined with the letters "a", "b", "c" and "d" placed under each underlined word respectively. A fifth choice, "e" (pas d'erreur) is given at the end of each sentence. Out of the five possible choices the testee must find the error if there is one and place the appropriate letter indicating the error in the space provided. Part 3 is made up of complete sentences, some of which may contain an error. The testee must find the error, if there is one, and then write the corrected word in the space provided. The scores obtained are then converted to a numerical rating ranging from 0-5.

Rationale For the Study and Techniques

If an acceptable definition of language proficiency could be established which could unequivocally account for the numerous variables involved in determining that proficiency, then a valid and reliable

language test would be relatively easy to devise. New tests claiming to be valid and reliable measures of language proficiency would then be verified more easily. Unfortunately such a utopian definition does not exist. Elizabeth Ingram (1978) states that "language tests are not based on any coherent or explicit psycholinguistic theory" (p. 7). A language test's validity and reliability partially rests on the fact that over a period of time it is found to work, that is "one can make better decisions on the basis of the information that they provide than one would without the information" (Ingram, 1978, p. 7). One way to establish the reliability and validity of a new test is to compare the scores obtained on the new test to those obtained on a standardized language test which has been used successfully over a period of time as a measure of the learner's language proficiency.

This brings up a possible problem regarding the kind of inter-relationship that should exist between a pragmatic language proficiency test and a discrete-point language proficiency test. An assumption underlying the discrete-point approach to testing is that each of the components and skills of language is distinct. It follows then that the assessment of the different components and skills should use different tests that include appropriate techniques. The differences between the components and skills of language has led to the accepted notion that a sub-test such as a vocabulary test would not necessarily be highly correlated with any other sub-test such as a grammar test. If a high correlation did result, the construct validity of one or both of the sub-tests would be seriously questioned because different sub-tests supposedly measured different constructs. Furthermore, a sub-test on

a discrete-point test which purported to measure a particular skill or component should correlate highly with the sub-test of the same skill on a different language test. Oller (1976) cites several studies (Darnell, 1968; Oller & Conrad, 1971; Pike, 1973) that cast doubt on the construct validity of several discrete-point standardized tests, namely, TOEFL, Form TEF4, UCLA ESLPE Form 2C. For example, in the study done by Darnell, a cloze test correlated more highly with the listening comprehension section of the TOEFL than with the reading comprehension section. In the Oller and Conrad study (1971) the cloze test correlated more highly with the dictation test than with the reading test scores.

In answer to these criticisms Upshur (1976) offered an interpretation which gave strong support to the discrete-point approach in language testing. Oller (1976) had cited the Pike study to prove that the construct validity between the reading comprehension section and listening comprehension section was not sustained. Pike's (1973) study used the TOEFL (Form TEF4) as well as three experimental measures, a cloze test, an essay and an oral interview. Upshur (1976) took issue with the statistical procedure used by Oller. He said, "the data admit of another interpretation, however, if one considers, not obtained correlations, but correlations 'corrected' for unreliability of the measures" (Upshur, p. 168). By so doing, Upshur was able to show a high correlation between cloze and reading comprehension, "just what one would expect if listening and reading are distinct dimensions of foreign language competence" (p. 169). Upshur (1976) also stated "it is not, therefore, true that the psychometric-structuralist proponent (discrete-point approach) would consider that he had found or made a highly valid

structure test because it had a low correlation with a vocabulary test" (p. 170). He added, "tests which purport to measure the contents of a single cell should exhibit even higher correlations with each other than the high correlations they should exhibit with tests which purport to test the contents of a different cell" (Upshur, p. 170), (see page 6 & 7 for description of a cell). The controversy between what kinds of correlations are expected in language tests dissipates if Upshur's views on the kinds of correlations between sub-tests on a discrete-point test are considered correct. It seems that the current opinion of testing specialists in the field of ESL is that both approaches to language tests do not differ as far as the kinds of correlations they are looking for, that is, language tests which are valid should correlate highly regardless of the mode, skill, component, reception or production aspect being emphasized in any particular test. As well, those tests which are judged as similar or the same as far as mode, skill, component, reception or production aspect should correlate even higher with each other than with other language tests.

The initial step in examining dictation as a valid measure of overall language proficiency was to use the statistical procedure of computing the mean, the standard deviation, the range and the frequency distribution. These statistics gave an indication of the discriminating power of the dictation, that is, the way in which dictation spreads out the individual scores and creates a rank order. Secondly, an independent validating criterion was established. For the purposes of this study the scores on the dictation were compared with the scores attained on the standardized language proficiency test battery

administered by the LSCD for the Department of National Defense (DND).
The Canadian Forces Language Proficiency test is described in more
detail on pages 24-27. What should be noted at this point, however, is
that the validating criterion is not a discrete-point test in terms of
the classical definition of discrete-point test items. In fact, the
LSCD language proficiency battery contains some highly integrative sub-
sections. For example, in the speaking test, Parts 3 and 4 can be con-
sidered a free speech portion which would meet the two naturalness
criteria established by Oller (1979). Part 2 of the aural comprehension
test could also be considered integrative. A testee hears a short
dialogue and must choose the best response (written) out of a possible
four choices. Part 4 of the reading comprehension test is a multiple
choice cloze, described as giving a measure of global comprehension.
Thus what is being used as a validating criterion is a battery of four
tests whose various parts fall on a continuum ranging from discrete-
point to pragmatic-integrative in nature.

In order to be able to infer that dictation is a valid and reliable
measure of language proficiency the scores obtained on the dictation
test must exhibit a high degree of correlation with the validating
criterion. Thus a Pearson product-moment correlation was computed
between the Exit Dictation scores and the part and total scores of the
Exit Canadian Forces Standardized Language Proficiency Test. A multiple
regression analysis using the Exit Dictation score for Group I and II
and the exit test score of the Canadian Forces Standardized Language
Proficiency Test for Group I and II was also computed in order to reveal
statistically the interrelationship between the two tests. Those

statistical results provided evidence for or against hypothesis I:

The correlation between the scores obtained on a dictation test
and the Canadian Forces Standardized Language Proficiency Test
will be sufficiently high to suggest that the same construct is
being tested by each test.

Hypothesis II restated here reads:

The correlation between the entry dictation scores and the Exit
Canadian Forces Standardized Language Proficiency Test scores
will be sufficiently high to suggest that the entry dictation
scores are a strong predictor of the scores obtained on the
Exit Canadian Forces Standardized Language Proficiency Test.

Hypothesis II was tested by correlating (Pearson product-moment formula)
the Entry Dictation scores for Group II with the exit scores on the
Canadian Forces Standardized Language Proficiency Test. The Entry
Dictation scores were also correlated with the improvement score the
subjects made on the Canadian Forces Standardized Language Proficiency
Test. The improvement score, called an X score, was derived in the
following way:

$$\text{Exit Score} - \text{Entry Score} = \text{X Score}$$

(on the validating criterion)

This analysis provided statistical evidence on the degree to which dic-
tation is able to predict the overall ability to learn and the ability
to learn in any of the four skills, given the training the subjects
received at the Language School. The raw scores were converted to Z-
scores for these statistical procedures and the .05 level of probability
was utilized. The Statistical Package for the Social Sciences (Nie,
Hull, Jenkins, Steinbrenner, Bent, 1975), was used for processing all
data.

RESULTS AND DISCUSSION

Descriptive Statistical Procedures

Prior to testing the hypotheses a number of descriptive statistical procedures were carried out in order to summarize, in a very general way, the overall character of the data. Table 3 shows the frequency distribution of the three dictation tests.

Table 3

Frequency Distribution of Test Scores For the Three Dictation Tests

Class Interval 1-5	Exit Dictation Group I	Entry Dictation Group II	Exit Dictation Group II
137 – 141	6	3	4
132 – 136	6	1	3
127 – 131	4	1	4
122 – 126	3	4	0
117 – 121	0	2	2
112 – 116	0	1	1
107 – 111		0	3
102 – 106		2	1
97 – 101		1	1
92 – 96		2	
87 – 91		1	
82 – 86	1	0	
77 – 81		0	
72 – 76		0	
67 – 71		1	
Total	20	19	19

The entry dictation for Group II, that is, the dictation administered at the beginning of the Advanced English Course created a much wider distribution of scores than the Exit Dictation for the same group. The Exit Dictation scores for Group II tended to cluster in the upper range of possible scores. These results initially suggested that the dictation test may be assumed to be sensitive to the improvement in the learners' overall language proficiency, which in turn was related to the effectiveness of the instruction. The scores obtained from Group I, who wrote the Exit Dictation only, also tended to cluster in the upper range of possible scores but to an even greater extent than Group II. Only one subject in Group I scored below 122 points out of a possible 143 points.

Table 4 gives the standard deviation, the mean and the range for the three separate administrations of the dictation test.

Table 4

The Mean, Standard Deviation and Range
For the Three Dictation Tests

	Mean	Standard Deviation	Range
Entry Dictation Group II n - 19	115	19	70 (69 - 139)
Exit Dictation Group II n - 19	124	13.8	43 (98 - 141)
Exit Dictation Group I n - 20	130	12.09	56 (85 - 141)

The findings of these statistical procedures concur with the results shown in Table 3. The difference between the means of the Entry Dictation ($\overline{X}$ - 115) and the Exit Dictation ($\overline{X}$ - 124) suggests that the dictation is sensitive to the students' improvements in the target language. The results of a T-Test on the means provides statistical support for this notion (p - .003). The decrease in the standard deviation from 13.08 - 12.09 is another way of showing that the scores were less spread out on the second administration of the dictation than the first. The one low score of 85 in Group I tended to distort the range as well as lower the mean toward the extreme score. The results shown in Table 3 and 4 suggest that the dictation may not have provided an adequate means for discriminating among the learners, particularly on the Exit Dictation Test. It is also possible that a ceiling effect was operating even though no one achieved a perfect score.

Reliability Testing

The main evidence for dictation as a reliable test comes from the empirical research of Valette (1964), Oller (1971), Johansson (1972), Oller (1972a), and Oller and Streiff (1975). These researchers obtained correlations between dictation and the total score of various discrete-point tests that ranged from .79 (Valette, 1964) to .96 (Oller, 1972a). The researchers concluded from the correlations that the dictation test was valid and from the consistency of these findings they also assumed that the dictation test was reliable.

To add further support to the reliability of the dictation test used in this study, a split-half reliability on the results of the three administrations of the dictation test was carried out. The selected

material was divided in half so that there was a total of 71 words or
points in the first half and 72 words or points in the second half.
Scores were calculated for each half and then correlated (Pearson
product-moment). Reliability was imputed using the following formula:
$R = \frac{2r}{1 + r}$ (Crocker, 1969, p. 38). Table 5 shows the results of the
reliability test.

Table 5

Reliability Coefficients For the Three
Dictation Tests

	Reliability Coefficient (using split-half method)
Entry Dictation Group II n-19	.90
Exit Dictation Group II n-19	.94
Exit Dictation Group I n-20	.87

The reliability coefficients of any language test must be high if
the test can be justifiably used in testing language proficiency.
Crocker (1969) points out that any test with a correlation coefficient
of reliability less than 0.9 is suspect. According to this standard the
reliability is confirmed in two out of three of the dictation tests.
The .87 reliability coefficient, while reasonably high, is somewhat
below the suggested standard. However, questions can be raised about
the way in which the reliability coefficiency was derived. Using the
split-half reliability method it is assumed that each word is of equal

value and that each word conveys as much information or meaning essential to the overall understanding of the dictation as every other word. It is also assumed that each half tests a subject's language proficiency equally as well as the other half. At this point in time there is little evidence to confirm or deny this. Furthermore, there is some controversy as to whether each error is as severe as every other error in such tasks as dictation. It has been suggested that a learner's errors should be judged according to the degree to which they interfere with communication (Bacheller, 1979). If the split-half method carried out in the current study is considered valid then the reliability of this dictation is generally confirmed with some minor doubts being caused by the one reliability coefficient of .87. These results also support the research, mentioned earlier, that infer the reliability of dictation from the correlations obtained between dictation and total scores of discrete-point tests. If the degree of severity of an error is taken into consideration then reliability must be reconfirmed based on the scores obtained using a scoring system which is sensitive to error severity.

Scattergram Analysis

The strength and direction of the relationship between the Exit Dictation and the exit scores of the validating criterion was initially demonstrated by obtaining scattergrams between the Exit Dictation and the part and total scores of the Canadian Forces Standardized Language Proficiency Test. The Exit Dictation scores defined the vertical axis while the part scores (speaking, listening, reading and writing) and the total score defined the horizontal axis. The scattergram for the Exit Dictation and the exit total score of the validating criterion is illustrated in Table 6.

Table 6

Scattergram Between the Exit Dictation and the Exit Total
Score For Group II (n-19)
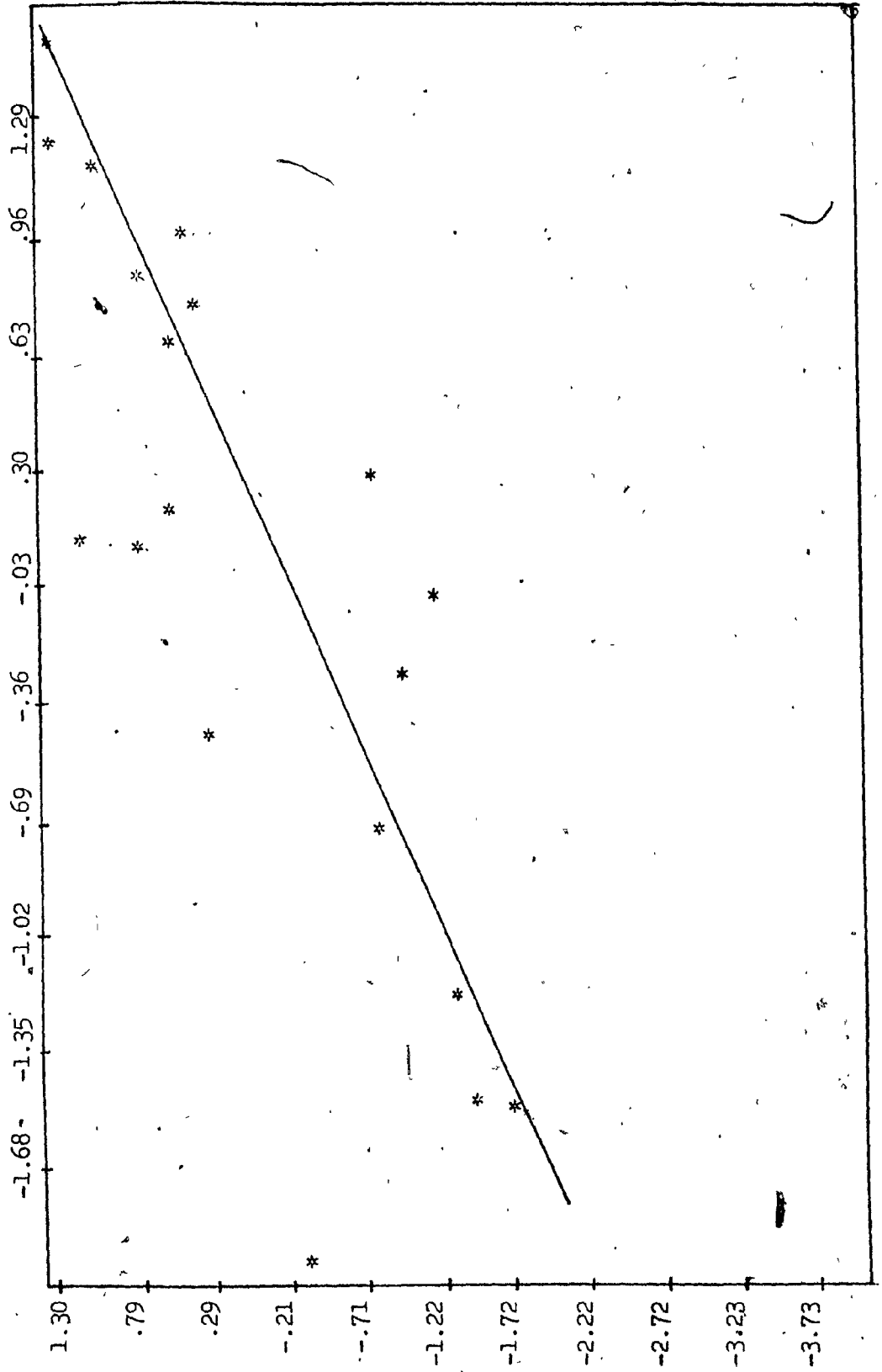
A strong positive linear relationship was obtained between the Exit Dictation and the exit total score. Each of the other variables also demonstrated a positive, though somewhat weaker, linear relationship with the Exit Dictation. These results suggest that as the learners increased their scores in dictation there tended to be a concomitant increase in their exit total score on the validating criterion. This trend was also evident between the Exit Dictation and the part scores, but to a lesser degree. Other statistical tests will help determine just how strong this relationship is.

T-Test Analysis

The final preliminary statistical procedure carried out was a T-Test to determine if the means of the Exit Dictation test obtained from Group I and II were statistically alike so that the Exit Dictation scores from Group I and II could be combined to form a single sample for the testing of Hypothesis I. The results of the T-Test, p = .16, indicated that there was no significant difference between the two means. This allows us to conclude that the groups represented samples belonging to the same population. The data from Group I and II were therefore combined for the testing of Hypothesis I.

Hypothesis I

Table 7 shows the correlation between the Exit Dictation and the part and total scores of the Exit Canadian Forces Proficiency Test. The highest correlation of .75 occurred between the dictation and the total score of the proficiency test. This finding was consistent with the empirical research reported earlier, that is, dictation has

Table 7

Pearson Correlation Between the Exit Dictation Test and the Exit
Canadian Forces Standardized Language Proficiency Test (n=39)

	Speaking	Listening	Reading	Writing	Total
Exit Dictation	.66	.59	.73	.63	.75

p - .001

consistently correlated more highly with the total scores of the various

discrete-point tests than any other part score (Valette, 1964; Oller, 1971;

Johansson, 1972; Oller, 1972a; Oller and Streiff, 1975). An analysis of

the correlation between the dictation and part scores show that the dic-

tation correlated better with the reading test than any of the other

part scores. This finding was also consistent with research discussed

earlier. Dictation has consistently shown a high correlation with cloze

tests (Oller and Conrad, 1971; Oller, 1972b; Irwin, Atai and Oller, 1974;

Stump, 1978) and cloze tests, in turn have been closely related to the

skill of reading comprehension. But, the results suggest that the dic-

tation measures an integrative language proficiency factor better than

any one of the separate skills of the validating criterion. The corre-

lation coefficient of .75 between the dictation and the total score is

high enough to infer that there is a distinct element of commonality

between the two tests.

The results of a multiple regression analysis are shown in Table 8.

Table 8

Multiple Regression Analysis Between Exit Dictation
and the Four Sub-tests of the Canadian Forces
Language Proficiency Test (n-39)

Dependent Variable (Exit Dictation)	R Square (step-wise procedure)
Reading	.53
Speaking	.57
Writing	.58
Listening	.59

The reading test exhibited the strongest relationship with the dictation and accounted for .53 of the variance. The combined contribution of the remaining three parts (speaking, listening and writing) effected a change in the variance to .59. If these remaining three parts were not mainly contributing to the measurement of a common construct they would have made a much greater contribution to the total amount of the variance. Therefore the results of the multiple regression analysis provides further support to the findings of Table 7. Adopting Oller's interpretation of variance the overlapping variance of .59 indicates the degree to which the dictation and the proficiency test measure a common construct. It also appears to be the case that the reading test provides almost as much information about the measurement of the construct common to both tests as the combination of the four parts of the validating criterion. Put another way, 53% of the differences among the subjects in terms of the dictation test scores are predictable on the

basis of the differences in their reading test scores and 59% of the differences of the dictation are predictable on the basis of the differences in the total scores.

It was stated in the hypotheses that a "sufficiently high" correlation must obtain between the dictation and the Canadian Forces Standardized Language Proficiency Test in order to provide support for the hypotheses. A measure was needed to quantitatively characterize the term "sufficiently high". It was determined in the following way. The correlation coefficient of reliability was used to determine the systematic variance of a test. It was assumed that the systematic variance is related to the construct being assessed. If the dictation test was measuring the same construct as the validating criterion the reliability would be approximately equal to the correlation between the two measures (Oller, 1978, p. 69). The lowest reliability coefficient (.87) of the three dictation tests was taken as the standard to quantitatively characterize the term "sufficiently high". Therefore, the dictation test had to correlate at .87 or better with the validating criterion before the hypotheses could be supported. It was felt that such a rigorous standard must be set to establish the reliability and validity of dictation and before dictation could convincingly be used as an alternate measuring device of a learner's global language proficiency. Although the correlation between the dictation and the Canadian Forces Proficiency Test was reasonable strong and positive (.75) it did not meet the sufficiently high standard of .87. Thus Hypothesis I is not supported by the experiment.

Hypothesis II

Hypothesis II was designed to test the ability of the dictation to predict a learner's scores on a subsequent and alternate language proficiency test, given the training the learners received. Table 9 shows the results of the correlation between the Entry Dictation and the Exit Canadian Forces Proficiency Test for Group II. The correlation of .76 between the Entry Dictation and the total score indicates that there is a strong positive relationship between the two tests. The relationship between the dictation test and the part scores shows that dictation correlated better with the listening test at .77 than any other part score. This result is somewhat perplexing as the relationship between the exit listening test and the Exit Dictation test was .59, the lowest relationship obtained between the dictation and any other part score on the exit tests. The data at this point seem to suggest that the dictation may show less stability in measurement with the individual skills tests than with the total score. What may also be occurring here is that the Entry Dictation, because the learners' language proficiency before instruction is lower, was able to discriminate better among the learners whereas in the Exit Dictation the bunching or clustering of scores tends to suggest that the dictation may not have provided the same discriminating power and this accounts for the results. Because these results raised more questions than they answered other statistical analyses were carried out to reveal more clearly just how accurately the dictation could predict final results. First, a Y-score was derived for Group II by subtracting the entry scores from the exit scores of the Canadian Forces Proficiency Test. The Y-scores

Table 9

Pearson Correlation Between the Entry Dictation Test and
the Part Scores and Total Score of the Exit Canadian
Forces Standardized Language Proficiency Test (n-19)

	Speaking	Listening	Reading	Writing	Total
Entry Dictation	.67	.77	.68	.69	.76

p - .001

represented the improvement a learner made in the four skills test and

the total score. The Y-score was then correlated (Pearson product-

moment formula) with the Entry Dictation test and the results are seen

in Table 10.

Table 10

Pearson Correlation Between the Entry Dictation Test and
the Improvement Scores (Y) on the Canadian Forces
Standardized Language Proficiency Test (n-19)

	Y1 Speaking	Y2 Listening	Y3 Reading	Y4 Writing	Y5 Total
Entry Dictation	-.29	.21	-.50	.10	-.21

p - .001

Clearly, the Entry Dictation test failed to show a strong relation-

ship with the Y or improvement scores and we can state that the dicta-

tion was not a good predictor of the improvement the learners made in

the part and total exit tests, given the training they received. With

these results known it was decided to correlate the Y-score with the

improvement the students made between the first and second administra-
tion of the dictation test. An X-score represented the improvement
the learners made in the dictation. The correlation between the X and
Y scores is shown in Table 11.

Table 11

Pearson Correlation Between the Dictation Improvement Score (X)
and the Improvement Score (Y) on the Canadian Forces
Standardized Language Proficiency Test (n-19)

	Y1 Speaking	Y2 Listening	Y3 Reading	Y4 Writing	Y5 Total
X scores	.06	-.29	.55	.42	.29

p - .001

The improvement made in the dictation test correlated weakly at .29
with the total improvement made in the validating criterion and it
related most strongly with the improvement the learners made in reading
comprehension at .55. This finding is not surprising when the findings
of Table 7 are reviewed. Dictation related more strongly to the read-
ing test than to any other part score. The results shown in Table 9
suggest that while there was a positive correlation between the Entry
Dictation and the Exit total score of the validating criterion the
relationship was not sufficiently high to support Hypothesis II. And
the results in Table 10 show that the Entry dictation test could not
be used to predict the improvement in the final outcome of the part or
total scores of the validating criterion. Although the improvement
made on the dictation tests cannot be used as a predictor of outcome

on the validating criterion it serves to support the findings of Table
10. There was a weak relationship with the total score and a moderately
positive relationship with reading. The statistical evidence for
Hypothesis II suggests that a dictation test in this experiment was not
a good predictor of the part or total scores of the validating criterion.
Thus Hypothesis II is not supported.

Dictation has been put forth as a global language measure which
yields substantial information about the learners' overall language pro-
ficiency. The total score of the validating criterion is also assumed
to give a measure of the overall language proficiency of the learners.
One would therefore have expected a higher correlation between the two
measures. The most obvious explanation as to why this did not occur
comes from Oller (1979) who posits that pragmatic tests such as dic-
tation tap an underlying proficiency factor that traditional discrete-
point tests cannot get at. As was discussed earlier in this research,
the validating criterion is both discrete-point in the traditional
sense and pragmatic as described by Oller, leaning heavily toward the
latter kind of test than the former. This fact, that there were
discrete-point elements, may partially account for the lower than
expected correlation but an examination of the dictation material, its
administration and scoring reveal a more practical and tangible
explanation.

It is possible that the dictation tests, particularly the Exit
Dictations did not sufficiently challenge the learners' current language
proficiency or what Oller calls, the learners' developing expectancy
grammar. The clustering of the scores within a narrow upper range

seems to support this notion. Oller (1979) states that "fairly wide ranges of difficulty levels [in dictation] are known to work about equally well in a given population of examinees" (p. 89). However, this research seems to show that the selected material may have been too easy. More careful consideration may be needed when selecting the dictation material in terms of syntactic complexity. It is also possible that the way in which the dictation was administered affected the results. The discriminating ability of the dictation might have been better had the size of the dictated passage been longer and had the dictation been read once or perhaps twice, but probably not three times particularly for advanced learners. "

The scoring procedure used on the dictation test can also be questioned. There are some doubts about the importance of certain kinds of errors made in tests such as dictation. The dictation is usually scored by giving one point for every correct word in the correct order. A missing word, an incorrect word, an incorrect ordering of a word is deemed to be as severe an error as a missing past tense morpheme, or a missing third person plural morpheme. The redundancy that is present at the higher-order level in a dictation is such that a missing "ed" or "s" morpheme is not essential to convey the correct meaning. On the other hand, a missing word, a wrong word or a word out of order may distort the intended meaning, and consequently such errors may be a better indication of the learner's ability to demonstrate his overall language proficiency than a missing morpheme. The discrete-point marking system does not seem to reflect the intended nature of the task which is to assess language globally. Also, the discrete-point system seems to be

less sensitive to assessing the extent to which the meaning intended by the message was received correctly. It seems to be more concerned with the medium itself than with the message. - Noting the inadequacies of this scoring procedure, perhaps consideration should be given to an alternate method of scoring the dictation. Bacheller (1980) proposes a scale of communicative effectiveness (SCE) which grades the severity of the error in the dictated phrases on a five point scale. Such a scale is "intended to be a measure in rendering surface forms of segments of text dictated between pauses" (Bacheller, 1980, p. 68). A grading system like Bacheller's might provide a better measure of the learners' overall language proficiency than the discrete-point marking system. While this possibility was not addressed in the present study it could be the subject of future research.

An alternate way of explaining the resulting correlation in this research is found in Krashen's theoretical model of adult second language learning. Krashen (1978) posits that adults have two separate systems for the development of the learners' language proficiency. Adults can "acquire" language as does a child learning his first language, and the adult can "learn" the target language which can be equated with conscious rule learning that characterizes second language classrooms. Acquisition is viewed as the central component of language performance and as the initiator of all utterances. According to Krashen's conceptualization of language performance, the learner, under certain conditions is able to monitor, using the learned system, the output of the acquired system. Monitoring can occur when there is sufficient time to apply the consciously learned rules and when the

learner is focusing his attention on grammatical form and structure.
What is being suggested here is that the dictation test provided the
opportunity for the learners to monitor their output to their satisfac-
tion. The subjects were given a third reading of the dictation (that
is, the time) to go back over their work as well as being told to check
their work (that is, to attend to grammatical form and structure).
Such was not always the case in the validating criterion. Monitoring
time was available in differing degrees in the Canadian Forces Pro-
ficiency Test. In the reading and writing test subjects would have had
time to review, check and change their answers. On the listening test
and even more so on the speaking test little if any time was available
for monitoring. The differences in total performance between the
dictation test and the validating criterion can partially be explained
by the use of monitoring. The dictation test was measuring a language
proficiency factor which was generated by the acquired system and
monitored by the learned system; the validating criterion was mainly
measuring the acquired system in the listening and speaking tests but
the acquired and possibly the learned system in the reading and writing
test. Thus the learners were able to monitor their performance only
part of the time on the Canadian Forces Proficiency Test and this in
turn would effect and explain some of the differences in correlations
between the dictation and the part and total scores of the validating
criterion.

Another possible factor that can be used to explain the relation-
ship that obtained between the dictation and the validating criterion
is the phenomenon of avoidance. Richards (1973) points out that in

communicative situations learners are able to exert a fair degree of control over their output to suit their intentions. They may use less complex sentence structures and vocabulary in order to find a simpler way of expressing an idea they cannot express in a more idiomatic way. In the speaking test it is possible that the learners are able to avoid certain structures in which they feel an error is more likely to occur and to use those structures and words of which they are more certain. In the dictation test the subjects must demonstrate their language ability by writing verbatim what they believe they hear. As pointed out by Oller (1979) the errors made in a dictation test reveal that analyzing and synthesizing the incoming stream of speech is a highly complex task. In a dictation test the learners cannot avoid using certain structures and words as they must replicate, in the written form, what they hear in the spoken form. The avoidance phenomenon (Schacter, 1974) is possibly a contributary factor which can be used to explain the lower than expected relationship that obtained between the dictation and the validating criterion.

CONCLUSION

One of the major goals of the thesis was to determine if a dictation test was a good predictor of a learner's future language proficiency after having received instruction. The results of the experiment demonstrated that dictation proved to be a moderate but not a sufficiently high predictor of overall language proficiency to be used with confidence in estimating the learner's subsequent performance. Moreover, dictation was not a good predictor of the improvement a learner would make over the duration of the course of instruction as measured by the validating criterion. The tests results may have been confounded because of the ceiling effect in the dictation and before unequivocal statements regarding the predictive abilities of dictation are made, other research would be needed to substantiate or negate the results of this research.

Another major goal of the thesis was to provide empirical evidence in support of dictation as a reliable and valid global language proficiency test. Positive support would imply that a dictation test could be used in place of a longer and more difficult to prepare discrete-point test. Even though some degree of support was obtained for the dictation test the findings do not lead to the conclusion that the dictation test would be an adequate substitute for the validating criterion used in this research. However, two observations resulting from this research suggest that certain factors need to be taken into

consideration if the dictation is to be used as a reliable and valid test of global language proficiency. First, the effect of the administration and the selected material on the learner's performance appear to be critical in determining the reliability and validity of the dictation test. Careful consideration of these two factors implies knowledge of the learners who are to be tested by a dictation so that the dictation can sufficiently challenge those to be tested. The person most likely to decide the most suitable material and how best to administer the dictation is probably the classroom teacher. Thus dictation could find its most useful function as part of the classroom learning/teaching situation. Second, the effects of such phenomena as avoidance and monitoring may need to be considered and controlled for when administering a dictation, particularly if the dictation is to be compared with another language test. Further research that pays particular attention to level of difficulty of the dictation and the administration as well as the effect of monitoring and avoidance is needed before the final conclusion can be made on the dictation as a test of overall language proficiency.

References

Allen, H. B. & Campbell, R. R. (Eds.). Teaching English as a second language: a book of readings. New York: McGraw-Hill, 1972.

Bacheller, F. Communicative effectiveness as predicted by judgments of the severity of learner errors in dictations. In J. W. Oller, Jr. & K. Perkins (Eds.), Research in language testing. Rowley, Massachusetts: Newbury House, 1980.

Brière, E. J. & Hinofotis, F. B. Current trends in second language testing. In L. Palmer & B. Spolsky (Eds.), Papers on language testing, 1966-1974. Washington, D. C.: TESOL, 1975.

Brière, E. J. & Hinofotis, F. B. (Eds.). Concepts in language testing: some recent studies. Washington, D. C.: TESOL, 1979.

Brown, H. D. Papers in second language acquisition. Language Learning, 1976, Special Issue 4.

Carroll, J. B. Fundamental considerations in testing for English language proficiency of foreign students. In H. B. Allen & R. N. Campbell (Eds.), Teaching English as a second language: a book of readings (2nd ed.). New York: McGraw-Hill, 1972.

Cartledge, H. A. A defence of dictation. English Language Teaching, 1968, 22,(3), 226-231.

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

Clark, J. L. D. Direct vs. semi-direct tests of speaking ability. In E. J. Brière & F. B. Hinofotis (Eds.), Concepts in language testing. Washington, D. C.: TESOL, 1979.

Clark, J. L. D. Psychometric considerations in language testing. In B. Spolsky, Approaches to language testing: Series 2. Washington, D. C.: Center for Applied Linguistics, 1979.

Cooper, R. L. Testing. In H. B. Allen & R. N. Campbell (Eds.), Teaching English as a second language: a book of readings. New York: McGraw-Hill, 1972.

Cooper, R. L. & Fishman, J. A. The sociolinguistic foundation of language testing. In B. Spolsky (Ed.), Approaches to language testing: Series 2. Washington, D. C.: Center for Applied Linguistics, 1979.

Crocker, A. C. Statistics for the teacher. Windsor, Berk.: NFER Publ. Co., 1974.

Davies, A. (Ed.). Language testing symposium: a psycholinguistic approach. London: Oxford University Press, 1968.

Davies, Alan. Language testing: survey article. Language Teaching and Linguistic Abstracts, 1978, 11(3), 145-231.

Deyes, A. F. Learning from dictation. English Language Teaching, 1972, 26, 149-154.

Groot, J. M. Validation of language tests. In L. Palmer & B. Spolsky (Eds.), Papers on language testing 1967-74. TESOL, 1975.

Gunnarsson, Bjarni. A look at the content similarities between intelligence, achievement, personality and language tests. In J. W. Oller, Jr. & K. Perkins (Eds.), Language in education: testing the tests. Massachusetts: Newbury House, 1978.

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

Harris, D. P. Report on an experimental group-administered memory span test. In L. Palmer & B. Spolsky (Eds.), Papers on language testing 1967-74. Washington, D. C.: TESOL, 1975.

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

Huck, S. W., Cormier, W. H. & Bounds, W. G. Reading statistics and research. New York: Harper & Row, 1974.

Ingram, E. The psycholinguistic basis. In B. Spolsky, Approaches to language testing: Series 2. Washington, D. C.: Center for Applied

Irvine, P., Atai, P. & Oller, J. W., Jr. Cloze, dictation and the TOEFL. Language Learning, 1974, 24(2), 245-252.

Isaac, S. and Michael, W. B. Handbook in research and evaluation. San Diego, California: EdITS, 1978.

Jones, R. L. Performance testing of second language proficiency. In E. J. Briere & F. B. Hinofotis, Concepts in language testing. Washington, D. C.: TESOL, 1979.

Jones, R. & Spolsky, L. B. (Eds.). Testing language proficiency. Washington, D. C.: Center for Applied Linguistics, 1975.

Krashen, S. Formal and informal linguistic environments in language acquisition and language learning. TESOL, 1974, 10(2), 158-168.

Krashen, S. Relating theory and practice in adult second language acquisition. SPEAQ Journal, 1978, 2, 9-32.

Lado, R. Language testing: the construction and use of foreign language tests. London: Longman, 1964.

Oller, J. W., Jr. Linguistics and the pragmatics of communication. Workpapers UCLA, 1970, IV, 107-115.

Oller, J. W., Jr. Language, communication and second language learning. Workpapers UCLA, 1970, IV, 99-107.

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

Oller, J. W., Jr. Scoring methods and difficulty levels for cloze tests of ESL proficiency. Modern Language Journal, 1972, 56, 151-157.

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

Oller, J. W., Jr. Cloze tests of second language proficiency. Language Learning, 1974, 23(1), 106-118.

Oller, J. W., Jr. Assessing competence in ESL: reading. In L. Palmer & B. Spolsky (Eds.), Papers on language testing 1967-74. Washington, D. C.: TESOL, 1975.

Oller, J. W., Jr. A program for language testing research. Language Learning, 1976, 4, 141-163.

Oller, J. W., Jr. Language testing. In R. Wardhaugh & H. D. Brown (Eds.), A survey of applied linguistics. Ann Arbor, Michigan: University of Michigan Press, 1976.

Oller, J. W., Jr. Language testing today: an interview with J. Oller. English Teaching Forum, 1976, 14(3), 22-37.

Oller, J. W., Jr. A further comment on language proficiency as a source of variance in certain affective measures. Language Learning, 1978, 28(2), 417-423.

Oller, J. W., Jr. Communicative competence: can it be measured? Revised version of a paper presented at the Los Angeles Second Language Research Forum. University of Southern California, Oct. 1978.

Oller, J. W., Jr. Pragmatics and language testing. In B. Spolsky (Ed.), Advances in language testing, Series: 1. Arlington, Virginia: Center for Applied Linguistics, 1979.

Oller, J. W., Jr. Language tests at school: a pragmatic approach. London: Longman, 1980.

Oller, J. W., Jr. & Conrad, C. The cloze technique and ESL proficiency. Language Learning, 1971, 21(2), 183-196.

Oller, J. W., Jr. & Perkins, K. Language in education: testing the tests. Rowley, Massachusetts: Newbury House, 1978.

Oller, J. W., Jr. & Perkins, K. Research in language testing. Rowley, Massachusetts: Newbury House, 1980.

Oller, J. W., Jr. & Spolsky, B. The test of English as a foreign language. In B. Spolsky (Ed.), Advances in language testing, Series: 1. Arlington, Virginia: Center for Applied Linguistics, 1979.

Oller, J. W., Jr. & Streiff, V. Dictation: a test of grammar based expectancies. In R. Jones & L. Spolsky (Eds.), Testing language proficiency. Washington, D. C.: Center for Applied Linguistics, 1975.

Palmer, L. & Spolsky, B. (Eds.). Papers on language test: 1967-74. Washington, D. C.: TESOL, 1975.

Pimsleur, Paul. Testing foreign language learning. In A. Valdman (Ed.), Trends in language teaching. New York: McGraw-Hill, 1966.

Richards, J. Error analysis and second language strategies. In J. W. Oller, Jr. & J. Richards (Eds.), Focus on the Learner. Massachusetts: Newbury House, 1973.

Sawrey, J. & Silver, S. K. Dictation in language learning. In H. B. Allen & R. N. Campbell (Eds.), Teaching English as a second language: a book of readings. New York: McGraw-Hill, 1970.

Schacter, J. An error in error analysis. Language Learning, 1974, 24(2), 205-213.

Spolsky, B. Language testing — the problem of validation. In K. Croft (Ed.), Readings on English as a second language. Cambridge, Massachusetts: Winthrop, 1972.

Spolsky, B. What does it mean to know a language or how do you get someone to perform his competence? In J. W. Oller, Jr. & J. C. Richards (Eds.), Focus on the learner. Rowley, Massachusetts: Newbury House, 1973.

Spolsky, B., Murphy, P., Holm, W. & Ferrel, A. Three functional tests of oral proficiency. In L. Palmer & B. Spolsky (Eds.), Papers on language testing, 1967-1974. Washington, D. C.: TESOL, 1975.

Spolsky, B. (Ed.). Advances in language testing, Series: 1. Arlington, Virginia: Center for Applied Linguistics, 1979.

Spolsky, B. (Ed.). Current trends in language testing. Arlington, Virginia: Center for Applied Linguistics, 1979.

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

Upshur, J. A. Language proficiency and the contrastive analysis dilemma. Language Learning, 1962, 62, 123-127.

Upshur, J. A. Discussion of "A program for language testing research," by J. W. Oller, Jr. Language Learning, 1976, Special Issue 4, 167-174.

Upshur, J. A., Acton, W., Arthur, B. & Guiora, A. Z. Causation or correlation: a reply to Oller and Perkins. Language Learning, 1978, 28(1), 85-91.

Upshur, J. A. Functional proficiency theory and a research role for language tests. In E. J. Briére & F. B. Hinofotis (Eds.), Concepts in language testing: some recent studies. Washington, D. C.: TESOL, 1979.

Valette, R. The use of the dictée in the French language classroom. Modern Language Journal, 1964, 48, 431-434.

Valette, R. Modern language testing (2nd ed.). New York: Harcourt, Brace, Johanovitch, 1977.

Whitaker, S. F. What is the status of dictation? Audio Visual Language Journal, 1976, 14, 87-93.

APPENDIX A

The Proficiency Levels for the Canadian Forces
Standardized Language Proficiency Test

PROFICIENCY LEVELS

SECTION I - GENERAL

201. INTRODUCTION

1.   LSCD has varied froms of language tests in each applicable skill.
Tests vary in duration, number of questions and degree of difficulty.
In order to report and record test results in a meaningful and descrip-
tive format, scores are converted into levels of proficiency. Levels
are graded from 0 which means no significant proficiency to 5 which
means excellent proficiency.

(202 to 205 inclusive: not allocated)

SECTION 2 - AURAL COMPREHENSION

206. LEVEL ZERO (No significant proficiency)

1. General. No ability to recognize words or phrases beyond a few
courtesy forms such as: "Hello, how are you'", or "Bonjour", "merci" and
isolated vocabulary items. A level 0 is unable to grasp the meaning
of a sentence when that meaning is not suggested by his immediate
situation.

2. Linguistic criteria:

    a.   Vocabulary. Isolated common words or phrases only.

    b.   Grammar. No appreciation of sentences or tenses.

    c.   Speed of comprehension. Not applicable.

207.   LEVEL ONE (Elementary proficiency)

1.   General. Comprehension ranges from courtesy forms to simple sen-
tences, directions and basic dialogue relating to age, time, weather,
health, colour, prices, etc.

- 58 -

2. Linguistic criteria:

    a. Vocabulary. Usual nouns, verbs and adjectives associated with above topics. Simple or frequently-used numbers. Should recognize some cognates and high frequency vocabulary.

    b. Grammar. Understands simple affirmative, negative and interrogative utterances using simple tenses.

    c. Speed of Comprehension. Needs slower than normal delivery and/or repetition.

208. LEVEL TWO (Fair proficiency)

1. General. Comprehension is adequate for sentences and questions dealing with topics such as hotel accommodations, shopping, ordering meals, introductions, casual and routine conversations. His comprehension of news reports ranges from isolated ideas, words and sentences to a general grasp of news items, although he may still miss many specific points. He would usually have difficulty dealing with telephone calls or following a conversation between two native speakers. His comprehension is seriously affected by any noise or interference.

2. Linguistic criteria:

    a. Vocabulary. Understands most common words dealing with above topics although often handicapped by low frequency vocabulary.

    b. Grammar. Has fair comprehension of simple tenses (present, past, future and progressive). He understands common modals such as "can, would", "pouvoir, vouloir". Has some understanding of more complex high frequency structures.

    c. Speed of Comprehension. Should understand short uncomplicated utterances at normal speed but will need slower than normal delivery or repetition of more complex or longer statements.

209. LEVEL THREE (Good proficiency)

1. General. Ability to understand the language with sufficient accuracy to satisfy most normal requirements but may need time to adapt to specialized occupational terminology. Can understand most general conversation. He should be able to understand news reports, although he may miss some vocabulary and idioms. Normally has little difficulty with telephone conversations. Rarely experiences difficulty in grasping nuances in meaning, overtones and insinuations. He can usually infer the meaning of unfamiliar communications from the general context.

2. Linguistic criteria:

   a. Vocabulary. Good understanding of most everyday vocabulary which is adequate for all practical and social conversations.

   b. Grammar. Full comprehension of basic structures and good understanding of the more complex structures.

   c. Speed of Comprehension. His comprehension of speech at a normal rate is good. Rarely needs repetition in general conversation.

210. LEVEL FOUR (Very Good proficiency)

1. General. Has full understanding of general conversation and has a grasp of subleties and nuances in meaning, overtones and insinuations. Normally has good understanding of linguistically based jokes and has no difficulty with telephone conversations or news reports.

2. Linguistic criteria:

   a. Vocabulary. Comprehensive and precise understanding of general vocabulary appropriate to the situation as well as good understanding of idiomatic expressions and low frequency words.

   b. Grammar. Understands most of the complex structures.

   c. Speed of Comprehension. May find rapid rate of delivery difficult to follow and may experience difficulty with colloquial and/or regional dialects.

211. LEVEL FIVE (Excellent proficiency)

1. General. Full understanding of idiomatic language for all general purposes. Normally grasps subtleties and nuances in meaning as well as linguistically based jokes. Comprehension is as full and as versatile as that of an educated native speaker.

2. This level of proficiency can only be attained after many years of constant usage in the appropriate environment.

SECTION 3 - ORAL EXPRESSION

212. LEVEL ZERO (No significant proficiency)

1. General. Displays little knowledge beyond a few memorized words and phrases such as: "Bonjour, merci, oui, non, ça va", "Thank you,

good morning, fine, thanks", and other isolated terms. He generally communicates only as a result of questions in stimulus-response situations.

2. Linguistics criteria:

   a. Vocabulary. Isolated common words and phrases only.

   b. Grammar. No ability to produce sentences and tenses.

   c. Fluency. Nil.

213. LEVEL ONE (Elementary proficiency)

1. General. Can handle minimum courtesy requirements with some degree of consistency. He can cope with limited communication dealing with such topics as weather, health, colour, age, prices, time, etc. but errors are frequent.

2. Linguistic criteria:

   a. Vocabulary. Limited to above elementary topics. Consists mostly of simple nouns, cognates, a few common adjectives and verbs. Frequently needs to revert to mother tongue.

   b. Grammar. Can produce but may not be able to manipulate simple, basic structures such as "subject - verb - complement". Has limited use of infinitive or present tense of few simple verbs. Uses appropriate question words in "who, what, where, when, how" structures. Often has problems with word order.

   c. Pronunciation. Meaning of utterances should be generally intelligible to a native speaker, although misunderstanding may occur and repetition may be required.

   d. Fluency. Communication broken by many pauses for reflection. Isolated ideas given in telegraphic style.

214. LEVEL TWO (Fair proficiency)

1. General. Can cope with social demands such as giving directions, requesting hotel accommodation, shopping, ordering meals and introductions. Can carry on limited casual conversations about current events, family and other familiar topics. A strong level 2 can handle most routine requirements although his limited proficiency and some interference from his mother tongue will at times impede communication.

2.   Linguistic criteria:

    a. Vocabulary. Only sufficient to handle above subjects. The
       average level 2 will often stop in mid-sentence to rephrase
       or to grope for words. His limited vocabulary will interfere
       with communication and will sometimes cause misunderstanding.

    b. Grammar. Generally a level 2 will use mainly the present
       tense but will normally also have a limited control of past
       and future tenses. He should have some control of simple
       modals such as: "can, may, would", "vouloir, devoir, pou-
       voir". Attempts to use more complex "high frequency" struc-
       tures. Should be able to ask questions reasonably well
       although not always correctly.

    c. Pronunciation. Pronunciation may interfere with communi-
       cation but despite strange accent he will usually be under-
       standable to a native speaker.

    d. Fluency. Ranging from exchange of basic sentences to more
       continuous expression of ideas, descriptions and explana-
       tions. Some hesitation caused by limited vocabulary and/or
       structural control. Speaking rate slower and more halting
       than that of a native speaker.

215. LEVEL THREE (Good proficiency)

1. General. This level is considered to be the "functional" or
"working knowledge" level. His knowledge of the language enables
him to handle most conversational topics of a general nature. He can
discuss everyday routine matters such as social events, hobbies, family
but will run into difficulty when arguing on fine distinctions or when
expressing abstract ideas. Flow of speech is fairly regular on fami-
liar topics and errors seldom interfere with understanding.

2. A level 3 may be awarded to individuals who do not meet all the
linguistic criteria mentioned below but who compensate with a high de-
gree of fluency and ease of communication.

3. Linguistic criteria:

    a. Vocabulary. Sufficient to handle all topics in normal con-
       versational situations. Can go into more detailed explana-
       tions of familiar subjects and can rephrase to cover lack of
       specific vocabulary.

    b. Grammar. Control of basic grammatical structures is good.
       Should be able to use which reasonable accuracy more complex
       structures such as connectives and subordinate clauses, etc.
       although there may still be some interference from mother

tongue. Structural mistakes do occur but they do not inter-
fere with communication.

c. Pronunciation. A level 3 still has difficulty in mastering
typical phonemes which do not exist in his own language but
this rarely causes misunderstanding.

c. Fluency. Can speak with ease on everyday topics with some
circumlocution, but may grope for words when discussing less
familiar subjects.

216. LEVEL FOUR (Very good proficiency)

1. General. Able to use language accurately, confidently, and
fluently for all requirements but would rarely be taken for a native
speaker. He uses more sophisticated language and is more aware of
nuances, distinctions, etc. He can communicate in discussions at
any level but could experience difficulty in conveying subtleties
and shades of meaning.

2. Level 4 normally indicates extensive usage of the target language
in an environment where it is the primary means of communication.

3. Linguistic criteria:

a. Vocabulary. Comprehensive and precise control of general voca-
bulary and idiomatic expressions appropriate to the situation.
Shows more versatility in choice of vocabulary but occasionally
misuses more abstract or specialized terminology.

b. Grammar. Errors are rare. Normally uses complex structures
accurately and links ideas in an organized, logical fashion
using low frequency connectives. Occasional interference
from mother tongue.

c. Pronunciation. Pronunciation does not interfere with commu-
nication although individual will not normally be taken for a
native speaker.

d. Fluency. Normal rate of speech in dealing with all general
requirements, but still may be hesitant under stress situa-
tions.

217. LEVEL FIVE (Excellent proficiency)

1. General. The individual's command of the target language equals
that of an educated native speaker. His performance is native-like in
all situations, including debates, speeches and conferences. He can
use the subtleties of language to express moods, attitudes, and other
abstract concepts.

2. This level of proficiency can only be attained after many years of constant usage in the appropriate environment.

SECTION 4 - READING

218. LEVEL ZERO (No significant proficiency)

1. General. Candidate's proficiency ranges from no ability to understand the printed work to minimal ability to comprehend isolated words or phrases in the simplest of texts.

2. Linguistic criteria:

    a. Vocabulary. Isolated common words or phrases.

    b. Grammar. No significant comprehension of structures.

219. LEVEL ONE (Elementary proficiency)

1. General. The candidate is able to understand reading material of a very basic nature such as very simple stories, second language beginners' texts, etc.

2. Linguistic criteria:

    a. Vocabulary. Can understand high frequency words and cognates.

    b. Grammar. Can understand basic "subject - verb - complement" structures, simple interrogatives, negations, prepositional structures, and simple verb forms.

220. LEVEL TWO (Fair proficiency)

1. General. Can understand reading material more complex than simple subject - verb - complement structures. Comprehension in some detail of simple memos, menues, labels, short texts and stories, but will require the extensive use of a dictionary and/or translation.

2. Linguistic criteria:

    a. Vocabulary. Recognizes and understands cognates and middle to high frequency range vocabulary within the realm of his experience and interests.

    b. Grammar. Structural elements of an intermediate level of

difficulty including an adequate differentiation of past,
present and future tenses plus common modals and simple
conditionals.

221. LEVEL THREE (Good proficiency)

1. General. This level identifies the functional reader who is able
to understand most texts of a general nature. He may still experience
some degree of difficulty with specific vocabulary, false cognates and
areas of linguistic contrast between native and target language. The
candidate has a good understanding of material directed to the general
reader. Moderate use of dictionary may be required.

2. Linguistic criteria:

   a. Vocabulary. Should be able to understand all vocabulary
      within the realm of his experience and interests. Will still
      experience difficulty with abstract concepts, specialized
      terminology and low frequency lexical items.

   b. Grammar. Should understand most grammatical structures in-
      cluding long sentences with various connectives, subordinate
      clauses, interjections, etc. Mother tongue interference may
      still cause some misunderstanding.

222. LEVEL FOUR (Very good proficiency)

1. General. The candidate fully understands standard reading ma-
terial normally found in magazines, newspapers and textbooks. He can
read technical material of a general nature and abstract prose with a
minimum of difficulty. At this level, the individual has considerable
flexibility in understanding the different styles of writing although
he may miss some subtleties. His ability to infer meaning from con-
text is well developed.

2. Linguistic criteria:

   a. Vocabulary. Recognizes and understands most words, even
      those taken out of their normal context. Recognizes diffe-
      rences in shades of meaning but may not recognize low fre-
      quency words and expressions.

   b. Grammar. Readily recognizes and understands practically all
      grammatical forms but may occasionally encounter difficulty
      with the most complex structural elements.

223. LEVEL FIVE (Excellent proficiency)

1. General. The candidate's comprehension equals that of an edu-
cated native speaker. He can understand many different styles of
writing, even extremely difficult and abstract prose, with full com-
prehension of the subtleties and nuances.

SECTION 5 - WRITING

224. LEVEL ZERO (No significant proficiency)

1. General. May be able to write isolated words or phrases correct-
ly, but has no practical control of target language grammatical
patterns and lexical items.

225. LEVEL ONE (Elementary proficiency)

1. General. Limited lexical and structural control restricts pro-
duction to single ideas contained in short, simple phrases and senten-
ces. Attempts at more complicated structures are usually the result
of translation and often confuse rather than clarify the idea being
communicated. Errors are numerous and he is able to express only the
most basic ideas.

2. Linguistic criteria:

    a. Vocabulary. Limited to high frequency words and cognates.

    b. Grammar. Grammatical control limited to the most basic
        forms; often translates literally; in general, usage of pre-
        sent tense and infinitives predominates.

    c. Mechanics (spelling, punctuation and capitalization) and or-
        ganization. Skills in these areas are practically non-
        existent.

226. LEVEL TWO (Fair proficiency)

1. General. Has limited control of simple structural forms and can
often incorporate more than a single idea within a sentence. Still
dependent on translation, particularly for more complex structural
forms and low frequency expressions. Ideas are fairly comprehensible
despite many errors. Can fill in routine forms, write short, simple
notes and memoranda but normally requires the use of a dictionary
and/or the assistance of a more proficient writer. Interference from
native language still obvious.

2. Linguistic criteria:

    a. Vocabulary. Able to use most high frequency words and cognates. Attempts to use lower frequency lexical items with reference to a dictionary.

    b. Grammar. Has a fair control of the basic structural elements including the present, past and future tenses. Can also use high frequency modals, connectives and modifiers. Errors will still persist. Some attempts at lower frequency structural patterns are evident.

    c. Mechanics and organization. Still highly dependent on dictionary for spelling and on native language usage for punctuation and capitalization. Shows progress towards logical organization of ideas.

227. LEVEL THREE (Good proficiency)

1. General. Can write language with sufficient structural accuracy and vocabulary to communicate ideas and opinions on most general topics. Grammatical errors may discourage attempts to correspond officially with target language public. Influence of native language is still noticeable.

2. Linguistic criteria:

    a. Vocabulary. Sufficient to meet routine and general requirements.

    b. Grammar. Control of basic grammatical structures is good. Should be able to use, with reasonable accuracy, more complex structures such as connective and subordinate clauses. Structural mistakes do occur but they do not interfere with communication.

    c. Mechanics and organization. Errors are less evident and mechanics are adequate for day to day communication. Ideas are normally presented in a logical and easily readable fashion.

228. LEVEL FOUR (Very good proficiency)

1. General. Can express ideas and opinions accurately and with ease for all requirements. Can use more sophisticated lexical and structural language elements. Errors are rare although would rarely be taken for a native writer. Can produce memoranda, letters, reports, etc., but material may require editing if to be used for official purposes.

2. Linguistic criteria:

   a. Vocabulary. Broad range and precision of vocabulary including abstract and specialized terms. Although able to express most nuances, he may occasionally choose a term or expression inappropriate to a particular context.

   b. Grammar. Errors are few but may occur in areas such as sequence of tenses, gender and choice of articles, prepositions and pronouns.

   c. Mechanics and organization. Errors in mechanics are rare but those that occur are indicative of mother tongue influence. Smooth and logical flow of ideas indicate that the individual is "thinking in the target language" and is developing a personal style.

229. LEVEL FIVE (Excellent proficiency)

1. General. The individual's command of the target language equals that of an educated native writer. Can write on any topic in a variety of styles depending on the purpose of communication. Familiar with conventional formalities of different types of written communication. Can recognize and manipulate subtle distinctions in use of vocabulary and structures. Can produce any written documents expected of his level of education and experience, to be used as such for any official purposes required.

APPENDIX B

The Dictation Material

Drunk Gets Orange Streak

A downtown drunk who dropped his trousers in front/ of a lady house-painter/ is sporting an orange stripe for his efforts./ Carol Carl said she was touching up/ the exterior of a downtown hotel this week/ when an elderly man staggered up to her/ and pulled down his pants/ in the middle of the street./ "The only weapon I had was my paint-roller/ so I painted him bright orange/ from the top of his bald head/ down to, well, you know where," she said./ "The man pulled up his pants in haste/ and wandered down the street, wiping the streak of paint off his face as he went," she said./ A city policeman confirmed her story./ "The man wasn't too hard to spot./ There are lots of drunks there/ but only one of them was sporting a bright orange paint job."/

APPENDIX C

The Normed Data for the Canadian Forces
Standardized Language Proficiency Test

Table 12

Normed data for the LSCD Language Proficiency Test, administered to Group I, on August 17, 1979 (Exit Language Proficiency Test)

	MEAN	STANDARD DEVIATION	STANDARD ERROR OF MEASUREMENT	AVERAGE DISCRIMINATION INDEX	PERSON R	KUDER-RICHARDSON FORMULA	SPLIT ITEM	SPLIT POPULATION
AURAL COMPREHENSION (ENG A 178)	34.03	13.38	3.1	.52	.96 (with ENG A 72)	.95	.91	.96
SPEAKING ABILITY (ENG B 175)	50.14	28.37	2.4	.63	.97 (with ENG B 71)	—	.99	—
READING ABILITY COMPREHENSION (ENG C 1 78)	34.30	13.96	—	.57	.92 (with ENG C 72)	.95	.91	.92
WRITING ABILITY (ENG D 1 78)	34.10	16.90	—	.65	.96 (with ENG D 72)	.97	—	—

Table 13

Normed data for the LSCB Language Proficiency Test administered to Group II,
on September 7, 1979 (Entry Language Proficiency Test)

	MEAN	STANDARD DEVIATION	STANDARD ERROR OF MEASUREMENT	AVERAGE DIS-CRIMINATION INDEX	PERSON R	KUDER-RICHARDSON FORMULA	SPLIT ITEM	SPLIT POPULATION
AURAL COMPRE-HENSION (ENG A 1 78)	34.03	13.38	3.1	.52	.96 (with ENG A 72)	.95	.91	.96
SPEAKING ABILITY (ENG B 78)	50.47	28.25	2.38	—	.98 (with ENG B 1 75)	—	—	—
READING ABILITY COMPRE-HENSION (ENG C 72)	31.70	13.36	—	.50	.92 (with ENG C 1 78)	.94	.91	.93
WRITING ABILITY (ENG D 72)	36.94	16.65	—	.54	.97 (with ENG D 1 78)	.96	—	—