NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

UNIVERSITY OF ALBERTA

LANGUAGE OF TESTING EFFECTS FOR ACADEMIC ACHIEVEMENT OF FRENCH IMMERSION STUDENTS

by

MICHELE J. SAMUEL

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF EDUCATION

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

EDMONTON, ALBERTA

SPRING, 1990

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

ISBN 0-315-60362-3

Alberta
EDUCATION

Devonian Building, West Tower, 11160 Jasper Avenue, Edmonton, Alberta, Canada T5K 0L2

May 3, 1990

Graduate Studies and Research
2-8 University Hall
University of Alberta
Edmonton, Alberta
T6G 2E1

Dear Sir or Madam:

This letter is to confirm that permission was granted to Michele Samuel to use and include in her thesis copies of the following tests:

Grade 6 Social Studies Achievement Test, Part A: Multiple Choice (1985)

Test de Rendement Etudes sociales 6e année, Partie A: Choix multiples (1985)

If you require more information, please feel free to call me at 427-2948.

Sincerely

Darlene Montgomery

for Frank G. Horvath
Director
Student Evaluation and Records

FGH:2222-A

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR MICHELE J. SAMUEL

TITLE OF THESIS LANGUAGE OF TESTING EFFECTS FOR ACADEMIC ACHIEVEMENT OF FRENCH IMMERSION STUDENTS

DEGREE FOR WHICH THESIS WAS PRESENTED MASTER OF EDUCATION

YEAR THIS DEGREE GRANTED SPRING, 1990

(SIGNED) M. J. Samuel

PERMANENT ADDRESS:

#7, 8616 - 148 St

Edmonton, Alberta

T6E 4M4

DATED June 23 1990

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled LANGUAGE OF TESTING EFFECTS FOR ACADEMIC ACHIEVEMENT OF FRENCH IMMERSION STUDENTS submitted by MICHELE J. SAMUEL in partial fulfilment of the requirements for the degree of MASTER OF EDUCATION.

Dr. S. Carey, Co-Supervisor

Dr. T.O. Maguire, Co-Supervisor

Dr. J. Edwards

Dr. R. F. Mulcahy

Date April 23/90

To the memory of my grandmother,

Desirée Blaquiere

ABSTRACT

The primary purpose of this study was to investigate the effect that language of testing has upon the scores of French Immersion students writing a standardized test of Grade 6 social studies achievement. It also examined the extent to which time is a variable that affects the test performance of French Immersion students.

Two experiments were run. In the first experiment French and English forms of a standardized test of social studies achievement were randomly assigned to Grade 6 French Immersion students. All of the standardized conditions of administration established for the test were followed, including the time limits. All of the conditions of Experiment 1 were replicated in Experiment 2 except that examinees were given unlimited time to complete the test.

In both experiments the scores of students writing the French form (F group) and those of students writing the English version (E group) were analysed according to the classification scheme used to report the results of the 1985 provincial achievement testing administration (Student Evaluation Branch, 1985c). The results from each experiment were interpreted separately and then comparatively in terms of the research questions that were posed.

The results revealed that French Immersion students achieve significantly lower scores when they write an achievement test in French as compared to English. While the differences in scores across all reporting categories are significant, the size of those differences is not constant. In particular, the effect sizes on topic specific data-based questions are all larger than the effect sizes on the same topic discrete item reporting categories. This indicates that the amount that scores are depressed is related to the type and topic of the items.

The scores of those students who wrote under timed conditions do not differ in any significant or important way from those obtained by students who had unlimited time to write the tests. This suggests that time pressures are not a *major* contributing factor in the

depression of F group scores.

The results from this study imply that it does matter in which language French Immersion students are tested and that this variable must be taken into account when interpreting test data.

ACKNOWLEDGEMENTS

I would like to sincerely thank all those people who have assisted and supported me in my effort to complete this thesis. In particular, I wish to acknowledge and express my appreciation to:

Dr. S. Carey, for suggesting a thesis topic that is still fascinating to me and for simultaneously challenging and supporting my thinking about the issue every step of the way.

Dr. T. O. Maguire, for dealing with me in a consistently respectful and supportive manner.

Drs. J. Edwards and R. F. Mulcahy, for so generously agreeing to serve on my committee.

Robert Runté, my friend and confidant, for his never-ending support and encouragement, including hands-on help when it was desperately needed.

Dr. G.W. Nixon, for reflecting back to me that which I needed to see to be able to finish this project.

Roy Jaffray and David Wasserman, for volunteering their time and sharing their expertise with me.

David Raboud and Jan Taylor, for caring about me and showing it throughout this project.

Guy and Helen Raboud, for teaching me to value truth and for helping me to develop the strength and courage that is required when seeking it.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

Introduction

Background to the Study

It is not idle curiosity that motivates educators and parents to find out how well students are achieving the objectives of school curricula. Instead, what underlies achievement testing programs is a desire to use such information for decision-making purposes (Bloom, Madaus, & Hastings, 1981; Sax, 1974). Because educational decisions have far-reaching pedagogical, social, and political implications, it is important to ensure that the conclusions one draws about levels of achievement are accurate. The accuracy of one's conclusions is founded on valid interpretations of test data.

The interpretation of achievement test data is a challenging task. According to Cronbach, (1971) tests are generally assumed to measure only the traits or constructs under study. In reality, extraneous variables may account, in part or in whole, for examinees' responses to test questions. These extraneous variables need to be controlled, or their effects accounted for, to have data interpretations that are valid. This requires the ability to distinguish relevant variables from the many elements sensed by the observer. This, in turn, depends on the ability to conceptualize certain elements as having a confounding effect on test outcomes (Jones, 1971).

Purpose of the Study

This study is about the validity of interpretations made about French Immersion program achievement test data. In particular, it examines Carey's (1980) hypothesis that language of testing has an influence on outcomes thereby invalidating or confounding conventional interpretations of test data.

In a report to the Department of Education, Carey (1980) recommended that the only valid way to measure French Immersion student achievement was to test that achievement in both languages. He based his recommendation on a belief that when French Immersion students, like any students, answer test questions, their responses are determined not only by their subject-matter knowledge and skill (the attributes under study) but also by the quality or nature of the test and by the students' ability to read the test questions. He argued that because the quality of the test and/or the students' ability to read might vary, depending on the language of testing, these extraneous variables could have an effect on how students respond to the test questions. In short, factors associated with the choice of language of testing could systematically shape how students respond to the test questions.

Carey's (1980) point is a significant one because, according to Capell and de Porcel (1979), differences in the scores generated by "parallel" achievement tests in two languages "signal the likely presence of some form of differential validity" (p.103). In other words, the conclusions one draws about levels of achievement and their implications for program change could differ depending on the language of testing.

In spite of its significance, it would be inappropriate to act on Carey's (1980) recommendation without further investigation because the evidence to support or refute his contention that language of testing affects outcomes is inconclusive. Secondary findings from a study by Swain and Lapkin (1981) do not support Carey's prediction that language of testing will affect outcomes. They reported that Grade 4 French Immersion students performed in French as they had in English to parallel forms of a test of social studies achievement. In other words, language of testing had no observable impact on outcomes.

Swain's and Lapkins's (1981) findings must be accepted with caution, however, for two reasons. First, there is no way of determining the internal validity of their study because their description of the instrumentation and method used is very limited. This means that uncontrolled variables could have accounted for their results. For example, nonequivalence of the tests could have produced their results if one test was more difficult than the other.

Similarly, subject selection differences could have confounded the outcomes. This possibility cannot be eliminated because there is scant information provided about the method of assigning subjects to conditions.

The second reason for questioning Swain's and Lapkin's (1981) conclusion is that their findings contradict the results of an American study which was much broader in scope. Willig (1985) conducted a meta-analysis of the results of studies of second language programs in the United States. She found that 63% of the total variance of effect sizes across studies could be accounted for in terms of six extraneous variables. One of these variables was the language of testing ($p.<.0001$). The significance of Willig's finding is not only that it contradicts Swain's and Lapkin's (1981) results but also that it is consistent with Carey's (1980) prediction.

The purpose of this study was to test Carey's (1980) hypothesis that language of testing is a variable that systematically affects how French Immersion students respond to test questions. This goal was approached by comparing Grade 6 French Immersion student performance on French and English forms of a standardized test of social studies achievement.

Importance of the Study

The need for an empirical verification of Carey's (1980) hypothesis can be seen in Cummins' (1983) comment that French Immersion programs have spread in Canada "not so much because they have succeeded in transmitting high levels of French proficiency to students at no cost to other academic skills, but because they have been *seen* to have succeeded" (p. 118). Given the apparent power that these studies have to shape educational programs in Canada, it is particularly important that valid inferences and conclusions be drawn from test data. A review of the literature indicates, however, that few if any Canadian studies of French Immersion student achievement identify language of testing as a variable that may confound data interpretation. The implications of this failure to control or account

for the effects of language of testing cannot be determined without a firm understanding of the impact of this variable on the nature of what is being measured.

Capell and de Porcel (1979) argue that proposed strategies for a second language program should be carefully scrutinized before they take on the institutional status of those routinely applied to monolingual programs. Their rationale for this assertion is that it is difficult to adjust any strategies, including inappropriate ones, once they have been instituted. It would seem, given Capell's and de Porcel's comments, that now is a particularly appropriate time to be addressing the question of whether or not language of testing is a variable that affects the scores of French Immersion students. Decisions about the nature of a French Immersion testing program are currently being made in Alberta (Student Evaluation & Records Branch, in press). An investigation of the effect that language of testing has on the scores of French Immersion students could provide needed direction, not only in the selection of appropriate instruments, but also in the development of data interpretation and reporting strategies.

Strictly speaking, the findings from this study will only be generalizable to situations where the same or similar testing instruments are being used. Nevertheless, the benefits to be gained by doing a study of the equivalence of French Immersion students' responses to French and English forms of a test are not limited to the Alberta situation. The results may serve to heighten the awareness of other researchers in Canada that language of testing has the potential to affect outcomes when French Immersion students are tested and thus should be considered when interpreting data or generalizing from one study to another.

Limitations Of The Study

This study does not address the question of whether or not standardized achievement tests that are blueprinted and fieldtested for use with regular English-language program students provide valid measures of achievement in French Immersion programs (whether those tests are presented in French or English). While this question is an important one, it goes

beyond the scope of the present investigation.

A number of variables related to language of testing that could affect outcomes are discussed in this study. The separate effect of these variables has not been isolated. As will be shown in the review of the literature, the overlapping nature of these variables, as well as technical limitations, preclude such a separation and analysis of effects.

The "degree of bilingualism" of the subjects in this study has not been measured. The lack of control of this variable may limit the generalizability of the findings.

Hypothesis To Be Tested

The hypothesis of interest to this study is that there will be a significant difference in French Immersion students' scores when they are tested in French as compared to English. The purpose of the next chapter is to determine a priori support for this hypothesis.

CHAPTER II

Review of the Literature

Introduction

Constructs such as social studies achievement are not directly measurable because they are not observable traits or behaviors. Instead, their presence must be inferred from test scores (Jones, 1971). These inferences about the attribute under study are based on an assumed relationship between the presence or absence of the construct being measured and the adequacy of responses to questions demanding the use of specific skills or items of information (Thorndike & Hagen, 1977).

If the assumed relationship between construct and response were a perfect one, then test scores would accurately reflect the attribute being measured and data interpretations would always be valid. Unfortunately, test scores often reflect information about traits or behaviors other than, or in addition to, the construct intended to be measured (Cronbach, 1971). As a consequence, assumptions about the relationship between the construct purportedly being measured and responses to test questions need to be challenged to ensure that data interpretations are valid. These assumptions are challenged through a process referred to as construct validation.

Construct validation begins with a claim that a given test measures a certain construct. The challenge consists of an attempt to prove a counterhypothesis. The counterhypothesis is an alternative explanation to account for test behavior in whole or in part. If the attempt to fit the data to the counterhypothesis fails, then the original hypothesis of what was being tested cannot be rejected (Cronbach, 1971). More importantly, it can be assumed that one's original inferences about the meaning of test scores are valid.

This study investigates the validity of the inferences one makes about levels of social studies achievement when French Immersion students are tested using French and English

versions of a standardized instrument. Its purpose is not to establish the absolute worth of the instruments as indicators of social studies achievement, but rather to determine the similarity of what is being measured by the two versions of the test. In other words, it is the construct equivalence of the two measures that is being studied.

The design of this study is similar to that which would be used to establish construct validity. First, an hypothesis is established about what is measured when French and English forms of a test are administered to French Immersion students. Then an attempt is made to prove a counterhypothesis. The hypothesis in this study is that the same body of knowledge and skills is measured when French Immersion students write French and English forms of a standardized test of social studies achievement. Since examinees' levels of social studies knowledge and skill remain constant, regardless of the language of testing, the operationalized form of this hypothesis is that their scores will be the same under both conditions of testing.

The alternative hypothesis to be examined is that the body of knowledge and skills assessed by the English form of the test differs, in whole or in part, from that which is assessed by its translated (French) version. In its operationalized form, the counterhypothesis is that French Immersion students' scores will differ under the two conditions of testing.

Carey (1980) alluded to two possible sources of difference in what is measured when French and English forms of a test are administered to French Immersion students. These sources of difference include (a) the effect of translation on the nature of test questions, and (b) the effects of first and second language reading abilities on test comprehension. Each of these conditions is examined below to determine if there is evidence to support the assumption that its effect is to change the nature of what is being measured by French and English forms of a test. Each section begins with an hypothesis about how the variable under study could cause what is being measured to differ under the two conditions of testing. Then the literature is examined to determine if there is support for this hypothesized source of difference.

Translation and Test Equivalence

Test Variables That Can Be Altered Through Translation

One of the variables that has a direct effect on the probability of selecting a correct answer to a test question is the quality or nature of that question. From this it follows that if, as a result of translation, the quality or nature of French forms of items is altered from that of the originals, then the probability of answering those questions correctly could vary. As a consequence of this variation, one's scores on a test could differ. For this reason, the quality or nature of English items and their French translations is of interest to this study.

Translation and Text Meaning.

One way of defining the quality or nature of an item is to consider the clarity of its meaning or purpose. Clarity of meaning is an essential attribute of an item because the writer's precision in selecting words is crucial in conveying the exact problem or task that the examinee must deal with. Often the choice of a particular word over a synonym can subtly change the meaning or emotional tone of an item (Bloom, Madaus, & Hastings, 1981). This in turn can affect how examinees respond to that item.

In considering the equivalence of the quality or nature of French and English forms of a test, a question that arises is whether it is ever possible to express the same meaning in two languages. This question is fundamental to this study because if the meaning of test questions is significantly altered through the process of translation then the probability of selecting correct answers to those questions could be affected. As a consequence, test scores could differ depending on the language of testing.

Language relativists hold that differences in the way various languages have come to encode meanings strongly influence the way in which members of that language group come to experience and know their world. Sapir (1961) wrote that the "real world" is to a great extent unconsciously built upon the language habits of individual groups. According to Sapir,

no two languages are ever sufficiently similar to be considered as representing the same social reality. As a result, "The worlds in which different societies live are distinct worlds, not merely the same world with different labels attached" (p. 69). From the relativist's point of view then, translation of meaning is not truly possible.

Others who have dealt extensively with the subject of translation hold a different view. That view is summed up in the following quotation from Katz (1972): "Natural languages are capable of providing a sentence to express any thought a speaker might wish to communicate. . . . For any example in English, a fluent speaker of any language could provide a parallel" (p. 12). Thus, from Katz's perspective the central conceptual meaning of utterances in one language can be translated into another language. This implies that it is at least theoretically possible to have French and English forms of a test that are parallel in meaning. It is necessary to apply the qualification of theoretical possibility to this assertion because the task of producing valid translations is a complex and difficult one.

One of the factors that makes translation a challenging task is that the meaning of a message is not "in the words". Thus the translator must discern not only what the *words* mean, but also what the *writer* means (Pergnier, 1978). As Graham (1985) notes, "the translator is under pressure not simply to produce a version of the original that reads or sounds well in the target language but also to understand and interpret the original masterfully so as to reproduce its message faithfully" (p. 37). What makes this interpretative phase of translation difficult is that words have no exact and constant equivalent in other languages (Pergnier, 1978). Because words have no constant equivalent, meanings often become distorted or blurred. Distortion occurs because the translator has to reconcile several possible meanings, including the author's intended meaning, the dictionary definition, and his or her own interpretation of a word or phrase (Duff, 1981).

To produce a faithful translation, the translator must do more than just interpret the original work accurately. She must also convey that meaning in a way that resembles the original writing. In short, the translator must maintain the style established by the author.

Maintaining that style is difficult because certain characteristics of one language are untranslatable into another. Often it is the syntactic requirements of the language such as specifying number or gender in nouns that cannot be translated (Scoon, 1974).

Translation and Item Equivalence.

Capturing and expressing the meaning of the source message of any piece of text is challenging enough for any translator, but much more is required when translating test items. Oller (1979) compares the task of translating test items to the translation of jokes, puns, riddles or poems. While it is the poetry in poems that must be captured when translating them, (Kolers, 1968) it is the meaning and the relationships among the stem and alternatives that must be preserved in test items. Preserving meaning and relationships is particularly important because these variables provide important cues to the test taker. Repetition of key words, grammatical inconsistencies, and unequal lengths among the alternatives all provide clues to, or pulls away from, the correct answer (Bloom, Madaus, & Hastings, 1981). Because these cues consciously or unconsciously shape how examinees respond to the test items, they can affect how difficult those items are. In essence, they shape the probability of selecting a correct answer.

Cues embedded in stems and alternatives of test items are not the only variables to affect the probability of selecting a correct answer. Test items are generally multidimensional in what they measure (Reckase, 1981). As a result they often tap skills and abilities other than those intended to be tested. In the case of multiple choice testing, it is difficult to obtain measures of factors independent of verbal comprehension because, no matter what skill or ability is being measured by an item, the achievement of a correct answer depends on proficient encoding and processing of the words and sentences of the test question (Horst, 1968; Nunnally, 1967). How proficient is one's encoding and processing of textual material depends, among other things, on the nature of the textual material being read, that is, on the text's readability (Duncan, 1986). From this it follows that one's comprehension of test

questions will depend, in part, on the readability of that material.

What this implies, in terms of this study, is that the responses of French Immersion students, to tests of social studies achievement, will be shaped not only by the meanings of, or cues embedded in, the items but also by the readability of the instruments used to measure that achievement. More importantly, it suggests that for students to respond equivalently to the two forms of the test, it is necessary for those tests to be equally readable. In short, the translator must not only have interpreted the meaning of the original test faithfully but must also have expressed that message in a way that is stylistically similar.

In summary, there are at least two ways that translation could alter the nature of test questions. First, the difficulty of items could be altered because of differences in the meaning of or presence of cues in the original and translated items. Secondly, the readability of the original items could be altered. Since readability affects how well students comprehend what is being asked by test questions, this alteration could affect the way students respond to those questions.

Understanding the impact of test translation on the nature of test questions is important to this study because if the probability of selecting a correct answer is altered through translation, then there is reason to question the null hypothesis that students' scores will not differ depending on the language of testing. Both of these issues are therefore examined in more detail below.

Translation and Item Cues

According to Oller, (1979) to achieve the required similarity in meaning and relationship when translating test questions one must maintain roughly the same style, the same usage of vocabulary and idiom, and comparable phrasing. Because of the problems inherent in translation, it is not always possible to achieve this similarity. The result is that for any item, translation will "produce (in principle and of necessity) a substantially different item" (p.93).

A study completed at the University of New Mexico investigated the feasibility of translating the *Boehm Test of Basic Concepts* from English into Navajo (Scoon, 1974). The scores of Navajo children who wrote the translated version were compared with those of an Albuquerque group of children who wrote the English version and with scores from the original Boehm norming group of English-speaking children. Scoon's results showed that the children who wrote the Navajo version had significantly lower scores than did either group who wrote the original (English) version. Based on these results she concluded that the Navajo form was not the same test as the original English version and from this that translation cannot be used to produce equivalent test forms.

Scoon's (1974) conclusion supports Oller's (1979) contention that translation produces a substantially different test. Her conclusion, however, may be a questionable one. Because a between-subject design was used, group differences and not item difficulty differences could have accounted for her findings. In other words, group variations in aptitude, experience, or even reading ability could have accounted for her findings. Consequently, it is difficult to accept her conclusion without further evidence.

Three English items and their French translations are presented below. These items demonstrate how subtle, yet potentially significant, changes can occur as a result of translation. They are presented and discussed here as a means of providing a context for interpreting Scoon's (1974) findings and as a way of judging whether or not it is possible to produce equivalent test items through translation.

The first set of items, presented in figures 2-1 and 2-2, are taken from a Grade 9 social studies achievement test (Student Evaluation Branch, 1987). They demonstrate how the meaning of an item can be changed in the process of being interpreted by the translator. The questions are based on four quotations and ask the examinee to identify which speaker fails to express a particular opinion. The English version asks the examinee to identify the speaker who fails to express an opinion about the *desirability* of using computers. The French version asks about the failure to express an opinion about the *advantages* of computer usage. These

Figure 2-1. Sample Item 1 (English)

SPEAKER I

Because of computer systems, it is now possible to monitor worker speed, accuracy, and length of rest periods. I favor the use of computers for two reasons: the number of managers needed to supervise work is reduced, and the problems with worker productivity can be identified more quickly.

SPEAKER II

With the continued automation of work, the skills and knowledge required to do the job are being transferred from the worker to the computer. Workers are reduced to watching machines. Work is becoming more monotonous, more routine, less challenging, and less rewarding. I think this is unhealthy.

SPEAKER III

Computer technology is changing the very nature of work. The result is that in some areas of the labor force, there is high unemployment as machines replace workers. In other areas there are skilled labor shortages. Significant adjustments to the labor force are needed to avoid a major crisis in the workplace.

SPEAKER IV

It is no longer necessary to assemble all workers at the same time and place. Portable computers create an office wherever the worker happens to be. The result is a lower expenditure of energy, time, and resources. You will never convince me that this is bad.

— Adapted from *Microtechnology*, 1982

15. Which speaker does NOT express an opinion about the desirability of using computers?

A. Speaker I
B. Speaker II
C. Speaker III
D. Speaker IV

Figure 2-2. Sample Item 1 (French)

INTERLOCUTEUR I
À cause des systèmes informatiques, il est maintenant possible de surveiller la vitesse et l'exactitude des travailleurs, et la longueur des temps de repos. Je suis en faveur de l'emploi des ordinateurs pour deux raisons: le nombre de chefs pour surveiller le travail est réduit et les problèmes que pose la productivité des travailleurs peuvent être identifiés plus vite.

INTERLOCUTEUR II
Avec l'automatisation permanente du travail, les aptitudes et les connaissances requises pour faire le travail sont transférées du travailleur à l'ordinateur. Les travailleurs en sont réduits à surveiller les machines. Le travail devient plus monotone, plus routinier, moins intéressant et moins valorisant. Je pense que c'est malsain.

INTERLOCUTEUR III
La technologie informatique est en train de changer la nature même du travail. Le résultat est que, dans certains domaines, il y a beaucoup de chômage parce que les machines remplacent les travailleurs. Dans d'autres domaines, il y a pénurie de main-d'oeuvre spécialisée. Des ajustements significatifs à la main-d'oeuvre sont nécessaires pour éviter une crise majeure dans le monde du travail.

INTERLOCUTEUR IV
Il n'est plus nécessaire de rassembler tous les travailleurs au même endroit et en même temps. Les ordinateurs portatifs créent un bureau là où le travailleur se trouve. Le résultat est une dépense moindre d'énergie, de temps et de ressources. On ne me convaincra jamais que c'est mal.

-- Adaptation de *Microtechnology*, 1982

15. Quel interlocuteur N'exprime PAS d'opinion sur les avantages d'employer des ordinateurs?

A. Interlocuteur I
B. Interlocuteur II
C. Interlocuteur III
D. Interlocuteur IV

items present different tasks to examinees because the terms "desirability" and "avantages" (advantages) have meanings that are quite different. Desirability refers to the attractiveness or advisability of an action or option (Websters, 1984) hence, in the English version, the examinee is asked to identify the speaker who fails to comment on this aspect of computer usage. The term "avantages" in the French translation of the item indicates that the student is to identify the speaker who fails to express an opinion about the *benefits* (Atkins, Duval, & Milne, 1987) accruing from this particular course of action.

The change introduced into the French version of the item is judged to be a function of interpretation and not of the ability of the target language to carry the meaning of the source language, because the French language includes the term "desirabilite" which, according to Atkins, Duval, and Milne (1987), translates to "desirability", the term used in the English version.

This change in word meaning has a direct effect on the correctness of the keyed response for this item. Of the four speakers, Speaker III is the only one who discusses an effect of computer usage without expressing an opinion about its desirability. Speakers I and IV both express positive opinions about the attractiveness or advisability of using computers. Speaker II offers the opinion that the use of computers is undesirable. The keyed response for the English version of the item is unquestionably alternative C. On the French version of the item, however, the keyed response is arguably either B or C, because neither Speaker II nor Speaker III discusses the benefits or advantages of using computers. In short, the keyed answer has changed due to differences in the wording of the item stems.

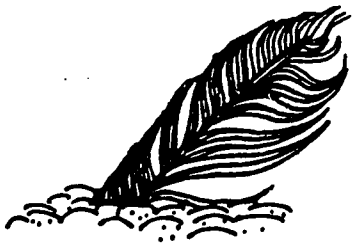
The next pair of items, presented in Figure 2-3 provide an example of how the nature of an item can be changed in the conveyance portion of the translation process. These items are taken from the test used as the criterion measure in this study (See Appendix). They form part of a provincial achievement test for Grade 6 social studies (Student Evaluation Branch, 1985a).

Figure 2-3. Sample Item 2 (English and French)

5. Which object found by archeologists would teach us the most about how people met basic needs?

A. A piece of volcanic rock

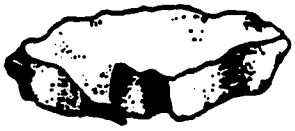
B. A rib from a buffalo



C. An eagle feather buried in sand

D. A needle made from a bone

5. Quel objet trouvé par les archéologues nous apprendrait le plus de choses sur la façon dont les habitants répondaient à leurs besoins essentiels?



A. Un morceau de roche volcanique

B. Une côte de bison

C. Une plume d'aigle enterrée dans le sable

D. Une aiguille en os

The portion of the items at issue is the keyed response, alternative D. The meaning of the correct alternative does not change from one version of the item to another. What does change is the wording of that alternative and this revised wording could affect how students respond to the item. The French version of the item has for alternative D, "Une aiguille en os" ("A bone needle") while the English version of the item says "A needle made from a bone". Since the critical feature that differentiates alternative D from the other three is that the object has been manipulated by man, the use of the word *made* in the English version makes this distinction more obvious then is the case with the French item. This item would predictably be more difficult in the French version.

The third pair of items, presented in figures 2-4 and 2-5, are taken from the achievement test used in this study. These items are based on six speakers' opinions about a law requiring the use of motorcycle helmets. What differs about the two items is the way in which the stem is worded. The term "comments" in the stem of the English question has been translated to "opinions" in the French item. Since the keyed answer is alternative A, "Opinions About the Helmet Law", the use of the word "opinions" rather than "commentairies" (comments) in the French stem is likely to make the answer to the French version more obvious. In other words, examinees are more likely to be cued to the correct answer if they are tested with the French rather than the English form of the item.

The differences illustrated by the three pairs of items discussed above support the assumption that the original meanings of, and relationships among, stems and alternatives can change when test items are translated. They also illustrate that item difficulty can be altered as a result of these changes. What can also be seen from these examples, however, is that the effect, on test difficulty, of these changes is not systematic. Two of the items are predicted to be *more difficult* in their translated form. The third item is predicted to be *easier* in its French form. Since support for Scoon's (1974) conclusion and the counterhypothesis in this study requires not only that items differ, but also that these differences yield results that are systematically different, this variation implies that one cannot assume that test scores will be

Figure 2-4. Sample Item 3 (English)

The law says that people riding motorcycles must wear helmets. Some people do not like this law and want to see it repealed (removed). Other people support the law.

These are some comments that citizens have made on this issue.

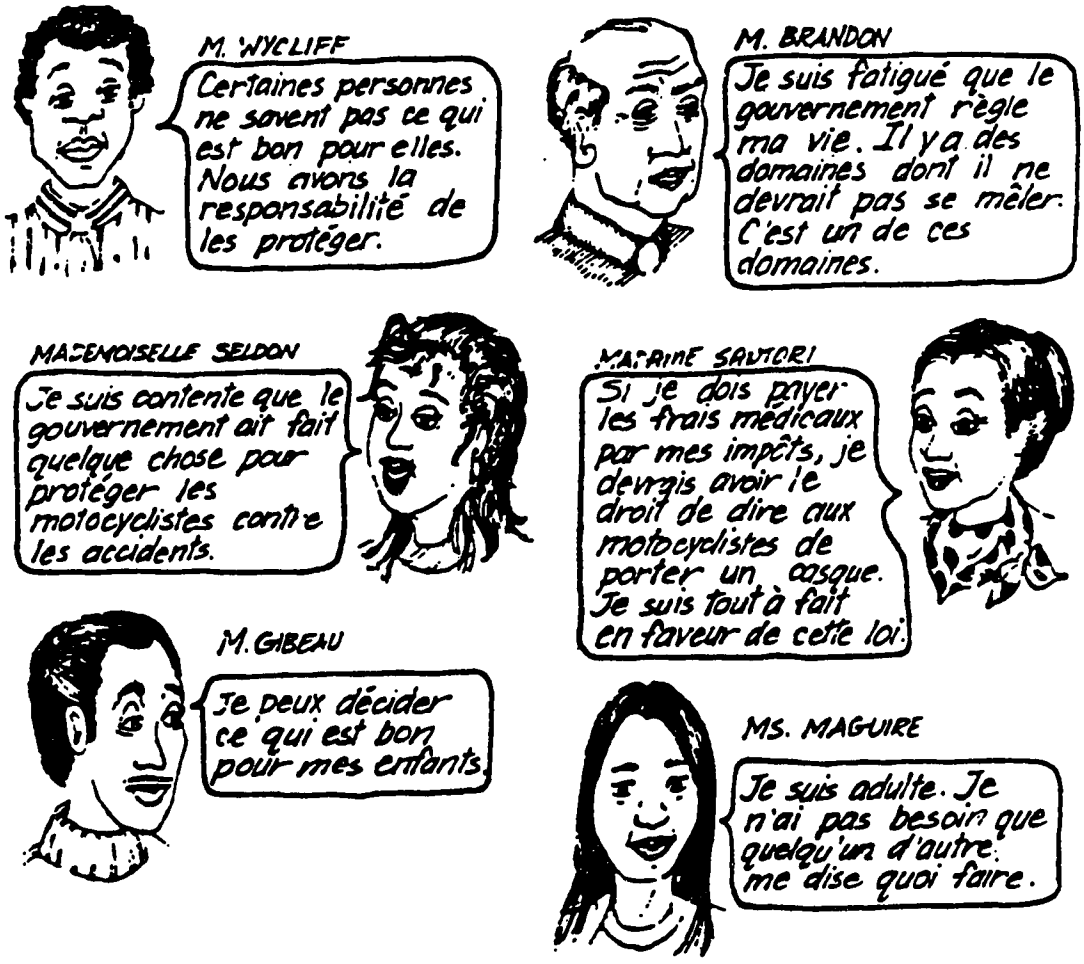
36. If all of the speakers' comments were being put on a chart, what would be the best title?

A. Opinions About the Helmet Law
B. Effects of Having the Helmet Law
C. Reasons for Keeping the Helmet Law
D. Persons Who Voted for the Helmet Law

Figure 2-5. Sample Item 3 (French)

La loi dit que les gens qui font de la motocyclette doivent porter un casque. Certaines personnes n'aiment pas cette loi et veulent la voir annulée (enlevée), d'autres soutiennent la loi.

Voici des commentaires faits par des citoyens sur cette question.



36. Si on rassemblait en un tableau toutes les opinions exprimées, quel serait le meilleur titre?

A. Opinions sur la loi sur le port du casque.
B. Effets de la loi sur le port du casque.
C. Raisons de maintenir la loi sur le port du casque.
D. Personnes qui ont voté pour la loi sur le port du casque.

constantly or systematically altered in the process of translation.

Translation and Readability

Researchers have determined that three broad classes of text-based variables influence reading comprehension. These variables include the physical characteristics of a piece of text, its content, and its linguistic style (Samuels & Eisenberg, 1981). Because these variables influence reading comprehension, then they must remain essentially unchanged for a source test and its translation to be equally readable. The likely effect that test translation will have on these variables is investigated below.

According to Samuels & Eisenberg, (1981) the physical variables that influence reading comprehension include such things as column widths, page and margin sizes, and the size and style of print. These variables can affect the speed of reading, the nature of eye movements and fixations, and the overall reading strategies used by a reader.

An examination of the two tests used in this study (See Appendix) reveals that in most ways the two forms are the same in terms of their physical characteristics. What does differ is the size and style of print. It seems unlikely, however, that this difference will have a significant effect on reader behavior because, according to Tinker (1966), most common typefaces are equally legible to an experienced reader. It is, therefore, possible to conclude that whatever physical differences exist between the two tests, these differences are unlikely to have a significant effect on examinee behavior.

A similar conclusion can be reached about the content variables that determine text, and therefore, test readability. Content variables include the specific subject matter of the text, the generality of the material and the abstractness of the material's presentation (Samuels & Eisenberg, 1981). While these variables are certain to have an effect on the difficulty of the tests under study, they are unlikely to be related to any differences in their relative ease or difficulty of comprehension. The reason that they are unlikely to be related to differences in difficulty is that the subject matter of the two tests will be the same, given a

highly qualified translator and a reasonable effort at translation.

The effect that translation may have on the third variable that affects reading comprehension, linguistic style, is less clear. Nida (1964) argued that when text is translated, no attempt should be made to preserve the original syntactic or semantic structure. Instead the message should be reduced to its kernel form and presented as a completely new utterance. Nida's assertion suggests that it is possible for the syntactic structure or choice of words used to express the original message to differ. Since word choice and syntactic structure are variables that have been correlated with ease of comprehension (McConnell, 1983), it could be that the translated version of a test is easier or more difficult to comprehend than the original form if the translator has introduced changes.

Dye (1971) conducted a study to determine the effects of translation on readability. He applied Flesch's (1948) readability formula to yield Reading Ease (RE) scores on the original (French) and translated (English) forms of sample passages taken from fourteen French originals and thirty translations of books and/or short stories. What Dye found was that scores consistently increased for text translated into English (higher scores indicate more easily read material). From this he concluded that the source documents became simpler to read when translated. He attributed the differences in RE scores to changes in the linguistic style of the passages as a result of translator changes.

Accepted at face value, Dye's (1971) conclusion lends support for the argument that translation alters levels of readability. Ironically, it is another finding from his research and the research he quoted that suggests Dye erred in his conclusion that the obtained differences in RE scores were an indication of changed levels of readability. Dye hypothesized that multiple translations of the same text would be reasonably consistent in their linguistic style because the style was predetermined by the original writer. He used the RE scores of the translations as indicators of linguistic style based on Klare's (1963) assertion that readability formulae measure difficulty of style. As predicted, Dye found that the RE scores of corresponding passages from four English translations of Voltaire's *Candide* were similar.

From this he concluded not only that the four translations were consistent in their linguistic style, but that this similarity was attributable to the fact that the translators had followed the style set by the original (French) writer.

What is noteworthy about Dye's (1971) discussion of his major findings is that he fails to see that his conclusions are illogical. On the one hand he argues that the style of the translations are similar because the translators were constrained by the style of the original writer. On the other hand he notes that the translations are not only equally readable but also consistently more readable than the original French version. The validity of Dye's second conclusion is dependent on the style of the translations being *consistently* different from that of the original. The reason that the translations have to be consistently different from the original in style is that linguistic style is the only variable that can cause readability to be the same across English forms but different from the French form, because in this case, the content of the message, the other critical variable determining readability, is constant.

The juxtaposition of Dye's (1971) two conclusions begs the following question. If the translators were sufficiently constrained by the style of the original writer to produce similar translations, why not sufficiently constrained as to produce translated text that parallels the source document? The only reasonable answer is that they were so constrained. In other words, in all likelihood the source and translated versions of the passages were stylistically similar and therefore, were, by definition, equally readable. What remains to be explained then, is why the RE scores of the source and translated versions were consistently different if their le els of readability were the same. The explanation appears obvious if one recalls that Flesch's Reading Ease Formula (1948) uses word length and average sentence length in words as its semantic and syntactic variables, respectively: What caused the RE scores to systemically and consistently vary was not stylistic differences but rather natural differences in the two languages. In other words, it simply took more and/or longer words to say in French what was said, in shorter form, in English.

An example of this phenomenon can be seen in the items presented in Figure 2-6. These items are taken from the tests used in this study. The translated item is faithful to the original in terms of sentence structure and word choice. In spite of this similarity in style, however, the stem and alternatives of the French form of the item are consistently longer than are those of the English version

Figure 2-6. Sample Item 4 (English and French)

12. In MOST early civilizations, wealth and power were

A. held mainly by the merchants and traders
B. held mainly by the nobles and priests
C. shared equally by the warriors
D. shared equally by all citizens

12. Dans LA PLUPART des civilisations du passé, la richesse et la puissance étaient

A. détenues principalement par les marchands et les commerçants
B. détenues principalement par les nobles et les prêtres
C. partagées également par les guerriers
D. partagées également par tous les citoyens

In conclusion, it appears that if translators have been faithful in maintaining the message of the source document, then the style of the original shapes the translation enough to maintain roughly the same linguistic style, at least as measured by counts of word and sentence length. Translators are unlikely to be any less constrained by the style of the original when translating tests than they are when translating prose. This implies that, except for minor variations in word choice or syntactic structure, the style of the English and French

versions of the tests should be similar. Research has indicated that text comprehension is influenced by a myriad of factors in addition to word choice and syntactic complexity (Koenke, 1987). As a consequence, minor text changes have little effect on readability. For example, Freebody and Anderson, (1983) have shown that a surprising number of difficult words have to be added to text before it becomes less readable. Similarly, Klare (1974-75) found that making sentences shorter did not necessarily lead to greater ease of comprehension. This suggests that the readability of a test is unlikely to be altered significantly as a result of minor word or syntactic changes introduced through translation.

One remaining issue requires resolution before it can be concluded that the readability of French and English forms of a test are likely to be approximately equal. This issue concerns the possible effect that the natural differences in the languages, referred to above, could have on comprehension. If the French form of a test consistently uses words of greater length than the English form, then word length as a variable (irrespective of its correlation to word meaning) must be ruled out as a determiner of comprehension to conclude that the tests are equally readable.

Evidence that word length does have an effect on reading behavior comes from the research related to eye-span behavior. When reading text, the eye moves in a series of discrete fixations with fast movements (saccades) in between (Just & Carpenter, 1987). Information is abstracted from text during these fixational pauses (Rayner, 1981). Just and Carpenter have shown that the time spent fixated on a word is directly related to its length; an average of 30 milliseconds more is spent on a word for each letter it contains. As well, Rayner has shown that the length of the word to the right of the word currently fixated influences the length of the following saccade.

The fact that every additional letter affects gaze duration and location implies that reader behavior may differ, at least in this way, when French Immersion students respond to tests in French as compared to English. It does not, however, prove that their level of comprehension will differ because of these differences in behavior.

To conclude that word length has a direct effect on the quality of text comprehension, evidence is needed that processing efficiency varies as a function of the number of letters in a word. This assertion is based on the widely held and tested assumption that one of the variables that determines one's degree of comprehension of textual material is the quality and efficiency of text processing in short-term memory (e.g., Jackson & McClelland, 1981; Perfetti, 1985).

Evidence suggests that it is unlikely that processing efficiency in short-term memory varies as a function of the number of letters in a word. Based on his research, Johnson (1981) concluded that featural characteristics of words (i.e., letter encodings) are the unit of representation only within the perceptual system of processing; the unit of representation within the cognitive system (i.e., working memory) is the encoded word. This point is significant because it suggests that the additional letters in the French text will only have an effect on reader behavior at preliminary stages of text processing. In short, while perceptual processing may be more complex for examinees reading a French translation than an English original, their higher order processing, including lexical access and semantic analysis, will occur only after the signals have been recoded into units which make the differences in letter counts irrelevant. Thus working memory capacity will be under no more strain when processing the longer French text than the shorter English version, assuming that any difference in length is a function of natural differences in the languages and not a result of differences in the linguistic style of the two tests.

In conclusion, there is little evidence to support the hypothesis that the French test will be systematically more or less readable than the original English version because of natural differences in the languages. A review of the two tests to be used in this study shows that in most cases, the French version of an item is longer than its English original. This extra length of the French text appears to be function of natural differences in the languages rather than due to stylistic differences, a factor than could have affected readability. Since letter encodings cease to be the unit of representation once text processing occurs within the

cognitive system, natural language differences should not affect text readability.

Conclusion

The purpose of this section was to determine if there is evidence to support the notion that scores could vary, depending on the language of testing, because of differences in the nature of the instruments being used. It was hypothesized that test scores could differ if test variables that have a direct effect on the probability of selecting a correct answer were altered as a result of the translation process. An assumption in this hypothesis was that these alterations would have to have a systematic effect on test difficulty for scores to be significantly different across forms. Two factors were examined: the effect that translation has on the meaning or cues provided by a test question and the effect of translation on readability.

A search of the literature indicated that neither of these factors has undergone much empirical investigation. Thus, any conclusions that can be drawn have to be based as much on logic as on hard evidence. Consequently, the conclusions are at best tentative. Nevertheless, based on what has been presented here it appears that there is insufficient evidence to support the notion that differences between carefully translated instruments will have a systematic effect on test outcomes.

Reading Ability Equivalence

The Relationship Between Reading Ability and Test Outcomes

Textual variables are not the only ones to affect the quality of comprehension processes. As Adams (1980) notes, the efficient operation of the system depends as much on the information in the reader's mind as on the information in the text. The reader plays a key role in text processing because the meaning of the discourse is something more than can be derived from a linguistic analysis of the text. According to Spiro (1980) what language

creates is a skeleton or a blueprint for the creation of meaning. It is the activity of the reader who by making an "effort after meaning" constructs a product that "makes sense within his or her individual view of the world" (p. 250).

The meaning that a reader constructs from text is shaped by two factors: the reader's knowledge (including world and language-related knowledge) and his or her skill at using that knowledge (Just & Carpenter, 1987; Perfetti, 1985). Readers differ in terms of these variables. Some readers have larger vocabularies and greater knowledge of lexical relationships than do others. Similarly, practised readers are distinquished by their greater ability to use the whole context to decode the meaning of unfamiliar words (Cooper, 1984), an ability that affects the comprehension process.

Because readers possess varying degrees of knowledge and skill, there are individual differences in what is comprehended from the same piece of text (Underwood, 1985). This implies, in the present context, that some test-takers will better understand the written form of a test question than will others. It also suggests that failure to perform adequately on an item could be more of a function of poor reading comprehension for some test-takers than for others. As a consequence, what a test question really measures (i.e., reading comprehension or the intended construct) will vary, depending on the knowledge and skill of the reader.

The processes involved in second language reading are similar to those required when reading in a native language (Block, 1986; Woytak, 1984). Readers construct meaning from text based on their level of knowledge and skill in that language. It follows from this that the more fluent is one's second language knowledge and skill, the greater or richer will be one's understanding of what is read in that language. Given this similarity between first and second language text-processing, it is possible to draw inferences about second language test-taking that parallel those presented above: Some second language test-takers will be better able to comprehend textual forms of test questions than will others and thus will be less likely to fail questions because of poor understanding. The implication is that the construct being

measured by test items will vary, depending on the level of knowledge and skill of the reader.

A question that arises in relation to this study is whether it is possible for an item to vary in terms of what it measures, depending on whether it is presented to an examinee in his or her first or second language. If French Immersion students are unequal in their ability to read in their two languages then it is possible that they will derive different meanings from French and English forms of a test, and therefore could respond differently to those questions. The remainder of this section is used to examine information related to first and second language knowledge and skill proficiency so that conclusions can be drawn about the first and second language reading abilities of French Immersion students.

Language-Related Knowledge

Much of the information that makes text understandable resides in the world knowledge shared by the writer and reader (Tighe & Hadaway, 1986). World knowledge includes such things as awareness of peoples' needs, wants, motivations, attitudes, plans, and values, and knowledge of specific content domains (Just & Carpenter, 1987).

In spite of its importance to the reading process, world or background knowledge is not a variable of interest to this study. It is not of interest because it is assumed that the background knowledge required to achieve comprehension will be constant across test forms, as the tests being used contain the same content, albeit in different languages. Moreover, it is assumed that the background knowledge of the examinees will be constant since a split-half design will be used.

What is of interest to this study is the body of knowledge required to read the test forms that is referred to as language-related knowledge. To understand written language, a reader has to encode the words and access their meanings in his internal lexicon (Just & Carpenter, 1987). The internal lexicon is a person's mental representation of word meanings (Underwood, 1985). Because lexical access plays such an important part in text processing its relationship to reading comprehension is one of the most robust and best documented

relationships in reading research (Just & Carpenter, 1987; Stahl, 1983). There is more to language-related knowledge than an awareness of what words mean, however, for as McKeown, Beck, Omanson & Perfetti (1983) note "A difference exists between acquiring knowledge of a word's meaning and knowing the word well enough to aid comprehension" (p.4).

What this difference entails is aptly described by Richards (1976). He lists a number of characteristics of "knowing" a word well enough to aid comprehension including: (a) knowing the probability of encountering that word in speech or print, (b) knowing the limitations imposed on the use of a word according to the variations of function and situation, (c) knowing the underlying form of a word and the derivations that can be made from it, (d) knowing the associations between a word and other words in the language (e.g., synonym, subordinate, and coorderinate relationships), (e) knowing the semantic value of a word, and (f) knowing the many different meanings associated with a word.

What Richards' (1976) list includes is not only the knowledge that is required to achieve lexical access but also the knowledge that is needed for syntactic analysis to occur. Syntax allows words to form higher order constituents such as phrases or clauses that provide part of the temporary structure required to organize words in memory until the underlying concepts are understood (Just & Carpenter, 1987). Syntactic analysis relies on cues in the text to indicate how words should be grouped into syntactic constituents. These cues include such things as word order, word class, function words, affixes, word meanings and punctuation. Awareness of what information is conveyed by a cue is referred to by Just and Carpenter as procedural knowledge, "a representation of the appropriate mental actions to be taken under a given set of circumstances" (p. 145).

The Transfer Of Language-Related Knowledge Across Languages.

In discussing the learning of a foreign language Beheydt (1987) pointed out that there is rarely a one-to-one match between the meanings of words in one language and the

meanings of words in another language. He concluded that in this respect, learning the vocabulary of a second language is really the acquiring of a new conceptual system along with the new verbal labels. What Beheydt's observations suggest is that the range and richness of understanding that French Immersion students have for English words does not automatically transfer when they learn the equivalent French verbal labels for words. Instead, to acquire that range and richness of meaning in their second language, French Immersion students must replicate their first language learning experiences.

Procedural knowledge, like lexical knowledge, is to a certain extent of limited transferability across languages. Because languages vary in the cues they use to signal the appropriate mental actions to be taken when reading, the expectancies set up by the reader when sampling syntactic clues in text must be related to one's knowledge of the structure of that language (Berman, 1984; Cowan, 1976; Just & Carpenter, 1987). To the extent that the languages are similar in structure, transfer of knowledge is facilitated and possible (Alderson, 1984). Confusion can occur, however, when predictions based on knowledge of the native language are used inappropriately when reading second language text. Yorio (1971) refers to this inappropriate transfer as language interference.

In the present situation where the languages of interest are fairly similar in structure there is likely to be a reasonable amount of knowledge transfer. Nevertheless, there are differences in the languages and thus to be equally fluent in their syntactic analysis of the two languages, French Immersion students must have frequent and varied exposure to both languages.

In conclusion, because there is limited transferability of lexical and procedural knowledge across languages, it is possible that French Immersion students will have unequal levels of language-related knowledge in French and English. If so, then they may have more trouble comprehending test questions presented in one language as compared to the other. It is therefore important to assess their relative levels of French and English language-related knowledge.

Assessing Language-Related Knowledge Equivalence.

The assessment of language-related knowledge is not easily achieved. Beheydt (1987) argues that the semantic values of words are only specified by their relatedness to and difference from words with adjacent meanings. What this implies is that to assess one's understanding of a word one must measure that understanding in relation to other words. In short, it is not adequate to assess language-related knowledge by having students select or provide adequate dictionary definitions of words because, according to Bussis and Chittenden (1987), simple tests of vocabulary recall cannot capture the full range and richness of meaning that a reader has for words in his or her lexicon. The process is even more complicated if it is the equivalence of language-related knowledge across languages that one is attempting to assess. The problems associated with this kind of assessment are obvious. Inequivalencies in the testing instruments or in examinees' abilities to articulate responses could confound the estimates of their actual level of knowledge in each language.

In discussing the evaluation of vocabulary understanding, Simpson (1987) noted that the issue in this type of assessment is not whether students know the words or not, but rather in what ways they know them. Simpson's comment hints at a method that can be used to infer the level of knowledge that French Immersion students have about their first and second languages: One can assess how French Immersion students have come to know what they know about their two languages. In short, one can assess their language acquisition histories. The rationale for this assumption is presented below.

The process of acquiring the procedural and lexical knowledge needed to achieve comprehension is a long and complicated one. Readers may initially know only some general features of a word, but over time they acquire a much more detailed representation of its meaning and usage. Developing a rich semantic and syntactic understanding of words requires much more than just time however. Progressive differentiation of word meaning and usage comes with frequent exposure to words in a variety of contexts (Anderson & Shifrin, 1980; Beheydt, 1987; Just & Carpenter, 1987; McKeown, Beck, Omanson, & Perfetti, 1983;

Simpson, 1987). This frequent and varied exposure is necessary because the "family of potential meanings" (Anderson & Shifrin, p. 332) that are associated with words are only articulated through contextual experience. This process is referred to as instantiation (Anderson and Shifrin, p. 334) or semantization (Beheydt, p. 55).

Because the meanings of words and the cues that they provide vary across languages, and those language specific meanings and cues are only acquired through frequent and varied contextual experiences, it follows that the quality and quantity of those experiences will have a direct relationship to the language-related knowledge that French Immersion students possess in each of their two languages. From this it can be inferred that the more equivalent is their first and second language acquisition histories, the more similar will be their levels of language-related knowledge. Consequently, one can predict their level of lexical and procedural knowledge in their two languages by assessing their experiences with those languages.

First and Second Language Acquisition Experiences.

West (1985) found that the parents of French Immersion students have an extraordinary degree of energy, enthusiasm, and commitment in regard to their children's education. They also have a higher socio-economic status, have greater confidence in their children's academic ability, and spend more time reading to their children than do parents of children in regular language programs (Carey, 1984). These characteristics are similar to those associated with high academic and reading achievement in English-language unilingual children. The parents of high academic reading achievement unilingual children are described by Friesen (1987) as having high aspirations and expectations for their children's achievement, considerable verbal interaction with them including time spent reading to them, and active involvement in their children's school programs.

Based on this similarity in parental characteristics it is possible to infer that French Immersion students will have a rich and varied experience with English, their mother tongue.

This does not appear to be the case when it comes to their experiences with French. When surveyed, parents and teachers from the French Immersion program of a large urban jurisdiction in Alberta expressed concern that their children had little opportunity to use their French language skills outside of the school (Acheson, 1986). Their concern seems legitimate for two reasons. First, the first and dominant language of 88 percent of the children in this French Immersion program was English. Only 7 percent of the children came from homes where French was the language currently spoken. Second, research shows that French Immersion students are more likely to read, watch television, and communicate with peers and adults in English rather than in French, when they are out of school (Cummins, 1987; McEwen, 1984; Swain & Lapkin, 1981).

That this discrepancy in the range and variety of experiences that French Immersion students have with their two languages will result in unequal development of those languages is evident from Bain's and Yu's (1987) comments about the Francophone experience with language in Western Canada. They note that even when Francophone parents speak French to their children "by preschool age the *lingua communis* . . . has become so dominant that it is but sentimental fiction to consider the language first spoken as the 'mother tongue'" (p. 221). If Francophones cannot maintain balance in their bilingualism, one wonders how French Immersion students from Anglophone backgrounds can be expected to.

In conclusion, it appears that in terms of their out-of-school experiences, French Immersion students are predominantly developing first language knowledge. The question then is whether their experience in a French Immersion classroom is sufficient to equalize their levels of L1 and L2 language-related knowledge. The answer, in the opinions of Carey (1987) and Swain (1974) is no; when second language learning is limited to school experiences, students rarely achieve a native-like command of that language. Part of the reason that classroom experience alone is insufficient to equalize French Immersion students' knowledge of French and English appears to be related to the quality of language children have in that setting. Different research studies have concluded that there is a considerably

higher proportion of teacher led lessons and much less small group work in Immersion programs than in regular English language programs (Cummins, 1987). As a result, students have little opportunity to use French in the classroom. Moreover, Chaudron (1983) has found that teachers faced with non-native speakers make greater efforts to simplify language than they would in a regular language classroom. This linguistic simplicity involves "less varied, more common and structurally more elemental or regularized material" (p. 128).

Children's experiences with written materials in French Immersion classrooms do not appear to be optimal either, according to the results of a survey conducted by Acheson (1986). He found that the limited availability of curricular materials in French was a paramount concern of teachers and principals. The existence of a shortage of French language resources was confirmed in an Alberta Education publication (Language Services Branch, 1985). In that document it was stated that "It will not surprise anyone that the French edition of approved English resources in not always available. Other appropriate French resources must consequently be identified to ensure that program objectives are met. At times, such resources cannot be found" (p. 13). Statements such as these strongly argue that French Immersion students are either using English resources or few resources at all. As a result, it is unlikely that students will have acquired as rich an understanding of the French language as they will have of English, their mother tongue and the language of the community.

Conclusion.

Vorhaus (1984) argues that the daily use of first language readers' own language across all communicative situations provides them with the advantage of being able to concentrate on comprehending related ideas and concepts represented by the words they are reading. In her view, however, readers in a second language feel constrained by their limited knowledge of vocabulary and grammatical rules and concentrate on segmenting meaning into understandable linguistic information. The effect, on comprehension, of these differences is

aptly described by Vorhaus:

> First language readers are interactors who use the author's language as a basis for developing concepts and an understanding of the author's idea while second language readers are mostly receivers who are constantly trying to develop more linguistic knowledge and insights about that particular author's language. . . . The first language reader has the linguistic resources that allow enough mental flexibility to understand what the author is *conveying*, while the second language reader can only use the available linguistic information to understand what the author is *saying*. (p. 413)

Given the previous description of French Immersion students' first and second language acquisition histories, it seems apparent that their communicative abilities in these languages will parallel those described by Vorhaus (1984). Simply put, French Immersion students will be less able to comprehend text presented in French than in English. Because comprehension is assessed each time students respond to multiple-choice questions it seems apparent that their unequal levels of comprehension will have an effect on how well they answer those questions. In short, there appears to be a compelling reason for arguing that their responses will vary, depending on the language of testing.

Skill Proficiency in First and Second Language Reading

In discussing the notion of language proficiency, Ingram (1985) argued that *knowledge* and *proficiency* are not the same thing. He noted that one can have considerable knowledge about a language including awareness of its grammatical rules and cues and yet not be proficient in the sense of being able to utilize that knowledge readily for practical communication purposes. A parallel distinction exists when it comes to reading. Even given an excellent command of the language, a reader will not achieve comprehension of text in the absence of proficient use of the skills that underlie the reading process (Just & Carpenter, 1987; Perfetti, 1985).

Reading skill proficiency is an issue of concern to this study because it has been consistently shown that foreign language readers perform more slowly in their second language than in their first language, for reasons not related to their knowledge of that language (Alderson, 1984; Favreau, Komoda, & Segalowitz, 1980; Favreau & Segalowitz, 1983; Woytak, 1984;). This slower second language reading rate suggests, at the very least, that when French

Immersion students write tests in their second language, they will require more time to process the test questions than if they had written the test in English. Thus some students who would be able to complete the English form of a test within the required time limits may be unable to complete the French form. This could artificially depress their scores on the test.

Their slower speed of reading may have a more deleterious effect on their test-taking ability than just an increase in testing time however. It may also affect how well they are able to comprehend the test questions. MacNamara (1967) found that the Irish-English bilingual students he tested were not only reading in their second language at a slower rate but also with lower comprehension. This lowered comprehension was not directly related to their level of language-related knowledge. Even when they understood the words and structures of the text under study (their understanding of the words and structures were tested separately) they were still less able to comprehend what they had read in their second language.

MacNamara (1967) assumed that his subjects' comprehension difficulties occurred because they required greater time and attention when decoding the semantic value of words in their second language. His hypothesis was that this increased time and attention added a burden to short-term working memory, thereby making it difficult for them to recall other parts of the message they were reading.

MacNamara's (1967) assumption that slower second language reading is a function of less efficient, more attention demanding, lower level text processing is supported by the findings of Favreau and Segalowitz (1983). They conducted a study with bilingual readers that was concerned with the use of automatic and controlled processing in a lexical decision task. Their results showed that bilinguals with *equal* first and second language reading rates responded in ways that suggested automatic processing in both languages. Bilinguals with *slower* second language reading rates showed a pattern of reaction times that suggested automatic processing in their first language but controlled or attention demanding processing in their second language. That this slower, attention demanding type of processing could cause second language readers such as MacNamara's to achieve poor comprehension fits with what is known about reading processes.

The relationship between automatic and controlled processing and skilled reading is a well documented one. Skilled reading involves the interactive processing of information from a number of information sources (Frederiksen, 1981; Perfetti, 1985). A fundamental component of this interactive processing system is the short-term or working memory (Masson & Miller, 1983). The working memory is where all of the information from the various sources is combined as evidence for or against hypotheses about meaning (Levy, 1981). Working memory is a limited capacity processor in terms of the amount of information it can process at any one time (Fletcher, 1981; Spiro, 1980). It is also limited in terms of the duration of time that traces can be held without active rehearsal (Lesgold & Perfetti, 1981).

Because of the limited capacity of short-term working memory, efficient comprehension can only be achieved by reducing competition for attentional resources among the component processes of reading. Competition is reduced by automatizing as many of the component processes as possible (Laberge & Samuels, 1974; Perfetti, 1985). The measure of automaticity is the extent to which an activity can be performed at the same time as a second activity to which attention must be directed (Underwood, 1985).

Not all of the component processes of reading are subject to automaticity of execution. For example, attention is demanded continuously if one is to integrate the meanings of individual words into a structure that corresponds to the underlying meaning of the text being read (Underwood, 1985). Similarly, a reader must consciously retain at the end of a segment of text what he or she read at the beginning for adequate comprehension to occur (Conrad, 1972; Curtis & Glaser, 1983; Masson & Miller, 1983). What is subject to automatic execution are the lower level activities of reading such as letter and word encoding and lexical access (Just & Carpenter, 1987). However, not all individuals are equal in their ability to perform these lower level activities automatically (Fredericksen, 1981; Levy, 1981; Samuels, 1987). In such cases, where increased attention must be allocated to specific lexical operations such as decoding, the higher-order processing of extended textual segments is jeopardized, resulting in poorer comprehension (Frederiksen, 1981; Laberge & Samuels, 1974;

Perfetti, 1985).

While it is clear from this description of the reading process that MacNamara's (1967) hypothesis is a valid one, one point requires clarification before it can be concluded that French Immersion students will be similarly disadvantaged when reading in their second language. That point relates to the fact that Favreau's and Segalowitz's (1983) subjects did not show uneven comprehension in their two languages even when their reading rates in those two languages differed. In other words, unlike MacNamara's subjects, their relatively inefficient lower order text processing did not appear to affect their understanding of what they read.

Favreau's and Segalowitz's (1983) subjects differed from MacNamara's (1967) in a very important way. Their subjects were fluent adult bilinguals who "read in each language at rates well within the range of normal monolingual readers" (p. 573). MacNamara's subjects were school aged children who were found to be "weaker than monolinguals in the monolinguals' language which was . . . the language of instruction" (p. 122). These descriptions suggest two reasons why the two groups of subjects differed in their abilities to comprehend what they had read. First, while Favreau's and Segalowitz's unequal reading rate subjects showed less automaticity in their second as compared to their first language, their overall efficiency may have been sufficient to permit adequate attention to what they were reading in their second language. In other words, their lower order abilities may have been sufficiently automatized as to free working memory capacity for the execution of the required higher-order processes.

The second point is that Favreau's and Segalowitz's (1983) subjects, unlike those of MacNamara (1967), were fluent in their second language, a fact which indicates that their lexical and procedural knowledge was well developed. This point is significant because it suggests that their higher level text processing abilities may have been sufficient to compensate for deficiencies at lower levels. According to Stanovich (1980) the processes of reading are not only interactive but also compensatory. This compensatory aspect of reading leaves open the possibility that higher level processes can actually compensate for deficiencies in lower

level processing. If, as Levy (1981) suggests, there is a threshold strength for determining when comprehension is achieved, then this threshold could have been achieved for Favreau's and Segalowitz's subjects through strong top-down support. MacNamara's subjects, on the other hand, may have had insufficient language-related knowledge to permit top-down compensation for weak lower level processing, with the result that their comprehension was incomplete in their second language.

Given their language acquisition histories, French Immersion students are more likely to have achieved the level of language-related knowledge acquired by MacNamara's (1967) subjects than that achieved by Favreau's and Segalowitz's (1983) subjects. Thus, it seems reasonable to conclude that, like MacNamara's subjects, French Immersion students will pay the price, in terms of comprehension, for their slower second language reading rate. That is, their comprehension of French text will be poorer than their understanding of English text. As a consequence, their ability to respond to French forms of a test will in all likelihood be more constrained by the need to read the test questions than it would be if they were taking the test in English. The predicted result is that their scores will be depressed, relative to their scores on an English version, when they respond to a test in French.

Summary and Conclusions

The purpose of this chapter was to examine support for the hypothesis that there will be no difference in French Immersion students' scores when they are tested in French as compared to English.

Two factors that affect how well students respond to test questions were examined: (a) the nature of the test questions and (b) the reading ability of the examinees. These factors were assumed to be relevant because if either of them were to vary with a change in the language of testing, then test scores could be affected.

In the case of the test questions, it was hypothesized that variables in the English items that affect item difficulty could be altered in the process of translating the questions into French. As a consequence of this variation, the probability of selecting a correct answer

could differ, depending on the language of testing.

From the literature it is apparent that the task of translating test items is a complex and difficult one; the translator must not only interpret the source message faithfully, but must also capture and then convey the linguistic and extralinguistic contexts that the text of an item calls to mind (Oller, 1979). These linguistic and extralinguistic contexts are fundamental to the nature of an item because they contain cues that affect examinee behavior. Similarly, the translator must also preserve the linguistic style established by the original test developer because examinee behavior is also determined by the readability of the items.

There has been very little research reported in the literature concerning how well translators are able to maintain the cues in, and readability levels of, source items when they translate them. Because of this absence of reported results, the possibility that the difficulty of test items will be altered through translation cannot be ruled out. Nevertheless, what little evidence is available suggests that the difficulty of test items will be affected by the translation process in a random rather than a systematic way.

In terms of the second factor, reading ability, it was hypothesized that French Immersion students could be unequal in their ability to read text presented in French and English. This unequal level of reading ability in their two languages could affect how examinees respond to test questions because reading comprehension is one of the factors being measured by paper and pencil tests of social studies achievement.

It was assumed that to achieve equal levels of comprehension of test questions presented in French and English, French Immersion students would have to have equivalent levels of language-related knowledge and skill proficiency in those two languages. Studies have shown that to have equal levels of language-related knowledge and skill proficiency in their two languages, readers must have had comparable experiences with them. Evidence suggests that Grade 6 French Immersion students in Alberta are unlikely to have had equivalent contextual experiences with French and English. This implies that their ability to comprehend test questions presented in French and English will differ. From this it is

possible to infer that their test scores could differ, depending on the language of testing.

CHAPTER III

Questions and Hypotheses

The major purpose of this study was to determine if French Immersion students performed in the same way when responding to French and English forms of a standardized test of social studies achievement. Two experiments were carried out. The research questions and hypotheses for each experiment are presented below.

Experiment 1

Questions

In Experiment 1 two research questions were studied. These were:

1.1 Do French Immersion students achieve similar scores when they respond to French and English forms of a standardized test of social studies achievement?

1.2 Are French Immersion students equally able to complete French and English forms of a standardized test of social studies achievement within the time limit established for the test?

These two questions were addressed by randomly assigning French and English forms of a standardized test of social studies achievement to Grade 6 French Immersion students. All of the standardized conditions of administration established for this test by its developers were followed, including the time limit. The responses of the group who wrote in English (E group) and the group who wrote in French (F group) were compared to determine if their performances were similar in terms of both their scores and their rates of completion.

Hypotheses

Three hypotheses were tested in Experiment 1. The first two pertained to Question 1.1. The last hypothesis was related to Question 1.2. These hypotheses and the rationale for each are presented, separately, below.

Question 1.1.

Research indicated that an examinee's performance on paper and pencil tests is shaped by his or her level of reading comprehension (Horst, 1968; Nunnally, 1967). This implied that the performances of French Immersion students, on tests presented in French and English, would be shaped by their abilities to comprehend text presented in those two languages. The research cited in Chapter II also indicated that the French Immersion students who would be tested in this study were likely to have less language-related knowledge and skill proficiency in their second (French) as compared to their first language (English). From this it was possible to predict that they would have unequal levels of comprehension when they read French and English text and thus that they would perform differently when responding to French and English forms of a test. Based on this prediction, the following hypothesis was put forth:

Hypothesis 1.1.1: The scores of examinees who write the English form of the test will be greater than those achieved by students who write the French version.

Research suggested that, as text difficulty increases, less skilled readers pay a proportionately higher price in processing efficiency than do skilled readers (Frederiksen, 1981). Since it it has been shown that reading comprehension is directly related to the efficiency of text processing in short-term memory (Laberge & Samuels, 1974; Perfetti, 1985), it was possible to infer that the greater the difficulty of the text and the less skilled the reader, the poorer would be his or her comprehension of that text. From this it was inferred that the greater the complexity of the text to be read on a test, and the more limited the language-related knowledge and skill of the reader, the poorer would be his or her comprehension of the test questions and the more likely he or she would be to answer the questions incorrectly.

It was concluded in the previous chapter that the French Immersion students in this study would have less language-related knowledge and skill in their second language (French) as compared to their first (English). As a result, they would be more like unskilled than skilled readers when reading in French. Given the aforementioned relationship between reader ability and text difficulty, it was possible to infer that as text difficulty increased, the students in the study would pay a proportionately higher price when reading text in their second language as compared to their first.

The questions on the social studies test that would be used in this study were of two different types: those that assessed recall and comprehension of previously learned information (knowledge-based items) and those that assessed the ability to process text-based information (skill-based items). The skill-based items differed from the knowledge-based items in two specific ways. First, unlike the knowledge-based items which "stood alone", the skill-based items were accompanied by graphic and/or textual data that needed to be read and interpreted for the questions to be answered. Second, the skill- but not the knowledge-based items contained information or content that was novel to the examinee. In short, skill-based items were more complex. They were, therefore, more difficult to comprehend than were knowledge-based items.

What this suggested, given the assumptions made above, was that (a) the processing efficiency of all examinees would be more taxed when reading skill-based as compared to knowledge-based items, and (b) that less skilled readers would have relatively more difficulty comprehending data-based questions than would those who were more skilled. In terms of the French Immersion students being tested in this study, this suggested that all examinees would have more difficulty reading data-based than discrete items, but that F group examinees would have relatively more difficulty than would E group students. Since performance on test questions would be related to one's ability to read and comprehend the questions, it was possible to infer that the performance of examinees in this study would be affected by the differing levels of readability of data-based and discrete items. Based on this assumption, the following hypothesis was put forth:

Hypothesis 1.1.2: Group differences in scores will be greater on data-based items than on discrete items.

Question 1.2.

The Grade 6 social studies achievement test used in this study was a power test which was designed and developed to assess the achievement of English language program students in Alberta (Student Evaluation Branch, 1984). Since the majority of students in this population were native speakers of English, it could be argued that the time limit established for the test was that which was appropriate for first language speakers of English.

Research indicated that foreign language readers typically require more time to read text presented in their second language as compared to their first (Alderson, 1984; Favreau, Komoda, & Segalowitz, 1980; Woytak, 1984). This suggested that the French Immersion students participating in this study would read French text more slowly than they would read text presented in English. It also suggested that the speed with which they would be able to read tests presented in French would be less than that in English. This suggested that they could be less able to complete French as compared to English forms of the test. Based on this assumption, the following hypothesis was put forth:

Hypothesis 1.2.1: The completion rate will be greater for examinees in the E group than in the F group.

Experiment 2

Questions

One must read and reflect on test questions to answer them correctly. This process requires time. If F group students in Experiment 1 ran out of time because of their slower reading speed, then they would have been unable to attempt questions that they may have been able to respond to correctly had they done the test in English. As a result, their scores could have been depressed relative to what they would have been had they written the test in English or had they had time to read and respond to all of the questions. This implied that all or a part of the language of testing effect predicted in Experiment 1 could have been the

result of F group students' pressure or inability to complete the test within the established time limit.

Experiment 2 was undertaken to examine this hypothesis. All of the conditions of Experiment 1 were replicated in Experiment 2 except that examinees were given unlimited time to complete the test. The following research questions were addressed:

2.1 Given unlimited writing time, do French Immersion students achieve similar scores when they respond to French and English forms of a standardized test of social studies achievement?

2.2 How do the scores of French Immersion students who were given unlimited time to write compare to those achieved by the groups who write with time limits?

Hypotheses

Three hypotheses were tested in this experiment. The first two pertained to Question 2.1. The last hypothesis was related to Question 2.2. These hypotheses and the rationale for each are presented, separately, below.

Question 2.1.

It was assumed that the French Immersion students who participated in the study would read and comprehend test questions less well in their second as compared to their first language. This deficiency in their reading ability was predicted to be sufficiently great as to cause F group students to achieve lower scores on the test than they would have obtained had they written the test in English. It was reasoned that if F group students were given unlimited time to write the test, this time would compensate for their relatively slow reading rate by allowing them to attempt all test questions. The provision of extra time would be unlikely to do anything, however, to compensate for their incomplete reading comprehension of the test questions, given that limited second language knowledge and skill proficiency was the assumed source of this deficiency. In this respect, the addition of extra writing time would have little or no effect on F group test outcomes.

Research indicated that reading comprehension is related to processing efficiency; the slower and less automatic the processing, the poorer the quality of comprehension (Laberge & Samuels, 1974; Perfetti, 1985). This implied that any F group student who was unable to complete the test because of his or her slower speed could also have had the poorest level of comprehension. This suggested that anyone who ran out of time might not have been able to understand the test questions completely even if he or she had had time to read them. From this it was inferred that the gain in scores that would accrue from having additional time to write the test would be minimal. Given this prediction, the following hypothesis was put forth:

Hypothesis 2.1.1: The scores of examinees who write the English form of the test will be greater than those achieved by students who write the French version.

The scores of F group students in Experiment 1 were expected to be more depressed, relative to E group scores, on data-based than on discrete items. As explained in the rationale for Hypothesis 2.1.1, the provision of extra writing time would not have a significant effect on the quality of F group students' reading comprehension. This implied that comprehension differences predicted to occur for F group students on discrete and data-based questions would be unaffected by the provision of more time. Unlimited writing time would, therefore, have influenced the difference in effect sizes on discrete and data-based questions only if F group students were able to correctly answer data-based questions that they would have been unable to respond to if they had a shortage of time. Given this prediction, the following hypothesis was put forth:

Hypothesis 2.1.2: Group differences in scores will be greater on data-based than on discrete knowledge items.

Question 2.2.

It was hypothesized earlier that, because of their lessened ability to comprehend the test questions, F group students would achieve lower scores on the test than would E group students. It was also argued that providing unlimited time to write the test would not alleviate this difference in scores because time, as such, would have had little or no effect on

the ability of F group students to comprehend what they had read. This implied that the scores of F group students writing with no time limit would not be significantly different from those of F group students who wrote with time limits.

E group students writing under standardized timed conditions would be unlikely to experience difficulty completing the test because the time limit set for the test was that which was appropriate for native speakers of English, and E group students in this study were native speakers of English. This implied that their scores would be unaffected by the time limit. Thus, all other things being equal, their scores and those of E group students in Experiment 2 should have been the same. Given these assumptions about the effect that unlimited time will have on the scores of E and F group students in Experiment 2 the following hypothesis was put forth:

Hypothesis 2.2.1: There will be no significant difference in the main effects for experiments 1 and 2.

CHAPTER IV

Methods and Results

The two experiments that were run and the results that were achieved in each, are presented and discussed in this chapter.

Experiment 1

Subjects

Six urban elementary schools in central Alberta provided the setting for Experiment 1. Permission to carry out this research was obtained from each school's respective central office administration. The schools that were selected for the study drew children from families of similar middle to upper middle class socio-economic backgrounds. None of these schools could be considered to have had children in their French Immersion programs who were from disadvantaged families.

Most, if not all, of the children in the classes under study had followed the usual pattern of early total immersion in which Kindergarten and Grade 1 were totally taught in French, followed by the introduction of English language arts in grades 2 or 3. At the time of the study, at least 60% of all of their school subjects were being taught in French, including social studies.

Instruments

Two tests were administered to determine if social studies achievement as measured in French differed from social studies achievement as measured in English: *The Grade 6 Social Studies Achievement Test Part A: Multiple Choice* (Student Evaluation Branch, 1985a) and its French translation, *Test de Rendement Etudes sociales 6e année Partie A: Choix multiples*

(Student Evaluation Branch, 1985b). These tests are described below.

Grade 6 Social Studies Achievement Test Part A: Multiple Choice.

The *Grade 6 Social Studies Achievement Test Part A: Multiple Choice* was developed by this author under the auspices of Alberta Education. Its purpose was to provide educators, trustees, and others with information about levels of social studies achievement at local and provincial levels. The test measures student knowledge and skills in relation to social studies program objectives (Student Evaluation Branch, 1984). Its content emphasis is derived from the *Grade 6 Social Studies Curriculum Specifications* (Curriculum Branch, 1984). The test has 50 items with an administration time of 50 minutes. Examinees are required to use separate machine-scorable answer sheets.

Although the design mean of the test was 62.5%, the provincial average, when administered in 1985, was 29.9 out of 50 or 59.8%. The standard deviation was 8.5. Items ranged in difficulty (p-value) from .36 to .88 and had discrimination values no lower than .200 (Student Evaluation Branch, 1985c). For reporting purposes, the test items were grouped into the following categories (subtests):

1. Topic A: All questions related to how people in ancient times met their physical, psychological, and social needs.
2. Topic B: All questions related to how people in Eastern societies meet their physical, psychological, and social needs.
3. Topic C: All questions related to meeting physical, psychological, and social needs through local, provincial, and federal government.
4. Recall & Comp. A: Recalls and understands facts, concepts, and generalizations related to how people in ancient times met their needs.
5. Recall & Comp. B: Recalls and understands facts, concepts, and generalizations related to how people in Eastern societies meet their needs today.
6. Recall & Comp. C: Recalls and understands facts, concepts, and generalizations related to meeting needs through local, provincial, and federal government.
7. Values: Recalls and understands competing values and uses skills to analyse competing value positions.
8. Inquiry I: Uses skills related to identifying elements of an issue, formulating research questions and procedures, and gathering data.

9. Inquiry II: Uses skills related to analysing, evaluating, and synthesizing data.
10. Inquiry III: Uses skills related to resolving issues, planning courses of action, and evaluating decisions and courses of action.

The distribution of questions, by category, is provided, along with the tests, in the Appendix.

Prior to its administration, the test was reviewed for content validity, accuracy, and technical merit by Grade 6 social studies teachers from all parts of the province and by a test review committee. No evidence of the construct validity of the reporting categories has been provided.

Test de Rendement Etudes sociales 6e année Partie A: Choix multiples.

Following final review by the Test Review Committee, the *Grade 6 Social Studies Achievement Test Part A: Multiple Choice* was professionally translated by Alberta Education as a service to school jurisdictions offering Grade 6 social studies in French. Students who were taught social studies in French were exempted from writing the provincial achievement test in that subject in 1985 (Student Evaluation Branch, 1984). However, a number of schools offering it in French opted to have their students write the achievement test in French. The scores of the French Immersion students who wrote the *Test de Rendement Etudes sociales 6e année Partie A: Choix multiples* in 1985 are not available.

Data Collection Procedures

The *Grade 6 Social Studies Achievement Test Part A: Multiple Choice* and its French translation *Test de Rendement Etudes sociales 6e année Partie A: Choix multiples* were administered to French Immersion students by their social studies teachers during the first two weeks of June, 1986. Each version of the test was randomly distributed to half of the students in each of the eight classrooms tested. In all, 95 students wrote the English version and 84 students wrote the test in French. Teachers were instructed to follow the standardized administration procedures developed for these tests, including the 50 minute time limit. The only variation from the original administration procedures was that student instructions and sample questions were presented in both French and English. Consistent with the provincial

administration of these tests, the students were told that their marks would not count toward their final grades but that it was important that they do their very best. It was indicated that the purpose of the study was to determine if the scores of those who wrote the French version would be the same as those achieved by the students who wrote the English form.

Data Analysis

In order to test the mean differences in student scores on French and English versions of this social studies achievement test, t tests for independent samples were performed on the total test and reporting category mean scores of the E and F groups. The .05 level of significance was used in testing the hypotheses. Effect sizes were used as a means of comparing group differences in performance across reporting categories. These effect sizes were calculated using either the standard deviations from the 1985 provincial administration of the achievement test or the pooled standard deviations of the E and F groups. Any difference in effect size greater than .25 was treated as important.

Results

Two questions were examined in this experiment. The results pertaining to each question are addressed, separately, below.

Question 1.1.

Do French Immersion students achieve similar scores when they respond to French and English forms of a standardized test of social studies achievement?

The means and standard deviations of the E and F groups on the total test and the subtest reporting categories used in reporting the 1985 provincial achievement testing results are presented in Table 4-1. The results show that, as predicted in Hypothesis 1.1.1, examinees who wrote the English form of the test achieved significantly higher scores on the total test and on all subtest reporting categories than did those who wrote the French version.

What is apparent from the data is that these differences in scores are systematic (i.e., unidirectional), [illegible]ose who wrote in English, across all reporting categories. This

Table 4-1
Summary Results Using Alberta Education Reporting Categories
Experiment 1

Reporting Category	Mean			Standard Deviation			Significance[1]	Effect Size[2]
	E Group	F Group	Prov.[3]	E Group	F Group	Prov.		
Total Test	31.5	24.5	29.9	6.9	7.3	8.5	***	.83
Topic A	11.0	7.8	10.7	2.5	2.9	3.2	***	.98
Topic B	10.3	8.8	9.7	2.7	3.0	3.1	***	.48
Topic C	10.3	7.8	.5	3.0	3.3	3.5	***	.70
Recall & Comp. A	4.8	4.0	5.0	1.4	1.6	1.4	***	.51
Recall & Comp. B	4.3	3.7	4.1	1.3	1.7	1.6	*	.34
Recall & Comp. C	4.3	3.2	4.1	1.8	2.1	2.0	***	.55
Values	3.9	2.6	3.2	1.3	1.1	1.5	***	.84
Inquiry I	5.4	4.0	5.0	1.6	1.4	1.7	***	.81
Inquiry II	6.0	4.5	5.4	1.6	1.9	2.0	***	.76
Inquiry III	3.3	2.8	3.1	1.4	1.4	1.5	*	.29

[1] Probability levels of t tests of significance.
[2] Effect sizes are calculated in relation to the standard deviations obtained when the Grade 6 Social Studies Achievement Test was administered provincially to English program students in Alberta in 1985. Any difference greater than .25 is treated as important.
[3] Mean scores achieved when the Grade 6 Social Studies Achievement Test was administered provincially to English program students in Alberta in 1985.
*p<.05. **p<.01. ***p<.001.

systematicity in the results may indicate that one or more language of testing variables had a pervasive influence on F and/or E group behavior across the whole test.

While the differences in scores across the tests are systematic in their direction, they are not constant in their magnitude; relative to provincial standard deviations, the effect sizes range from .29 to .98 across reporting categories. Since variations in effect sizes of this magnitude are unlikely to have occurred by chance, it can be inferred from these results that there is a relationship between the category of question being asked and the size of the differences in E and F group scores. This implies that variables related to the way these items were grouped affected how students responded to those questions in French and English. In short, while there may have been general factors contributing to the differences in scores across all items, factors specific to individual reporting categories may also have had varying degrees of influence on examinee behavior.

That there would be a variation in effect sizes relative to item groupings was anticipated and expressed in Hypothesis 1.1.2. Specifically, it was hypothesized that effect sizes would be greater on items that were data-based than on those that were discrete. This hypothesis was based on the assumption that reading ease would differ across these groupings of items and that F group behavior would be more significantly affected by this difference in reading ease than would E group behavior.

The test questions used in this study were grouped, for reporting purposes, as they were for the 1985 provincial administration of the test. The results in 1985 were not specifically reported in relation to item type, that is, according to whether the items were discrete or data based. To be consistent, the data from this experiment were not analysed in this manner either. Nevertheless, it is possible to infer from the results what the magnitude of the group differences in scores is on discrete and data-based items because three reporting categories (i.e., Recall & Comp. A; Recall & Comp. B; and Recall & Comp. C) are composed of discrete items only and all of the questions in reporting categories Values, Inquiry I, Inquiry II, and Inquiry III are data-based. When the results in Table 4-1 are examined in relation to these item groupings, it can be seen that there is some support for

Hypothesis 1.1.2: Except for the reporting category Inquiry Skills III, the effect sizes on reporting categories composed of data-based items are larger than are those for categories composed of discrete items. This may indicate that E and F group performances on the questions were related to item type.

Another trend in the data suggests that item type may not have been the only variable related to the differences in E and F group performance. The results show that there are considerable variations in the size of the language of testing effects among the three topic-specific reporting categories: The effect size is greatest on those questions covering material from Topic A (ancient civilizations), followed by those from Topic C (governments in Canada), and finally those from Topic B (Southeast Asian societies). These results seem to suggest that the ease with which students were able to answer the questions in one language as compared to the other was related to the conceptual content of the question, as defined by its curriculum topic of study. In other words, the magnitude of the language of testing effect seems to have varied in relation to what the questions were about.

One other trend in the figures in Table 4-1 is worth noting. The data reveal that the pattern of effect sizes across the topic-specific comprehension reporting categories is not consistent with that which is present across the three reporting categories that reflect all questions (i.e., recall and comprehension, value, and skill items) in each topic. In particular, the effect sizes for Recall & Comp. A and Recall & Comp. C are more similar to each other than are those for Topic A and Topic C in their entirety.

This discrepancy is of interest because reporting categories Recall and Comp. A and Recall and Comp. C are subsets of Topics A and C, respectively. This means that for the effect sizes on topics A and C to have differed from each other as much as they do, then the language of testing effect must have been greater on the Topic A data-based questions than on either the Topic A knowledge-based or the Topic C data-based questions. In other words, effect sizes must have varied among the items within topics as well as among items within item types. This implies that either item topic and item type variables interacted to affect E and F group behavior differentially across reporting categories or that the apparent

relationships between examinee behavior and item topic and between examinee behavior and item type were in fact spurious and that some other variable or group of variables produced the results seen in Table 4-1.

So as to investigate these apparent relationships between effect size and item topic and item type further, the items were regrouped according to topic and type and new average scores were calculated for the E and F groups. These figures are presented in Table 4-2. Unlike those presented in Table 4-1, the effect sizes for these new groupings were calculated in relation to the pooled standard deviations of the E and F groups, and not the provincial standard deviations, as these figures were not available.

The composition of reporting categories Topic A, Topic B , and Topic C are identical in tables 4-1 and 4-2. This means that the same pattern of effect sizes are present for the reporting categories in both tables. Similarly, reporting categories Discrete A, Discrete B, and Discrete C in Table 4-2 are simply the reporting categories Recall & Comp A, Recall & Comp. B, and Recall & Comp. C, from Table 4-1, renamed. This renaming has been done to emphasize in what way the items in these categories differ from those in the data-based reporting categories. Because these reporting categories are identical to those in Table 4-1, there is no new information to be gained from these portions of Table 4-2. Instead, this data is presented as a way of providing a context for that information which is new in Table 4-2. What is unique in Table 4-2 is the grouping of all discrete items into one reporting category (Discrete), the grouping of all data-based items into another reporting category (Data), and the regrouping of items from the value and skill reporting categories in Table 4-1 into data-based categories that are topic specific (Data A, Data B, and Data C).

The data in Table 4-2 indicate that when items are pooled according to whether they are discrete or data-based, the resulting effect sizes are considerably different. These data seem to indicate support for Hypothesis 1.1.2 because it is apparent that group differences are greater on data-based than on discrete items. Closer scrutiny reveals, however, that the relationship between item type and effect size may not be as simple as that assumed by Hypothesis 1.1.2. When the items in reporting categories Discrete and Data are further

Table 4-2

Summary Results Using Reconstructed Reporting Categories

Experiment 1

Reporting Category[1]	Mean		Standard Deviation		Effect Size[2]
	E Group	F Group	E Group	F Group	
Topic A	11.0	7.8	2.5	2.9	1.19
Topic B	10.3	8.8	2.7	3.0	.53
Topic C	10.3	7.8	3.0	3.3	.80
Discrete	13.4	10.9	3.4	3.9	.69
Data	18.2	13.5	4.2	4.1	1.13
Discrete A	4.8	4.0	1.4	1.6	.54
Discrete B	4.3	3.7	1.3	1.7	.40
Discrete C	4.3	3.2	1.8	2.1	.57
Data A	6.2	3.8	1.7	1.8	1.37
Data B	6.0	5.1	1.9	2.0	.46
Data C	6.0	4.6	1.9	2.0	.72

[1]Discrete A; Discrete B; & Discrete C are the reporting categories Recall & Comp. A; Recall & Comp. B; and Recall & Comp. C from Table 4-1, renamed.

[2] Effect sizes are calculated in relation to the pooled standard deviations. Any difference greater than .25 is treated as important.

subdivided by topic, the resulting patterns of effect sizes are not consistent with the overall finding. In particular, there are no real differences in the effect sizes for Discrete B and Data B, and only a difference of .15 in the effect sizes for Discrete C and Data C. This indicates that the relationship specified by Hypothesis 1.1.2 only holds for items from Topic A.

It appears from the data in Table 4-2 that the situation may be somewhat similar in regard to the apparent relationship between item topics and effect sizes. There are important differences in effect sizes across the data-based reporting categories that are topic specific. However, no significant differences are present among the topic-specific reporting categories made up of discrete items. This suggests that variables related to item topics may be related to differences in E and F group scores on data-based but not on discrete items.

In summary, the scores of the two groups of French Immersion students participating in this study are consistently different from each other, with the E group achieving higher scores than the F group. While the differences in scores across all reporting categories are significant, the size of those differences varies. In particular, the effect sizes on topic specific data-based questions are all larger than the effect sizes on the same topic discrete item reporting categories. The most notable differences in E and F group scores seem to have occurred in relation to the data-based items from Topic A.

Question 1.2.

Are French Immersion students equally able to complete French and English forms of a standardized test of social studies achievement within the time limit established for the test?

The item analyses indicates that, while all of the E group students were able to complete the test in the time given, nine out of the 84 F group students who wrote were not. This finding suggests that it took examinees more time to read the French form of the test than the English version. However, no conclusion can be drawn about the cause of this apparent difference in reading rates. It may be that examinees had slower reading rates in their second language and therefore required more time to process the test questions in French as compared to English. On the other hand, it may be that the French form had more text to read than the English one. In other words, the French test may have had more and/or longer

text as a result of the translation process or simply as a function of natural differences in the two languages.

This unequal rate of completion across groups may explain why students had lower scores when they wrote the tests in French as compared to English. Students writing the French form may have felt a time pressure and may, therefore, have rushed to get through the test. This could have caused even those students who completed the test to score less well than they would have, had they written the test in English.

In order to determine whether or not the ability to finish the test in the time given was a crucial variable underlying the language of testing effect, it was decided to perform a second experiment. All of the conditions of Experiment 1 were replicated in this second experiment, except that examinees were given unlimited time to complete the test. The design of Experiment 2 and the results that were achieved are described below.

Experiment 2

Subjects

Six urban elementary schools in central Alberta provided the setting for Experiment 2. A process similar to that used to obtain permission to carry out Experiment 1 was undertaken. The schools that were selected for the study drew children who had the same or similar characteristics as the children who participated in Experiment 1. That is, the students came from families of middle to upper middle class socio-economic backgrounds. None of the schools had children in their French Immersion programs who were from disadvantaged families.

As with Experiment 1, most, if not all, of the children in the classes under study had followed the usual pattern of early total immersion in which Kindergarten and Grade 1 were totally taught in French, followed by the introduction of English language arts in grades 2 or 3. At the time of the study, at least 60% of all of their school subjects were being taught in French, including social studies.

Instruments

The same two tests that were administered in Experiment 1 were administered in Experiment 2. These were: *Grade 6 Social Studies Achievement Test Part A: Multiple Choice* and its French translation, *Test de Rendement Etudes sociales 6e année Partie A: Choix multiples*.

Data Collection Procedures

The tests were administered to French Immersion students by their social studies teachers during the first two weeks of June, 1987. Each version of the test was randomly distributed to half of the students in each classroom. In all 72 children wrote the English form and 75 wrote the French version. Teachers were instructed to follow the same standardized administration procedures as were used in Experiment 1. The only variation from the original procedures was that students were given unlimited time to complete the test. Consistent with the provincial administration of these tests, the students were told that their marks would not count toward their final grades but that it was important that they do their very best. It was indicated that the purpose of their writing the test was to determine if the students who wrote the French version would achieved the same scores as those who wrote in English.

Data Analysis

In order to test the mean differences in student scores on French and English versions of this social studies achievement test, t tests for independent samples were performed on the total test and reporting category mean scores of the E and F groups. The .05 level of significance was used in testing the hypotheses. Effect sizes were used as a means of comparing group differences in performance across reporting categories. Effect sizes were calculated using either the standard deviations from the 1985 provincial administration of the achievement test or the pooled standard deviations of the E and F groups. Any difference in effect size greater than .25 was treated as important.

Results

Two questions were examined in this experiment. The results pertaining to each question are examined, separately, below.

Question 2.1.

Given unlimited writing time, do French Immersion students achieve similar scores when they respond to French and English forms of a standardized test of social studies achievement?

The means and standard deviations of the E and F groups on the total test and the subtest reporting categories are presented in Table 4-3. These results show that in all cases the F group means are significantly lower than are those achieved by the E group. This indicates that, in spite of having unlimited time to complete the test, examinees who wrote the French form achieved significantly poorer performances on the total test and on all subtest reporting categories than did those who wrote the English version.

There is a broad range in the effect sizes across the test, indicating that the scores of the F group are more depressed, relative to those of the E group, on some categories of questions than on others. In particular, there seems to be a significantly greater depression of scores on Topic A items than on those from either topics B or C. This trend is present on both the total topic and the discrete item levels of reporting, suggesting that the effect size on Topic A data-based questions may also have been significantly different. These trends suggest that there is a relationship between the topic of the items and the magnitude of the discrepancy in E and F group scores.

The effect size on Inquiry III items is considerably smaller than are those on Inquiry I or Inquiry II items. These reporting categories consist of items involving increasing more complex data interpretation processes, with Inquiry I items requiring the least, and Inquiry III items the most, complex interpretions. These results may, therefore, indicate that effect sizes are related to the cognitive complexity of the items.

For the sake of clarification and ease of comparison with the results from Experiment 1, the items were regrouped according to their topic and type and new mean scores were calculated. Effect sizes for these new groupings were calculated using the pooled standard

Table 4-3
Summary Results Using Alberta Education Reporting Categories
Experiment 2

Reporting Category	Mean			Standard Deviation			Significance[1]	Effect Size[2]
	E Group	F Group	Prov.[3]	E Group	F Group	Prov.		
Total Test	32.1	24.8	29.9	7.4	7.8	8.5	***	.86
Topic A	10.9	7.6	10.7	2.9	3.3	3.2	***	1.03
Topic B	10.5	8.5	9.7	2.6	3.3	3.1	***	.65
Topic C	10.7	8.8	.5	3.1	2.7	3.5	***	.54
Recall & Comp. A	4.8	3.7	5.0	1.5	1.6	1.4	***	.79
Recall & Comp. B	4.2	3.6	4.1	1.5	1.7	1.6	*	.38
Recall & Comp. C	4.5	4.0	4.1	1.9	1.8	2.0		.25
Values	3.8	2.7	3.2	1.4	1.3	1.5	***	.73
Inquiry I	5.5	3.9	5.0	1.5	1.5	1.7	***	.94
Inquiry II	6.2	4.5	5.4	1.5	1.8	2.0	***	.85
Inquiry III	3.4	2.5	3.1	1.5	1.6	1.5	***	.60

[1] Probability levels of t tests of significance.
[2] Effect sizes are calculated in relation to the standard deviations obtained when the Grade 6 Social Studies Achievement Test was administered provincially to English program students in Alberta in 1985. Any difference greater than .25 is treated as important.
[3] Mean scores achieved when the Grade 6 Social Studies Achievement Test was administered provincially to English program students in Alberta in 1985.
*$p<.05$. **$p<.01$. ***$p<.001$.

deviations of the E and F groups, as was done in Experiment 1. These figures are presented in Table 4-4.

Two trends are noticeable in the data. First, when items are grouped by type, the effect size on the discrete items is considerably different from that on the items that are data-based. This trend is present across all three topics, thereby providing strong support for Hypothesis 2.1.2. Second, when items are grouped by topic, the effect size on those items pertaining to the study of ancient civiliations (Topic A) is significantly greater than are those for topics B or C. This trend holds regardless of whether the items are discrete or data-based.

In summary, there are significant differences in group scores across all reporting categories, in favor of those who wrote in English. The magnitude of these differences is greater on data-based items than on those that are discrete, regardless of which topics of study those items reflect. The effect sizes on Topic A discrete and data-based items are considerably greater than are those on the same type items from the other topics. Finally, when items are grouped by level of skill complexity, the effect size is smallest on those items judged by the test developers to be the most cognitively complex.

It is clear from these results that the French Immersion students participating in this study did not achieve the same scores when responding to French and English forms of the test, even though they had unlimited time to complete those forms.

Question 2.2.

How do the scores of French Immersion students who were given unlimited time to write compare to those achieved by the groups who wrote with time limits?

The total test mean scores of the E and F groups from experiments 1 and 2 (see tables 4-1 and 4-3, respectively) were compared using ANOVA. There is no significant difference in overall effect sizes across the two experiments. A glance back to the earlier tables reveals other similarities in the data across years. First, and most importantly, in both experiments F group scores are significantly and consistently lower than E group scores across all reporting categories. The amount that they differ varies across reporting categories with a similar range

Table 4-4

Summary Results Using Reconstructed Reporting Categories

Experiment 2

Reporting Category[1]	Mean		Standard Deviation		Effect Size[2]
	E Group	F Group	E Group	F Group	
Topic A	10.9	7.6	2.9	3.3	1.06
Topic B	10.5	8.5	2.6	3.3	.67
Topic C	10.7	8.8	3.1	2.7	.65
Discrete	13.4	11.4	3.7	4.2	.51
Data	18.7	13.5	4.3	4.4	1.20
Discrete A	4.8	3.7	1.5	1.6	.71
Discrete B	4.2	3.6	1.5	1.7	.37
Discrete C	4.5	4.0	1.9	1.8	.27
Data A	6.2	3.8	1.9	2.2	1.17
Data B	6.3	4.9	1.6	2.0	.77
Data C	6.2	4.7	1.7	1.6	.91

[1]Discrete A; Discrete B; & Discrete C are the reporting categories Recall & Comp. A; Recall & Comp. B; and Recall & Comp. C from Table 4-3, renamed.

[2] Effect sizes are calculated in relation to the pooled standard deviations. Any difference greater than .25 is treated as important.

in effect sizes across experimental conditions. Second, when skill-based items are grouped by their level of cognitive complexity, the effect size is smallest on those items that are most complex. Third, the effect sizes on topic-specific data-based questions are consistently greater than the effect sizes on the same topic discrete items. Fourth, the rank order of the effect sizes on the topic-specific data-based questions is the same across experiments. Finally, when items are grouped by topic and item type, the largest effect size is on data-based items from Topic A.

Given these similarities in the results across years, a number of tentative conclusions are possible. The first of these conclusions is that the performance of the E and F groups in Experiment 2 did not differ substantially from that of their counterparts in Experiment 1. The second conclusion is that, in both experiments, the magnitude of the effect sizes was related to the complexity of the cognitive processing requirements of the items. Finally, if one accepts the assumption of group comparability across experiments, then it can be concluded from these findings that the provision of extra writing time had no significant effect on the pattern of results in the second experiment. This implies that the differences in E and F group scores across experiments were not related to the ability to complete the test in the time given.

While there are many important similarities in the data from the two experiments there are also some notable differences. These differences need to be considered before it can reasonably be concluded that the provision of extra time had little or no effect on examinee behavior. The data show that the effect sizes for Discrete A and Discrete C reporting categories differ significantly in Experiment 2 but not in Experiment 1. At the same time, there are significant differences in the effect sizes for reporting categories Data B and Data C in Experiment 1 but not in Experiment 2. These trends are indicative of some of the remarkable changes that occurred in the effect sizes for specific reporting categories across experiments. Of particular note are the increases in effect size on reporting categories Inquiry III and Data B, and the decrease in effect size on Recall & Comp. C (Discrete C).

Increases in the effect size for any given reporting category, from Experiment 1 to Experiment 2, indicate that the discrepancy in E and F group scores was greater in Experiment 2 than it was in Experiment 1. This implies that F group scores decreased, or E group scores increased, or both. The opposite is true of a decrease in effect size: This implies an increase in F scores, a decrease in E scores, or both.

The results show that, in Experiment 2, the F group mean score on Inquiry III items is .3 lower than that achieved by the F group in Experiment 1; the E group mean is .1 higher than that of its counterpart. Similarly, the F group mean on Data B items is .2 lower than the F group mean on those items in Experiment 1, but the E group means in Experiment 2 increased by .3 over those in Experiment 1. In other words, in both of these situations, the Experiment 2 F group means decreased while the E group means increased. This implies that the increase in effect sizes across experiments, on these two reporting categories, can be attributed more to the relatively better performance of E group students than to the relatively poorer performance of F group students.

For two reasons, these results do not support the notion that time was responsible for any differences in E and F group performance. First, all E group students in Experiment 1 completed the test in the time given. Because experience has shown that French Immersion students have better than average first language reading abilities and that examinees who have this level of reading proficiency have no difficulty completing the test in the time given, it is assumed that this result indicates not only that they were able to complete the test but also that they were under no pressure to do so. This implies that the increase in E group performance in Experiment 2, relative to that in Experiment 1, was unrelated to the increase in writing time. Second, logic suggests that if time had had any effect on F group behavior in Experiment 2, that effect would have been to increase examinees' possibilities of selecting correct answers, not to decrease them. This means that their scores should have increased in Experiment 2. Given this premise, one can only account for the pattern of F group scores across experiments by assuming that time had no effect, or that that effect was masked by F group differences across years. This hypothesis seems unlikely, given that E group scores in

the second experiment were as good as or better than E group scores in Experiment 1. Thus, aside from the possibility of measurement error, the only reasonable conclusion is that, in Experiment 2, F group performance on these items was unaffected by the provision of extra time.

A somewhat different story emerges in relation to the effect size decrease on Recall & Comp. C (Discrete C) items. The data show that the mean scores of the F and E groups in Experiment 2 are higher than are those in Experiment 1 by .8 and .2, respectively. This pattern suggests that the decrease in effect size in Experiment 2 on this reporting category is, in large measure, attributable to the increase in the Experiment 2 F group mean on these items. The question is, was this increase related to the provision of unlimited writing time? It was noted earlier that eight out of the final nine questions on the test used in this study were recall and comprehension questions from Topic C. If F group students were pressured for time in Experiment 1, and the results say that at least some of them were, then the most obvious effect of this pressure would have been an inability to complete the final items on the test. Providing unlimited writing time in Experiment 2 would have meant that unlike those in Experiment 1, all F group students would have had the opportunity to complete all of these questions on the test, including those recall and comprehension questions from Topic C. This suggests that time may have been a factor contributing to the differences in F group scores on this reporting category across experiments.

In summary, there are a number of important similarities and consistencies in the data across the two experiments. These include the significant depression of F group scores, relative to those of the E group, across all reporting categories, as well as a similar pattern in the magnitude of group differences relative to the composition of reporting categories. These commonalities in the data suggest that the same factors were responsible for the depression of F group scores in both experiments. It appears that time pressures are not a *major* contributing factor in the depression of F group scores, because the provision of unlimited time does not seem to have made any significant difference in the scores.

There were, however, some notable differences in the data. While some of these differences could be attributed to group differences across experimental conditions, the lower F group mean in Experiment 1 compared to Experiment 2 on Discrete C items may have been the result of students' inability to complete all of the questions. This suggests that the provision of extra time may have had some impact on the responses of F group students in Experiment 2.

CHAPTER V

Conclusions and Implications

This chapter provides a summary of the conclusions that were reached in terms of the research questions that were posed. Theoretical and practical implications are also presented as are suggestions for further research.

Conclusions

The major finding of this study is that there are significant differences in the scores of Grade 6 French Immersion students who write a standardized test of social studies achievement in French and English, respectively. This difference is systematic across all reporting categories, in favor of those who write in English. The general conclusion reached from this finding is that the performance of French Immersion students, on a test of social studies achievement, is affected by the language of testing. In this respect, there is support for Carey's (1980) argument that language of testing differences could account for all or a portion of any observed discrepancy in the performances of French Immersion and English program students.

While the differences in E and F group scores are systematic across the test in terms of their direction, their magnitude varies across reporting categories. Because items were grouped into reporting categories according to apparent similarities in what they measured, and effect sizes varied across these categories, it is assumed that this indicates a relationship between what is being measured and students' abilities to respond to those items in French and English.

There are considerable differences in the effect sizes on discrete and data-based items. Data-based items differ from discrete items in that they require examinees to read and interpret novel information to answer them correctly. Discrete items typically assess an examinee's ability to recall previously learned information. This means that data-based items

measure more complex cognitive processes than do discrete items. Because data-based and discrete items differ in the complexity of what they measure, and effect sizes vary in relation to these items types, it can be concluded that the amount that F group scores are depressed, relative to those of the E group, is related to the cognitive processing demands of the item.

The results showed that while all students who wrote the English form of the test were able to finish in the time given, some of those who wrote the French form were not. From this it can be concluded that it takes French Immersion examinees more time to read the French form of the test than to read the English version.

When students were given unlimited time to write the test, their scores did not differ significantly or in any apparently important way from those of E and F group students who wrote under timed conditions. This implies that test-taking speed (i.e., the ability to complete the test in the time given) is not a significant factor underlying the language of testing effect. It must be noted, however, that some F group students in Experiment 1 were unable to complete the test in the time given. From this it can be concluded that, while the provision of extra writing time has no significant impact on the group score of those writing the French test, it can have a substantial effect for any given individual. At the same time, however, it must be acknowledged that anyone who is unable to finish the test because of slower reading speed, is also likely to have poor second language reading comprehension. As a consequence, he or she may not be able to increase his or her score significantly by being given more time to complete the test.

Implications

The contribution that this study can make to educational research in general is that it illustrates how extraneous variables can influence test outcomes thereby confounding data interpretations. It also demonstrates why it is so important to conceptualize and then control or account for the effects of such variables when conducting educational research. The E group scores in both experiments were equal to or higher than the averages obtained when the test was administered, provincially, to English program students in 1985. Based on these E

group scores, one might conclude that through instruction in a French Immersion program, it is possible to develop proficiency in French while at the same time achieving levels of performance that are equal to provincial levels.

One's conclusions would be quite different if one only had access to the F group scores and had no awareness of the apparent underestimation of social studies knowledge and skill that they provide. Given the differences in F group and provincial averages, one might assume that the students in this study had levels of performance that were considerably lower than were those of Grade 6 children in English language programs in 1985. An interpretation of these results might lead to the conclusion that, while French Immersion students are acquiring proficiency in French, they are doing so at the expense of their academic performance in social studies. Clearly this conclusion about the efficacy of French Immersion programs and the former one are widely discrepant, even though both are based on the performance of the same group of students.

In addition to showing how important it is to conceptualize and control confounding variables, the results of this study have a number of theoretical and practical implications that are more directly related to the issue of French Immersion testing. These implications are discussed below.

Theoretical Implications

While it is important to know that there is a language of testing effect on scores, it is as or more important to know what caused French Immersion students to achieve lower scores when they wrote in French compared to English. This understanding of the causal agents is important because it is only through knowing what the source of that depression is that one is able (a) to control or eliminate that element as a confounding factor or (b) to account for its impact on test outcomes.

Two factors that could be causally related to effect sizes were examined in Chapter II. These factors include the equivalence of the French and English forms of the test and the first and second language reading abilities of the examinees. Test equivalence was defined in

terms of item difficulty and text readability. First and second language reading abilities were considered in terms of levels of language-related knowledge and skill and the likelihood, given these levels, of achieving full comprehension of the textual material that was read.

Because of technical limitations, the equivalence of the two tests, in terms of their item difficulties and/or levels of readability could not be established. Similarly, the first and second language reading abilities of the students participating in this study were not determined. Consequently, it was not possible to determine what impact each of these variables had on E and F group performance. Nevertheless, given some of the trends in the data from the two studies, it is possible to develop some hunches about the effect of these variables on test outcomes.

The results showed that effect sizes were significantly larger on data-based skill items than they were on discrete knowledge-based items. This implies that F group scores were more depressed on the former type of question than the latter. Similarly, there seems to be some relationship between the topic of the items and the amount that F group scores were depressed relative to those of the E group. These results suggest that the discrepancy in E and F group scores can be attributed to test inequivalencies if, and only if, one can explain how these inquivalencies could be greater on some types and topics of items then on others. Conversely, to promote the argument that it is reading ability differences that account for the depression of scores, one has to explain how there could be greater reading ability differences on one type and topic of items then on others.

One could argue for the possibility of item difficulty and/or item readability differences across question types if one assumed that it was the amount of textual material and not the level of cognitive complexity that was the salient feature distinquishing performance on discrete knowledge and data-based skill items. Simply put, it could be that the more text associated with an item the harder it is for a translator to maintain the meaning and/or linguistic style of the original writer. Since meaning is related to item difficulty and linguistic style is a determiner of reading ease, then it could be assumed that the greater the difference in the quality of these variables across the French and English forms, the greater

the difference in E and F group scores.

A similar argument could be presented to explain the tentative relationship between effect size and item topic. That is, it could be that the extent to which a translator is able to remain faithful to the original item, in terms of its meaning and style, is related to the content that is being translated. Perhaps, for example, it is easier to retain meanings or levels of readability across languages (English and French) when dealing with abstractions about government (Topic C) than with concepts such as psychological needs or social equality (Topic A). As a result, the item difficulties and/or readability levels of the item from one topic may have been more similar to that of the originals than were those in another topic.

The limitation of both of these explanations is that they assume that the effect of the argued translation difficulties on item difficulty or readability would be systematic. In other words, they assume that all of the items of a certain type or from a certain topic would have been made either systematically more or systematically less difficult or readable than the English originals. The research cited in Chapter II did not support such an assumption.

The alternate explanation for the differences in E and F group scores put forth in this thesis was that French Immersion students have unequal levels of first and second language reading abilities and these differences systematically affect their comprehension of the questions and, therefore, their ability to answer those questions. This means that to argue for this explanation in a way that is consistent with the trends in the data, it must be shown that variables related to item type and item topic could be causally related to E and F group students' relative abilities to read and comprehend the test questions.

The rationale for an assumed relationship among the variables reading ability, item type, and effect size was presented in Chapter III. Briefly, it was argued that as text difficulty increases F group students would pay a proportionately greater price in processing efficiency then would E group students with the result that their level of text comprehension would be negatively correlated with the complexity of the question. E group students would not pay the same "price" as F group students when processing increasing complex text because relative to F group students they are skilled readers. According to Frederiksen (1981), the

price one pays is related to skill level.

To be consistent with the theory that it was reading ability differences rather than test differences that accounted for the discrepancies in E and F group scores, it must also be shown that the reading ability of either the E or the F group could vary according to the conceptual content of items. Such an explanation is possible and rests on the premise that reading comprehension is dependent among other things, on the reader's ability to achieve lexical access, and that this in turn depends on one's familiarity with the language. Since lexical access is related to the range and richness of one's understanding of the language (Just & Carpenter, 1987; Perfetti, 1985) an understanding that comes from a broad and varied exposure to that language, then it could be argued that French Immersion students in this study had different levels of language related knowledge about one topic of study than another. This could result from different exposures to the language of each topic because of differences in the text books and other resource materials, or because of differences in the teachers' familiarity with the terminology from one topic compared to the other.

In summary, arguments that are consistent with the data trends (i.e., a relationship between effect size and item topic and/or item type) can be made to support the notion that test inequivalencies accounted for group differences in performance. Arguments can also be made to show that it is reading ability differences that underlie the language of testing effect, and that the variation in effect sizes relative to item topics and types is a function of differences in E and F group levels of language related knowledge and skill in these areas. The weight of these arguments is not equal however. Instead the hypothesized relationship between reading ability and effect size variations seems more compelling for two reasons.

First, the reading ability arguments are based on assumptions that are supported in the literature. None of the assumptions underlying the test inequivalencies arguments are supported. For example, there is no evidence that linguistic style is easier to maintain with some topics than with others. The second point arguing in favor of the notion that it is reader differences that accounts for the apparent topic related variations in effect size is that some F group students were unable to finish the test in the time given. This finding is

meaningful because it suggests that the French Immersion students in this study had unequal first and second language reading speeds. Since reading speed has been correlated with processing efficiency and processing efficiency is fundamental to text comprehension, then this pattern implies that the F group students may have achieved incomplete comprehension of what they read.

Practical Implications

The results of this study indicate that it does matter in which language French Immersion students are tested; the scores of the students participating in this study were significantly lower when they wrote in French as compared to English. What is implied by these results is a need for educators to make a decision about the language in which French Immersion students should be tested.

Alberta Education exempts from its Achievement Testing Program those students for whom the test is inappropriate (Student Evaluation and Records Branch, in press). Included in the list of students who are eligble for exemption are those whose language of instruction is other than English and/or those students who are enrolled in an English as a Second Language program. Implicit in these categories of exemption is the principle that for participation to be appropriate, the language of instruction and students' language of fluency must match the language of testing.

On the basis of the evidence presented in Chapter II and the patterns of results obtained in experiments 1 and 2, it seems reasonable to conclude that a probable cause of the depression of F group scores in this study was reading ability differences. If this is the case, then it is not possible when testing French Immersion students to achieve the principles for appropriate participation as set out by Alberta Education (Student Evaluation and Records Branch, in press) because the language of instruction and the language of fluency are not the same thing.

This situation poses a dilemma in choosing a language of testing. It is clear from the number of jurisdictions choosing to have their French Immersion students write these optional

tests (Student Evaluation and Records Branch, in press), that educators want to know how well these students are achieving the goals and objectives of the programs of study. However, regardless of which language of testing is chosen there is a price to be paid (Carey, 1980). It appears from this study that choosing to test in the language of instruction when it is not the language of fluency will cause scores to be artifically low. Such low scores could have negative political and pedagogical implicatons if they are interpreted to mean that the level of French Immersion achievement is lower than expected. On the other hand, choosing to test in the language of fluency rather than the language of instruction suggests that the gap in fluency between the dominant and nondominant language is sufficiently great as to cast doubts on the efficacy of the entire French Immersion process. Moreover, the act of testing French Immersion students in English could lead some teachers to introduce English terminology into their instruction as a way of "preping" students to write that form of the test. This could further jeopardize the integrity of the program.

If one accepts that unequal levels of first and second language reading ability was causally related to the depression of F group scores, then one can infer that French Immersion students do not achieve complete comprehension of the school based texts that they read in their second language. This inference is not startling, given Carey's (1987) finding that high school students who have received all or most of their schooling in French Immersion programs achieve unequal levels of comprehension when they read excerpts from French and English versions of a textbook.

What does seem unusual and therefore warrants some consideration, is why the students in this study had levels of achievement that were equal to or better than provincial averages if they had achieved incomplete comprehension of the materials they had read in class. In other words, if their test performance was negatively affected by their second language reading comprehension, why wasn't their classroom performance (as measured by the English version of the test) equally depressed?

Two explanations are possible. One is that their performance was negatively affected. That is, their performance was at or above provincial levels, not because they achieved at

their optimal level, but because as a self-selected group (Carey, 1984; West, 1985) they were so superior to the regular program students that even though their social studies achievement was depressed by their inability to understand what they read in class (in French) they were still able to outperform the regular English program students.

A second possibility is that because these students were from Anglophone communities and French resources were not always available to them (Acheson, 1986), they were able to acquire some understanding of the concepts from the program through their experiences, in and out of school, in English. If this is true, then it provides a strong argument for ensuring that French Immersion students continue to have at least part of their resource materials provided in English so that their academic achievement is not jeopardized.

It is common to find researchers and educators who are willing to argue that parallel tests cannot be produced through translation and that it is this lack of parallelism that is responsible for effects such as that which were present in this study. The patterns of results in this study imply that contrary to this perspective, it is *not* the use of translations, per se, that is the problem. Rather it is the assumption that L2 students can perform equivalently in their nondominant and dominant languages that is problematic. In other words, it may be inappropriate to test French Immersion students using a translation, not because it diverges from the source (English) test but, ironically, because it remains too faithful: By maintaining the linguistic style of the original, the translator produces a version of the test that is appropriate for a native language reader. This is an important distinction and one worth pointing out to those who may have erroneously concluded from their findings (e.g., Scoon, 1974) that instrumen. differences and not unequal levels of language ability accounted for variations in group scores. In short, the use of transations may have been condemned in the past for the wrong reasons.

This inference has major consequences for test construction and administration. Given the increasingly multicultural nature of North American society, as well as the increased emphasis being placed on international assessments the demand for standardized tests in multiple languages is constantly growing. If equivalent tests in multiple languages can

be produced through translation then a significant economic burden is eliminated from such testing initiatives because test construction is a very costly and time consuming enterprise. Equally significant are the problems that would be eliminated in trying to equate parallel tests produced in separate languages.

Suggestions for Further Research

The ability to establish the parallelism of test forms in two languages is fundamentally related to the ability to answer an underlying question in this study which was "To what extent do test differences rather than reader differences account for the variation in scores across languages of testing?" The parallelism or equivalence of the two forms can only be established if one knows what text-related factors affect examinee behavior and then one is able to compare and contrast the tests in relation to these variables. This requires a technically and theoretically adequate enumeration and definition of the factors that contribute to or diminish test parallelism either within or across languages.

The variation in effect sizes in relation to item topic and type suggests that the amount that French Immersion students' scores will be depressed when they write a test in French rather than English will be inconstant across subject areas or even across test forms within a subject area. Similarly, the suspected relationship between first and second language reading abilities and effect sizes implies that the language of testing effect could vary according to the number of years students have spent in a French Immersion program. For these reasons it is suggested that this study be replicated with French Immersion students in different grade levels and in different subject areas.

V. REFERENCES

Acheson, J. B. (1986). *French immersion program survey.* Edmonton: Department of Program Services, Edmonton Catholic School District.

Adams, M. (1980). Failure to comprehend and levels of processing in reading. In R. J. Spiro, B. C. Bruce, & W. Brewer, (Eds.), *Theoretical issues in reading comprehension* (pp. 7-9). Hillsdale, NJ: Erlbaum.

Alderson, J. (1984). Reading in a foreign language: a reading problem or a language problem? In J. C. Alderson & A. H. Urquhart (Eds.), *Reading in a foreign language* pp. 1-24). New York: Longman.

Anderson, R. & Shifrin, Z. (1980). The meaning of words in context. In R. J. Spiro, B. C. Bruce, & W. Brewer, (Eds.), *Theoretical issues in reading comprehension* (pp. 331-348). Hillsdale, NJ: Erlbaum.

Atkins, B., Duval, A., & Milne, R. (1988). *Collins-Robert French-English English-French Dictionary* (2nd ed.). London: Collins.

Bain, B., and Yu, A. (1987). Issues in second-language education in Canada. In L. Stewin, & S. McCann (Eds.). *Contemporary educational issues - The Canadian mosaic* (pp. 215-225). Toronto: Copp Clark Pitman.

Barnett, M. (1986). Syntactic and lexical/semantic skill in foreign language reading: Importance and interaction. *The Modern Language Journal*, 70, 343-349.

Beheydt, L. (1987). The semantization of vocabulary in foreign language learning. *System*, 15, 55-67.

Berman, R. (1984). Syntactic components of the foreign language reading process. In J. C. Alderson & A. H. Urquhart (Eds.), *Reading in a foreign language* (pp. 139-159). New York: Longman.

Block, E. (1986). The comprehension strategies of second language readers. *TESOL Quarterly*, 20, 463-494.

Bloom, B., Madaus, G., & Hastings, J. (1981). *Evaluation to improve learning.* New York:

McGraw-Hill.

Bussis, A.M. and Chittenden, E. (1987). Research currents: What the reading tests neglect. *Language Arts*, 64, 302-308.

Capell, F. & de Porcel, A. (1979). Assessment of reading achievement in two languages: New methods for studying bias. In R. Silverstein, (Ed.), *Proceedings of the Third International Conference on Frontiers in Language Proficiency and Dominance Testing* (pp. 96-112). Carbondale: Southern Illinois University.

Carey, S. (1980). *Student evaluation in French programs in Alberta*. Edmonton, AB: Alberta Education.

Carey, S. (1984). Reflections on a decade of French immersion. *Canadian Modern Language Review*, 41, 246-259.

Carey, S. (1987). Reading comprehension in first and second languages of Immersion and Francophone students. *Canadian Journal for Exceptional Children*. 3, 103-108.

Chaudron, C. (1983). Foreigner talk in the classroom. An aid to learning? In H. Seliger & M. Long (Eds.), *Classroom oriented research in second language acquisition* (pp. 127-145). London: Newbury House.

Conrad, C. (1972). Cognitive economy in semantic memory. *Journal of Experimental Psychology*, 92, 149-154.

Cooper, M. (1984). Linguistic competence of practised and unpractised non-native readers of English. In J. C. Alderson & A. H. Urquhart (Eds.), *Reading in a foreign language* (pp. 122-138). New York: Longman.

Cowan, J.R. (1976). Reading, perceptual strategies and contrastive analysis. *Language Learning*. 26, 95-109.

Cronbach, L. (1971). Test validation. In R. Thorndike (Ed.), *Educational measurement* (2nd ed.) (pp. 443-507). Washington, D.C.: American Council on Education.

Cummins, J. (1983). Language proficiency, biliteracy and French immersion. *Canadian Journal of Education*, 8(2), 117-137.

Cummins, J. (1987). Immersion programs: Current issues and future directions. In L.

Stewin, & S. McCann (Eds.). *Contemporary educational issues - The Canadian mosaic* (pp. 192-206). Toronto: Copp Clark Pitman.

Curriculum Branch, (1984). *Grade 6 social studies curriculum specifications.* Edmonton: Alberta Education.

Curtis, M. & Glaser, R. (1983). Reading theory and the assessment of reading achievement. *Journal of Educational Measurement.* 20, 133-147.

Duncan, R. (1986). *Theoretically based test item readability: An approach to estimating the degree to which an item can be understood and answered correctly.* Unpublished doctoral dissertation, University of Texas, Austin.

Duff, A. (1981). *The third language: Recurrent problems of translation into English.* Toronto: Pergamon Press.

Dye, O. (1971). The effects of translation on readability. *Language and Speech,* 14, 392-397.

Favreau, M., Komoda, M., & Segalowitz, N. (1980). Second language reading: Implications of the word superiority effect in skilled bilinguals. *Canadian Journal of Psychology,* 34, 370-380.

Favreau, M. & Segalowitz, N. (1983). Automatic and [illegible] processes in the first and second language reading of fluent bilinguals. *Memory a[illegible]sition,* 11, 565-57[illegible].

Flesch, R. F. (1948). A new readability yardstick. *Journal of Applied Psychology,* 32, 221-233.

Fletcher, C. (1981). Short-term memory processes in text comprehension. *Journal of Verbal Learning and Verbal Behavior,* 20, 564-574.

Freebody, R. & Anderson, R. (1983). Effects on text comprehension of differing proportions and locations of difficult vocabulary. *Journal of Reading Behavior,* 15(3), 10-39.

Frederiksen, J. (1981). Sources of process interactions in reading. In A. M. Lesgold & C. A. Perfetti (Eds.), *Interactive processes in reading* (pp. 361-386). Hillsdale, NJ: Erlbaum.

Friesen, J. (1987). Family and school. An uneasy partnership. In L. Stewin, & S. McCann (Eds.). *Contemporary educational issues - The Canadian mosaic* (pp. 304-312). Toronto: Copp Clark Pitman.

Graham, J. (1985). *Differences in translation.* Ithaca, NY:Cornell University Press.

Horst, P. (1968). *Psychological measurement and prediction.* Belmont, CA: Wadsworth.

Ingram, D. (1985). Assessing proficiency: An overview on some aspects of testing. In K. Hyltenstam & A. Pienemann (Eds.), *Modelling and assessing second language acquisition* (pp. 215276). Clevedon, Avon: Multilingual Matters.

Jackson, J. & McClelland, J. (1981). Exploring the nature of a basic visual-processing component of reading ability. In O. Tzeng, & H. Singer (Eds.), *Perception of print: Reading research in experimental psychology* (pp 29-63). Hillsdale, NJ: Erlbaum.

Johnson, N. (1981). Integration processes in word recognition. O. Tzeng, & H. Singer (Eds.), *Perception of print: Reading research in experimental psychology* (pp 29-63). Hillsdale, NJ: Erlbaum.

Jones, J. (1984). Past, present, and future needs in immersion. *Canadian Modern Language Review*, 41, 260-267.

Jones, L. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.) (pp. 335-355). Washington, D.C.: American Council on Education.

Just, M. & Carpenter, P. (1987). *The psychology of reading and language comprehension.* Newton, MA: Allyn & Bacon.

Katz, J. (1972). *Semantic theory.* New York: Harper & Row.

Klare, G. (1963). *The measurement of readability.* Ames: Iowa State University Press.

Klare, G. (1974-75). Assessing readability. *Reading Research Quarterly*, 10, 62-102.

Klare, G. (1984). Readability. In P.D. Pearson, R. Barr, M. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 681-744). New York: Longman.

Koenke, K. (1987). Readability formulas: Use and misuse. *The Reading Teacher*, 40, 672-674.

Kolers, P. A. (1968). Bilingualism and information processing. *Scientific American*, 218,

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

UNIVERSITY OF ALBERTA

LANGUAGE OF TESTING EFFECTS FOR ACADEMIC ACHIEVEMENT OF FRENCH IMMERSION STUDENTS

by

MICHELE J. SAMUEL

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF EDUCATION

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

EDMONTON, ALBERTA

SPRING, 1990

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

ISBN 0-315-60362-3

Alberta
EDUCATION

Devonian Building, West Tower, 11160 Jasper Avenue, Edmonton, Alberta, Canada T5K 0L2

May 3, 1990

Graduate Studies and Research
2-8 University Hall
University of Alberta
Edmonton, Alberta
T6G 2E1

Dear Sir or Madam:

This letter is to confirm that permission was granted to Michele Samuel to use and include in her thesis copies of the following tests:

Grade 6 Social Studies Achievement Test, Part A: Multiple Choice (1985)

Test de Rendement Etudes sociales 6e année, Partie A: Choix multiples (1985)

If you require more information, please feel free to call me at 427-2948.

Sincerely

Darlene Montgomery

for Frank G. Horvath
Director
Student Evaluation and Records

FGH:2222-A

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR MICHELE J. SAMUEL

TITLE OF THESIS LANGUAGE OF TESTING EFFECTS FOR ACADEMIC ACHIEVEMENT OF FRENCH IMMERSION STUDENTS

DEGREE FOR WHICH THESIS WAS PRESENTED MASTER OF EDUCATION

YEAR THIS DEGREE GRANTED SPRING, 1990

(SIGNED) M. J. Samuel

PERMANENT ADDRESS:

#7, 8616 - 108 St.

Edmonton, Alberta

T6E 4N4

DATED June 23 1990

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled LANGUAGE OF TESTING EFFECTS FOR ACADEMIC ACHIEVEMENT OF FRENCH IMMERSION STUDENTS submitted by MICHELE J. SAMUEL in partial fulfilment of the requirements for the degree of MASTER OF EDUCATION.

Dr. S. Carey, Co-Supervisor

Dr. T.O. Maguire, Co-Supervisor

Dr. J. Edwards

Dr. R. F. Mulcahy

Date April 23/90

To the memory of my grandmother,

Desirée Blaquiere

ABSTRACT

The primary purpose of this study was to investigate the effect that language of testing has upon the scores of French Immersion students writing a standardized test of Grade 6 social studies achievement. It also examined the extent to which time is a variable that affects the test performance of French Immersion students.

Two experiments were run. In the first experiment French and English forms of a standardized test of social studies achievement were randomly assigned to Grade 6 French Immersion students. All of the standardized conditions of administration established for the test were followed, including the time limits. All of the conditions of Experiment 1 were replicated in Experiment 2 except that examinees were given unlimited time to complete the test.

In both experiments the scores of students writing the French form (F group) and those of students writing the English version (E group) were analysed according to the classification scheme used to report the results of the 1985 provincial achievement testing administration (Student Evaluation Branch, 1985c). The results from each experiment were interpreted separately and then comparatively in terms of the research questions that were posed.

The results revealed that French Immersion students achieve significantly lower scores when they write an achievement test in French as compared to English. While the differences in scores across all reporting categories are significant, the size of those differences is not constant. In particular, the effect sizes on topic specific data-based questions are all larger than the effect sizes on the same topic discrete item reporting categories. This indicates that the amount that scores are depressed is related to the type and topic of the items.

The scores of those students who wrote under timed conditions do not differ in any significant or important way from those obtained by students who had unlimited time to write the tests. This suggests that time pressures are not a *major* contributing factor in the

depression of F group scores.

The results from this study imply that it does matter in which language French Immersion students are tested and that this variable must be taken into account when interpreting test data.

ACKNOWLEDGEMENTS

I would like to sincerely thank all those people who have assisted and supported me in my effort to complete this thesis. In particular, I wish to acknowledge and express my appreciation to:

Dr. S. Carey, for suggesting a thesis topic that is still fascinating to me and for simultaneously challenging and supporting my thinking about the issue every step of the way.

Dr. T. O. Maguire, for dealing with me in a consistently respectful and supportive manner.

Drs. J. Edwards and R. F. Mulcahy, for so generously agreeing to serve on my committee.

Robert Runté, my friend and confidant, for his never-ending support and encouragement, including hands-on help when it was desperately needed.

Dr. G.W. Nixon, for reflecting back to me that which I needed to see to be able to finish this project.

Roy Jaffray and David Wasserman, for volunteering their time and sharing their expertise with me.

David Raboud and Jan Taylor, for caring about me and showing it throughout this project.

Guy and Helen Raboud, for teaching me to value truth and for helping me to develop the strength and courage that is required when seeking it.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

Introduction

Background to the Study

It is not idle curiosity that motivates educators and parents to find out how well students are achieving the objectives of school curricula. Instead, what underlies achievement testing programs is a desire to use such information for decision-making purposes (Bloom, Madaus, & Hastings, 1981; Sax, 1974). Because educational decisions have far-reaching pedagogical, social, and political implications, it is important to ensure that the conclusions one draws about levels of achievement are accurate. The accuracy of one's conclusions is founded on valid interpretations of test data.

The interpretation of achievement test data is a challenging task. According to Cronbach, (1971) tests are generally assumed to measure only the traits or constructs under study. In reality, extraneous variables may account, in part or in whole, for examinees' responses to test questions. These extraneous variables need to be controlled, or their effects accounted for, to have data interpretations that are valid. This requires the ability to distinguish relevant variables from the many elements sensed by the observer. This, in turn, depends on the ability to conceptualize certain elements as having a confounding effect on test outcomes (Jones, 1971).

Purpose of the Study

This study is about the validity of interpretations made about French Immersion program achievement test data. In particular, it examines Carey's (1980) hypothesis that language of testing has an influence on outcomes thereby invalidating or confounding conventional interpretations of test data.

In a report to the Department of Education, Carey (1980) recommended that the only valid way to measure French Immersion student achievement was to test that achievement in both languages. He based his recommendation on a belief that when French Immersion students, like any students, answer test questions, their responses are determined not only by their subject-matter knowledge and skill (the attributes under study) but also by the quality or nature of the test and by the students' ability to read the test questions. He argued that because the quality of the test and/or the students' ability to read might vary, depending on the language of testing, these extraneous variables could have an effect on how students respond to the test questions. In short, factors associated with the choice of language of testing could systematically shape how students respond to the test questions.

Carey's (1980) point is a significant one because, according to Capell and de Porcel (1979), differences in the scores generated by "parallel" achievement tests in two languages "signal the likely presence of some form of differential validity" (p.103). In other words, the conclusions one draws about levels of achievement and their implications for program change could differ depending on the language of testing.

In spite of its significance, it would be inappropriate to act on Carey's (1980) recommendation without further investigation because the evidence to support or refute his contention that language of testing affects outcomes is inconclusive. Secondary findings from a study by Swain and Lapkin (1981) do not support Carey's prediction that language of testing will affect outcomes. They reported that Grade 4 French Immersion students performed in French as they had in English to parallel forms of a test of social studies achievement. In other words, language of testing had no observable impact on outcomes.

Swain's and Lapkins's (1981) findings must be accepted with caution, however, for two reasons. First, there is no way of determining the internal validity of their study because their description of the instrumentation and method used is very limited. This means that uncontrolled variables could have accounted for their results. For example, nonequivalence of the tests could have produced their results if one test was more difficult than the other.

Similarly, subject selection differences could have confounded the outcomes. This possibility cannot be eliminated because there is scant information provided about the method of assigning subjects to conditions.

The second reason for questioning Swain's and Lapkin's (1981) conclusion is that their findings contradict the results of an American study which was much broader in scope. Willig (1985) conducted a meta-analysis of the results of studies of second language programs in the United States. She found that 63% of the total variance of effect sizes across studies could be accounted for in terms of six extraneous variables. One of these variables was the language of testing ($p.<.0001$). The significance of Willig's finding is not only that it contradicts Swain's and Lapkin's (1981) results but also that it is consistent with Carey's (1980) prediction.

The purpose of this study was to test Carey's (1980) hypothesis that language of testing is a variable that systematically affects how French Immersion students respond to test questions. This goal was approached by comparing Grade 6 French Immersion student performance on French and English forms of a standardized test of social studies achievement.

Importance of the Study

The need for an empirical verification of Carey's (1980) hypothesis can be seen in Cummins' (1983) comment that French Immersion programs have spread in Canada "not so much because they have succeeded in transmitting high levels of French proficiency to students at no cost to other academic skills, but because they have been *seen* to have succeeded" (p. 118). Given the apparent power that these studies have to shape educational programs in Canada, it is particularly important that valid inferences and conclusions be drawn from test data. A review of the literature indicates, however, that few if any Canadian studies of French Immersion student achievement identify language of testing as a variable that may confound data interpretation. The implications of this failure to control or account

for the effects of language of testing cannot be determined without a firm understanding of the impact of this variable on the nature of what is being measured.

Capell and de Porcel (1979) argue that proposed strategies for a second language program should be carefully scrutinized before they take on the institutional status of those routinely applied to monolingual programs. Their rationale for this assertion is that it is difficult to adjust any strategies, including inappropriate ones, once they have been instituted. It would seem, given Capell's and de Porcel's comments, that now is a particularly appropriate time to be addressing the question of whether or not language of testing is a variable that affects the scores of French Immersion students. Decisions about the nature of a French Immersion testing program are currently being made in Alberta (Student Evaluation & Records Branch, in press). An investigation of the effect that language of testing has on the scores of French Immersion students could provide needed direction, not only in the selection of appropriate instruments, but also in the development of data interpretation and reporting strategies.

Strictly speaking, the findings from this study will only be generalizable to situations where the same or similar testing instruments are being used. Nevertheless, the benefits to be gained by doing a study of the equivalence of French Immersion students' responses to French and English forms of a test are not limited to the Alberta situation. The results may serve to heighten the awareness of other researchers in Canada that language of testing has the potential to affect outcomes when French Immersion students are tested and thus should be considered when interpreting data or generalizing from one study to another.

Limitations Of The Study

This study does not address the question of whether or not standardized achievement tests that are blueprinted and fieldtested for use with regular English-language program students provide valid measures of achievement in French Immersion programs (whether those tests are presented in French or English). While this question is an important one, it goes

beyond the scope of the present investigation.

A number of variables related to language of testing that could affect outcomes are discussed in this study. The separate effect of these variables has not been isolated. As will be shown in the review of the literature, the overlapping nature of these variables, as well as technical limitations, preclude such a separation and analysis of effects.

The "degree of bilingualism" of the subjects in this study has not been measured. The lack of control of this variable may limit the generalizability of the findings.

Hypothesis To Be Tested

The hypothesis of interest to this study is that there will be a significant difference in French Immersion students' scores when they are tested in French as compared to English. The purpose of the next chapter is to determine a priori support for this hypothesis.

CHAPTER II

Review of the Literature

Introduction

Constructs such as social studies achievement are not directly measurable because they are not observable traits or behaviors. Instead, their presence must be inferred from test scores (Jones, 1971). These inferences about the attribute under study are based on an assumed relationship between the presence or absence of the construct being measured and the adequacy of responses to questions demanding the use of specific skills or items of information (Thorndike & Hagen, 1977).

If the assumed relationship between construct and response were a perfect one, then test scores would accurately reflect the attribute being measured and data interpretations would always be valid. Unfortunately, test scores often reflect information about traits or behaviors other than, or in addition to, the construct intended to be measured (Cronbach, 1971). As a consequence, assumptions about the relationship between the construct purportedly being measured and responses to test questions need to be challenged to ensure that data interpretations are valid. These assumptions are challenged through a process referred to as construct validation.

Construct validation begins with a claim that a given test measures a certain construct. The challenge consists of an attempt to prove a counterhypothesis. The counterhypothesis is an alternative explanation to account for test behavior in whole or in part. If the attempt to fit the data to the counterhypothesis fails, then the original hypothesis of what was being tested cannot be rejected (Cronbach, 1971). More importantly, it can be assumed that one's original inferences about the meaning of test scores are valid.

This study investigates the validity of the inferences one makes about levels of social studies achievement when French Immersion students are tested using French and English

versions of a standardized instrument. Its purpose is not to establish the absolute worth of the instruments as indicators of social studies achievement, but rather to determine the similarity of what is being measured by the two versions of the test. In other words, it is the construct equivalence of the two measures that is being studied.

The design of this study is similar to that which would be used to establish construct validity. First, an hypothesis is established about what is measured when French and English forms of a test are administered to French Immersion students. Then an attempt is made to prove a counterhypothesis. The hypothesis in this study is that the same body of knowledge and skills is measured when French Immersion students write French and English forms of a standardized test of social studies achievement. Since examinees' levels of social studies knowledge and skill remain constant, regardless of the language of testing, the operationalized form of this hypothesis is that their scores will be the same under both conditions of testing.

The alternative hypothesis to be examined is that the body of knowledge and skills assessed by the English form of the test differs, in whole or in part, from that which is assessed by its translated (French) version. In its operationalized form, the counterhypothesis is that French Immersion students' scores will differ under the two conditions of testing.

Carey (1980) alluded to two possible sources of difference in what is measured when French and English forms of a test are administered to French Immersion students. These sources of difference include (a) the effect of translation on the nature of test questions, and (b) the effects of first and second language reading abilities on test comprehension. Each of these conditions is examined below to determine if there is evidence to support the assumption that its effect is to change the nature of what is being measured by French and English forms of a test. Each section begins with an hypothesis about how the variable under study could cause what is being measured to differ under the two conditions of testing. Then the literature is examined to determine if there is support for this hypothesized source of difference.

Translation and Test Equivalence

Test Variables That Can Be Altered Through Translation

One of the variables that has a direct effect on the probability of selecting a correct answer to a test question is the quality or nature of that question. From this it follows that if, as a result of translation, the quality or nature of French forms of items is altered from that of the originals, then the probability of answering those questions correctly could vary. As a consequence of this variation, one's scores on a test could differ. For this reason, the quality or nature of English items and their French translations is of interest to this study.

Translation and Text Meaning.

One way of defining the quality or nature of an item is to consider the clarity of its meaning or purpose. Clarity of meaning is an essential attribute of an item because the writer's precision in selecting words is crucial in conveying the exact problem or task that the examinee must deal with. Often the choice of a particular word over a synonym can subtly change the meaning or emotional tone of an item (Bloom, Madaus, & Hastings, 1981). This in turn can affect how examinees respond to that item.

In considering the equivalence of the quality or nature of French and English forms of a test, a question that arises is whether it is ever possible to express the same meaning in two languages. This question is fundamental to this study because if the meaning of test questions is significantly altered through the process of translation then the probability of selecting correct answers to those questions could be affected. As a consequence, test scores could differ depending on the language of testing.

Language relativists hold that differences in the way various languages have come to encode meanings strongly influence the way in which members of that language group come to experience and know their world. Sapir (1961) wrote that the "real world" is to a great extent unconsciously built upon the language habits of individual groups. According to Sapir,

no two languages are ever sufficiently similar to be considered as representing the same social reality. As a result, "The worlds in which different societies live are distinct worlds, not merely the same world with different labels attached" (p. 69). From the relativist's point of view then, translation of meaning is not truly possible.

Others who have dealt extensively with the subject of translation hold a different view. That view is summed up in the following quotation from Katz (1972): "Natural languages are capable of providing a sentence to express any thought a speaker might wish to communicate. . . . For any example in English, a fluent speaker of any language could provide a parallel" (p. 12). Thus, from Katz's perspective the central conceptual meaning of utterances in one language can be translated into another language. This implies that it is at least theoretically possible to have French and English forms of a test that are parallel in meaning. It is necessary to apply the qualification of theoretical possibility to this assertion because the task of producing valid translations is a complex and difficult one.

One of the factors that makes translation a challenging task is that the meaning of a message is not "in the words". Thus the translator must discern not only what the *words* mean, but also what the *writer* means (Pergnier, 1978). As Graham (1985) notes, "the translator is under pressure not simply to produce a version of the original that reads or sounds well in the target language but also to understand and interpret the original masterfully so as to reproduce its message faithfully" (p. 37). What makes this interpretative phase of translation difficult is that words have no exact and constant equivalent in other languages (Pergnier, 1978). Because words have no constant equivalent, meanings often become distorted or blurred. Distortion occurs because the translator has to reconcile several possible meanings, including the author's intended meaning, the dictionary definition, and his or her own interpretation of a word or phrase (Duff, 1981).

To produce a faithful translation, the translator must do more than just interpret the original work accurately. She must also convey that meaning in a way that resembles the original writing. In short, the translator must maintain the style established by the author.

Maintaining that style is difficult because certain characteristics of one language are untranslatable into another. Often it is the syntactic requirements of the language such as specifying number or gender in nouns that cannot be translated (Scoon, 1974).

Translation and Item Equivalence.

Capturing and expressing the meaning of the source message of any piece of text is challenging enough for any translator, but much more is required when translating test items. Oller (1979) compares the task of translating test items to the translation of jokes, puns, riddles or poems. While it is the poetry in poems that must be captured when translating them, (Kolers, 1968) it is the meaning and the relationships among the stem and alternatives that must be preserved in test items. Preserving meaning and relationships is particularly important because these variables provide important cues to the test taker. Repetition of key words, grammatical inconsistencies, and unequal lengths among the alternatives all provide clues to, or pulls away from, the correct answer (Bloom, Madaus, & Hastings, 1981). Because these cues consciously or unconsciously shape how examinees respond to the test items, they can affect how difficult those items are. In essence, they shape the probability of selecting a correct answer.

Cues embedded in stems and alternatives of test items are not the only variables to affect the probability of selecting a correct answer. Test items are generally multidimensional in what they measure (Reckase, 1981). As a result they often tap skills and abilities other than those intended to be tested. In the case of multiple choice testing, it is difficult to obtain measures of factors independent of verbal comprehension because, no matter what skill or ability is being measured by an item, the achievement of a correct answer depends on proficient encoding and processing of the words and sentences of the test question (Horst, 1968; Nunnally, 1967). How proficient is one's encoding and processing of textual material depends, among other things, on the nature of the textual material being read, that is, on the text's readability (Duncan, 1986). From this it follows that one's comprehension of test

questions will depend, in part, on the readability of that material.

What this implies, in terms of this study, is that the responses of French Immersion students, to tests of social studies achievement, will be shaped not only by the meanings of, or cues embedded in, the items but also by the readability of the instruments used to measure that achievement. More importantly, it suggests that for students to respond equivalently to the two forms of the test, it is necessary for those tests to be equally readable. In short, the translator must not only have interpreted the meaning of the original test faithfully but must also have expressed that message in a way that is stylistically similar.

In summary, there are at least two ways that translation could alter the nature of test questions. First, the difficulty of items could be altered because of differences in the meaning of or presence of cues in the original and translated items. Secondly, the readability of the original items could be altered. Since readability affects how well students comprehend what is being asked by test questions, this alteration could affect the way students respond to those questions.

Understanding the impact of test translation on the nature of test questions is important to this study because if the probability of selecting a correct answer is altered through translation, then there is reason to question the null hypothesis that students' scores will not differ depending on the language of testing. Both of these issues are therefore examined in more detail below.

Translation and Item Cues

According to Oller, (1979) to achieve the required similarity in meaning and relationship when translating test questions one must maintain roughly the same style, the same usage of vocabulary and idiom, and comparable phrasing. Because of the problems inherent in translation, it is not always possible to achieve this similarity. The result is that for any item, translation will "produce (in principle and of necessity) a substantially different item" (p.93).

A study completed at the University of New Mexico investigated the feasibility of translating the *Boehm Test of Basic Concepts* from English into Navajo (Scoon, 1974). The scores of Navajo children who wrote the translated version were compared with those of an Albuquerque group of children who wrote the English version and with scores from the original Boehm norming group of English-speaking children. Scoon's results showed that the children who wrote the Navajo version had significantly lower scores than did either group who wrote the original (English) version. Based on these results she concluded that the Navajo form was not the same test as the original English version and from this that translation cannot be used to produce equivalent test forms.

Scoon's (1974) conclusion supports Oller's (1979) contention that translation produces a substantially different test. Her conclusion, however, may be a questionable one. Because a between-subject design was used, group differences and not item difficulty differences could have accounted for her findings. In other words, group variations in aptitude, experience, or even reading ability could have accounted for her findings. Consequently, it is difficult to accept her conclusion without further evidence.

Three English items and their French translations are presented below. These items demonstrate how subtle, yet potentially significant, changes can occur as a result of translation. They are presented and discussed here as a means of providing a context for interpreting Scoon's (1974) findings and as a way of judging whether or not it is possible to produce equivalent test items through translation.

The first set of items, presented in figures 2-1 and 2-2, are taken from a Grade 9 social studies achievement test (Student Evaluation Branch, 1987). They demonstrate how the meaning of an item can be changed in the process of being interpreted by the translator. The questions are based on four quotations and ask the examinee to identify which speaker fails to express a particular opinion. The English version asks the examinee to identify the speaker who fails to express an opinion about the *desirability* of using computers. The French version asks about the failure to express an opinion about the *advantages* of computer usage. These

Figure 2-1. Sample Item 1 (English)

SPEAKER I

Because of computer systems, it is now possible to monitor worker speed, accuracy, and length of rest periods. I favor the use of computers for two reasons: the number of managers needed to supervise work is reduced, and the problems with worker productivity can be identified more quickly.

SPEAKER II

With the continued automation of work, the skills and knowledge required to do the job are being transferred from the worker to the computer. Workers are reduced to watching machines. Work is becoming more monotonous, more routine, less challenging, and less rewarding. I think this is unhealthy.

SPEAKER III

Computer technology is changing the very nature of work. The result is that in some areas of the labor force, there is high unemployment as machines replace workers. In other areas there are skilled labor shortages. Significant adjustments to the labor force are needed to avoid a major crisis in the workplace.

SPEAKER IV

It is no longer necessary to assemble all workers at the same time and place. Portable computers create an office wherever the worker happens to be. The result is a lower expenditure of energy, time, and resources. You will never convince me that this is bad.

— Adapted from *Microtechnology*, 1982

15. Which speaker does NOT express an opinion about the desirability of using computers?

A. Speaker I
B. Speaker II
C. Speaker III
D. Speaker IV

Figure 2-2. Sample Item 1 (French)

INTERLOCUTEUR I
À cause des systèmes informatiques, il est maintenant possible de surveiller la vitesse et l'exactitude des travailleurs, et la longueur des temps de repos. Je suis en faveur de l'emploi des ordinateurs pour deux raisons: le nombre de chefs pour surveiller le travail est réduit et les problèmes que pose la productivité des travailleurs peuvent être identifiés plus vite.

INTERLOCUTEUR II
Avec l'automatisation permanente du travail, les aptitudes et les connaissances requises pour faire le travail sont transférées du travailleur à l'ordinateur. Les travailleurs en sont réduits à surveiller les machines. Le travail devient plus monotone, plus routinier, moins intéressant et moins valorisant. Je pense que c'est malsain.

INTERLOCUTEUR III
La technologie informatique est en train de changer la nature même du travail. Le résultat est que, dans certains domaines, il y a beaucoup de chômage parce que les machines remplacent les travailleurs. Dans d'autres domaines, il y a pénurie de main-d'oeuvre spécialisée. Des ajustements significatifs à la main-d'oeuvre sont nécessaires pour éviter une crise majeure dans le monde du travail.

INTERLOCUTEUR IV
Il n'est plus nécessaire de rassembler tous les travailleurs au même endroit et en même temps. Les ordinateurs portatifs créent un bureau là où le travailleur se trouve. Le résultat est une dépense moindre d'énergie, de temps et de ressources. On ne me convaincra jamais que c'est mal.

-- Adaptation de *Microtechnology*, 1982

15. Quel interlocuteur N'exprime PAS d'opinion sur les avantages d'employer des ordinateurs?

A. Interlocuteur I
B. Interlocuteur II
C. Interlocuteur III
D. Interlocuteur IV

items present different tasks to examinees because the terms "desirability" and "avantages" (advantages) have meanings that are quite different. Desirability refers to the attractiveness or advisability of an action or option (Websters, 1984) hence, in the English version, the examinee is asked to identify the speaker who fails to comment on this aspect of computer usage. The term "avantages" in the French translation of the item indicates that the student is to identify the speaker who fails to express an opinion about the *benefits* (Atkins, Duval, & Milne, 1987) accruing from this particular course of action.

The change introduced into the French version of the item is judged to be a function of interpretation and not of the ability of the target language to carry the meaning of the source language, because the French language includes the term "desirabilite" which, according to Atkins, Duval, and Milne (1987), translates to "desirability", the term used in the English version.

This change in word meaning has a direct effect on the correctness of the keyed response for this item. Of the four speakers, Speaker III is the only one who discusses an effect of computer usage without expressing an opinion about its desirability. Speakers I and IV both express positive opinions about the attractiveness or advisability of using computers. Speaker II offers the opinion that the use of computers is undesirable. The keyed response for the English version of the item is unquestionably alternative C. On the French version of the item, however, the keyed response is arguably either B or C, because neither Speaker II nor Speaker III discusses the benefits or advantages of using computers. In short, the keyed answer has changed due to differences in the wording of the item stems.

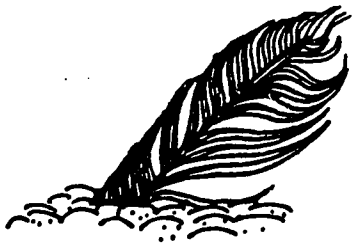
The next pair of items, presented in Figure 2-3 provide an example of how the nature of an item can be changed in the conveyance portion of the translation process. These items are taken from the test used as the criterion measure in this study (See Appendix). They form part of a provincial achievement test for Grade 6 social studies (Student Evaluation Branch, 1985a).

Figure 2-3. Sample Item 2 (English and French)

5. Which object found by archeologists would teach us the most about how people met basic needs?

A. A piece of volcanic rock

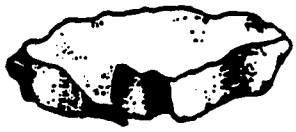
B. A rib from a buffalo



C. An eagle feather buried in sand

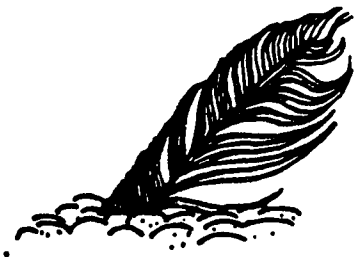
D. A needle made from a bone

5. Quel objet trouvé par les archéologues nous apprendrait le plus de choses sur la façon dont les habitants répondaient à leurs besoins essentiels?



A. Un morceau de roche volcanique

B. Une côte de bison



C. Une plume d'aigle enterrée dans le sable

D. Une aiguille en os

The portion of the items at issue is the keyed response, alternative D. The meaning of the correct alternative does not change from one version of the item to another. What does change is the wording of that alternative and this revised wording could affect how students respond to the item. The French version of the item has for alternative D, "Une aiguille en os" ("A bone needle") while the English version of the item says "A needle made from a bone". Since the critical feature that differentiates alternative D from the other three is that the object has been manipulated by man, the use of the word *made* in the English version makes this distinction more obvious then is the case with the French item. This item would predictably be more difficult in the French version.

The third pair of items, presented in figures 2-4 and 2-5, are taken from the achievement test used in this study. These items are based on six speakers' opinions about a law requiring the use of motorcycle helmets. What differs about the two items is the way in which the stem is worded. The term "comments" in the stem of the English question has been translated to "opinions" in the French item. Since the keyed answer is alternative A, "Opinions About the Helmet Law", the use of the word "opinions" rather than "commentairies" (comments) in the French stem is likely to make the answer to the French version more obvious. In other words, examinees are more likely to be cued to the correct answer if they are tested with the French rather than the English form of the item.

The differences illustrated by the three pairs of items discussed above support the assumption that the original meanings of, and relationships among, stems and alternatives can change when test items are translated. They also illustrate that item difficulty can be altered as a result of these changes. What can also be seen from these examples, however, is that the effect, on test difficulty, of these changes is not systematic. Two of the items are predicted to be *more difficult* in their translated form. The third item is predicted to be *easier* in its French form. Since support for Scoon's (1974) conclusion and the counterhypothesis in this study requires not only that items differ, but also that these differences yield results that are systematically different, this variation implies that one cannot assume that test scores will be

Figure 2-4. Sample Item 3 (English)

The law says that people riding motorcycles must wear helmets. Some people do not like this law and want to see it repealed (removed). Other people support the law.

These are some comments that citizens have made on this issue.

36. If all of the speakers' comments were being put on a chart, what would be the best title?

A. Opinions About the Helmet Law
B. Effects of Having the Helmet Law
C. Reasons for Keeping the Helmet Law
D. Persons Who Voted for the Helmet Law

Figure 2-5. Sample Item 3 (French)

La loi dit que les gens qui font de la motocyclette doivent porter un casque. Certaines personnes n'aiment pas cette loi et veulent la voir annulée (enlevée), d'autres soutiennent la loi.

Voici des commentaires faits par des citoyens sur cette question.

M. WYCLIFF

Certaines personnes ne savent pas ce qui est bon pour elles. Nous avons la responsabilité de les protéger.

M. BRANDON

Je suis fatigué que le gouvernement règle ma vie. Il y a des domaines dont il ne devrait pas se mêler. C'est un de ces domaines.

MADEMOISELLE SELDON

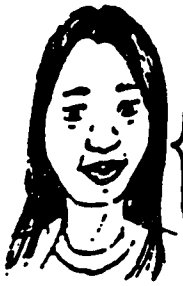
Je suis contente que le gouvernement ait fait quelque chose pour protéger les motocyclistes contre les accidents.

MARINE SAVTORI

Si je dois payer les frais médicaux par mes impôts, je devrais avoir le droit de dire aux motocyclistes de porter un casque. Je suis tout à fait en faveur de cette loi.

M. GIBEAU

Je peux décider ce qui est bon pour mes enfants.



MS. MAGUIRE

Je suis adulte. Je n'ai pas besoin que quelqu'un d'autre me dise quoi faire.

36. Si on rassemblait en un tableau toutes les opinions exprimées, quel serait le meilleur titre?

A. Opinions sur la loi sur le port du casque.
B. Effets de la loi sur le port du casque.
C. Raisons de maintenir la loi sur le port du casque.
D. Personnes qui ont voté pour la loi sur le port du casque.

constantly or systematically altered in the process of translation.

Translation and Readability

Researchers have determined that three broad classes of text-based variables influence reading comprehension. These variables include the physical characteristics of a piece of text, its content, and its linguistic style (Samuels & Eisenberg, 1981). Because these variables influence reading comprehension, then they must remain essentially unchanged for a source test and its translation to be equally readable. The likely effect that test translation will have on these variables is investigated below.

According to Samuels & Eisenberg, (1981) the physical variables that influence reading comprehension include such things as column widths, page and margin sizes, and the size and style of print. These variables can affect the speed of reading, the nature of eye movements and fixations, and the overall reading strategies used by a reader.

An examination of the two tests used in this study (See Appendix) reveals that in most ways the two forms are the same in terms of their physical characteristics. What does differ is the size and style of print. It seems unlikely, however, that this difference will have a significant effect on reader behavior because, according to Tinker (1966), most common typefaces are equally legible to an experienced reader. It is, therefore, possible to conclude that whatever physical differences exist between the two tests, these differences are unlikely to have a significant effect on examinee behavior.

A similar conclusion can be reached about the content variables that determine text, and therefore, test readability. Content variables include the specific subject matter of the text, the generality of the material and the abstractness of the material's presentation (Samuels & Eisenberg, 1981). While these variables are certain to have an effect on the difficulty of the tests under study, they are unlikely to be related to any differences in their relative ease or difficulty of comprehension. The reason that they are unlikely to be related to differences in difficulty is that the subject matter of the two tests will be the same, given a

highly qualified translator and a reasonable effort at translation.

The effect that translation may have on the third variable that affects reading comprehension, linguistic style, is less clear. Nida (1964) argued that when text is translated, no attempt should be made to preserve the original syntactic or semantic structure. Instead the message should be reduced to its kernel form and presented as a completely new utterance. Nida's assertion suggests that it is possible for the syntactic structure or choice of words used to express the original message to differ. Since word choice and syntactic structure are variables that have been correlated with ease of comprehension (McConnell, 1983), it could be that the translated version of a test is easier or more difficult to comprehend than the original form if the translator has introduced changes.

Dye (1971) conducted a study to determine the effects of translation on readability. He applied Flesch's (1948) readability formula to yield Reading Ease (RE) scores on the original (French) and translated (English) forms of sample passages taken from fourteen French originals and thirty translations of books and/or short stories. What Dye found was that scores consistently increased for text translated into English (higher scores indicate more easily read material). From this he concluded that the source documents became simpler to read when translated. He attributed the differences in RE scores to changes in the linguistic style of the passages as a result of translator changes.

Accepted at face value, Dye's (1971) conclusion lends support for the argument that translation alters levels of readability. Ironically, it is another finding from his research and the research he quoted that suggests Dye erred in his conclusion that the obtained differences in RE scores were an indication of changed levels of readability. Dye hypothesized that multiple translations of the same text would be reasonably consistent in their linguistic style because the style was predetermined by the original writer. He used the RE scores of the translations as indicators of linguistic style based on Klare's (1963) assertion that readability formulae measure difficulty of style. As predicted, Dye found that the RE scores of corresponding passages from four English translations of Voltaire's *Candide* were similar.

From this he concluded not only that the four translations were consistent in their linguistic style, but that this similarity was attributable to the fact that the translators had followed the style set by the original (French) writer.

What is noteworthy about Dye's (1971) discussion of his major findings is that he fails to see that his conclusions are illogical. On the one hand he argues that the style of the translations are similar because the translators were constrained by the style of the original writer. On the other hand he notes that the translations are not only equally readable but also consistently more readable than the original French version. The validity of Dye's second conclusion is dependent on the style of the translations being *consistently* different from that of the original. The reason that the translations have to be consistently different from the original in style is that linguistic style is the only variable that can cause readability to be the same across English forms but different from the French form, because in this case, the content of the message, the other critical variable determining readability, is constant.

The juxtaposition of Dye's (1971) two conclusions begs the following question. If the translators were sufficiently constrained by the style of the original writer to produce similar translations, why not sufficiently constrained as to produce translated text that parallels the source document? The only reasonable answer is that they were so constrained. In other words, in all likelihood the source and translated versions of the passages were stylistically similar and therefore, were, by definition, equally readable. What remains to be explained then, is why the RE scores of the source and translated versions were consistently different if their le els of readability were the same. The explanation appears obvious if one recalls that Flesch's Reading Ease Formula (1948) uses word length and average sentence length in words as its semantic and syntactic variables, respectively: What caused the RE scores to systemically and consistently vary was not stylistic differences but rather natural differences in the two languages. In other words, it simply took more and/or longer words to say in French what was said, in shorter form, in English.

An example of this phenomenon can be seen in the items presented in Figure 2-6. These items are taken from the tests used in this study. The translated item is faithful to the original in terms of sentence structure and word choice. In spite of this similarity in style, however, the stem and alternatives of the French form of the item are consistently longer than are those of the English version

Figure 2-6. Sample Item 4 (English and French)

12. In MOST early civilizations, wealth and power were

A. held mainly by the merchants and traders
B. held mainly by the nobles and priests
C. shared equally by the warriors
D. shared equally by all citizens

12. Dans LA PLUPART des civilisations du passé, la richesse et la puissance étaient

A. détenues principalement par les marchands et les commerçants
B. détenues principalement par les nobles et les prêtres
C. partagées également par les guerriers
D. partagées également par tous les citoyens

In conclusion, it appears that if translators have been faithful in maintaining the message of the source document, then the style of the original shapes the translation enough to maintain roughly the same linguistic style, at least as measured by counts of word and sentence length. Translators are unlikely to be any less constrained by the style of the original when translating tests than they are when translating prose. This implies that, except for minor variations in word choice or syntactic structure, the style of the English and French

versions of the tests should be similar. Research has indicated that text comprehension is influenced by a myriad of factors in addition to word choice and syntactic complexity (Koenke, 1987). As a consequence, minor text changes have little effect on readability. For example, Freebody and Anderson, (1983) have shown that a surprising number of difficult words have to be added to text before it becomes less readable. Similarly, Klare (1974-75) found that making sentences shorter did not necessarily lead to greater ease of comprehension. This suggests that the readability of a test is unlikely to be altered significantly as a result of minor word or syntactic changes introduced through translation.

One remaining issue requires resolution before it can be concluded that the readability of French and English forms of a test are likely to be approximately equal. This issue concerns the possible effect that the natural differences in the languages, referred to above, could have on comprehension. If the French form of a test consistently uses words of greater length than the English form, then word length as a variable (irrespective of its correlation to word meaning) must be ruled out as a determiner of comprehension to conclude that the tests are equally readable.

Evidence that word length does have an effect on reading behavior comes from the research related to eye-span behavior. When reading text, the eye moves in a series of discrete fixations with fast movements (saccades) in between (Just & Carpenter, 1987). Information is abstracted from text during these fixational pauses (Rayner, 1981). Just and Carpenter have shown that the time spent fixated on a word is directly related to its length; an average of 30 milliseconds more is spent on a word for each letter it contains. As well, Rayner has shown that the length of the word to the right of the word currently fixated influences the length of the following saccade.

The fact that every additional letter affects gaze duration and location implies that reader behavior may differ, at least in this way, when French Immersion students respond to tests in French as compared to English. It does not, however, prove that their level of comprehension will differ because of these differences in behavior.

To conclude that word length has a direct effect on the quality of text comprehension, evidence is needed that processing efficiency varies as a function of the number of letters in a word. This assertion is based on the widely held and tested assumption that one of the variables that determines one's degree of comprehension of textual material is the quality and efficiency of text processing in short-term memory (e.g., Jackson & McClelland, 1981; Perfetti, 1985).

Evidence suggests that it is unlikely that processing efficiency in short-term memory varies as a function of the number of letters in a word. Based on his research, Johnson (1981) concluded that featural characteristics of words (i.e., letter encodings) are the unit of representation only within the perceptual system of processing; the unit of representation within the cognitive system (i.e., working memory) is the encoded word. This point is significant because it suggests that the additional letters in the French text will only have an effect on reader behavior at preliminary stages of text processing. In short, while perceptual processing may be more complex for examinees reading a French translation than an English original, their higher order processing, including lexical access and semantic analysis, will occur only after the signals have been recoded into units which make the differences in letter counts irrelevant. Thus working memory capacity will be under no more strain when processing the longer French text than the shorter English version, assuming that any difference in length is a function of natural differences in the languages and not a result of differences in the linguistic style of the two tests.

In conclusion, there is little evidence to support the hypothesis that the French test will be systematically more or less readable than the original English version because of natural differences in the languages. A review of the two tests to be used in this study shows that in most cases, the French version of an item is longer than its English original. This extra length of the French text appears to be function of natural differences in the languages rather than due to stylistic differences, a factor than could have affected readability. Since letter encodings cease to be the unit of representation once text processing occurs within the

cognitive system, natural language differences should not affect text readability.

Conclusion

The purpose of this section was to determine if there is evidence to support the notion that scores could vary, depending on the language of testing, because of differences in the nature of the instruments being used. It was hypothesized that test scores could differ if test variables that have a direct effect on the probability of selecting a correct answer were altered as a result of the translation process. An assumption in this hypothesis was that these alterations would have to have a systematic effect on test difficulty for scores to be significantly different across forms. Two factors were examined: the effect that translation has on the meaning or cues provided by a test question and the effect of translation on readability.

A search of the literature indicated that neither of these factors has undergone much empirical investigation. Thus, any conclusions that can be drawn have to be based as much on logic as on hard evidence. Consequently, the conclusions are at best tentative. Nevertheless, based on what has been presented here it appears that there is insufficient evidence to support the notion that differences between carefully translated instruments will have a systematic effect on test outcomes.

Reading Ability Equivalence

The Relationship Between Reading Ability and Test Outcomes

Textual variables are not the only ones to affect the quality of comprehension processes. As Adams (1980) notes, the efficient operation of the system depends as much on the information in the reader's mind as on the information in the text. The reader plays a key role in text processing because the meaning of the discourse is something more than can be derived from a linguistic analysis of the text. According to Spiro (1980) what language

creates is a skeleton or a blueprint for the creation of meaning. It is the activity of the reader who by making an "effort after meaning" constructs a product that "makes sense within his or her individual view of the world" (p. 250).

The meaning that a reader constructs from text is shaped by two factors: the reader's knowledge (including world and language-related knowledge) and his or her skill at using that knowledge (Just & Carpenter, 1987; Perfetti, 1985). Readers differ in terms of these variables. Some readers have larger vocabularies and greater knowledge of lexical relationships than do others. Similarly, practised readers are distinquished by their greater ability to use the whole context to decode the meaning of unfamiliar words (Cooper, 1984), an ability that affects the comprehension process.

Because readers possess varying degrees of knowledge and skill, there are individual differences in what is comprehended from the same piece of text (Underwood, 1985). This implies, in the present context, that some test-takers will better understand the written form of a test question than will others. It also suggests that failure to perform adequately on an item could be more of a function of poor reading comprehension for some test-takers than for others. As a consequence, what a test question really measures (i.e., reading comprehension or the intended construct) will vary, depending on the knowledge and skill of the reader.

The processes involved in second language reading are similar to those required when reading in a native language (Block, 1986; Woytak, 1984). Readers construct meaning from text based on their level of knowledge and skill in that language. It follows from this that the more fluent is one's second language knowledge and skill, the greater or richer will be one's understanding of what is read in that language. Given this similarity between first and second language text-processing, it is possible to draw inferences about second language test-taking that parallel those presented above: Some second language test-takers will be better able to comprehend textual forms of test questions than will others and thus will be less likely to fail questions because of poor understanding. The implication is that the construct being

measured by test items will vary, depending on the level of knowledge and skill of the reader.

A question that arises in relation to this study is whether it is possible for an item to vary in terms of what it measures, depending on whether it is presented to an examinee in his or her first or second language. If French Immersion students are unequal in their ability to read in their two languages then it is possible that they will derive different meanings from French and English forms of a test, and therefore could respond differently to those questions. The remainder of this section is used to examine information related to first and second language knowledge and skill proficiency so that conclusions can be drawn about the first and second language reading abilities of French Immersion students.

Language-Related Knowledge

Much of the information that makes text understandable resides in the world knowledge shared by the writer and reader (Tighe & Hadaway, 1986). World knowledge includes such things as awareness of peoples' needs, wants, motivations, attitudes, plans, and values, and knowledge of specific content domains (Just & Carpenter, 1987).

In spite of its importance to the reading process, world or background knowledge is not a variable of interest to this study. It is not of interest because it is assumed that the background knowledge required to achieve comprehension will be constant across test forms, as the tests being used contain the same content, albeit in different languages. Moreover, it is assumed that the background knowledge of the examinees will be constant since a split-half design will be used.

What is of interest to this study is the body of knowledge required to read the test forms that is referred to as language-related knowledge. To understand written language, a reader has to encode the words and access their meanings in his internal lexicon (Just & Carpenter, 1987). The internal lexicon is a person's mental representation of word meanings (Underwood, 1985). Because lexical access plays such an important part in text processing its relationship to reading comprehension is one of the most robust and best documented

relationships in reading research (Just & Carpenter, 1987; Stahl, 1983). There is more to language-related knowledge than an awareness of what words mean, however, for as McKeown, Beck, Omanson & Perfetti (1983) note "A difference exists between acquiring knowledge of a word's meaning and knowing the word well enough to aid comprehension" (p.4).

What this difference entails is aptly described by Richards (1976). He lists a number of characteristics of "knowing" a word well enough to aid comprehension including: (a) knowing the probability of encountering that word in speech or print, (b) knowing the limitations imposed on the use of a word according to the variations of function and situation, (c) knowing the underlying form of a word and the derivations that can be made from it, (d) knowing the associations between a word and other words in the language (e.g., synonym, subordinate, and coorderinate relationships), (e) knowing the semantic value of a word, and (f) knowing the many different meanings associated with a word.

What Richards' (1976) list includes is not only the knowledge that is required to achieve lexical access but also the knowledge that is needed for syntactic analysis to occur. Syntax allows words to form higher order constituents such as phrases or clauses that provide part of the temporary structure required to organize words in memory until the underlying concepts are understood (Just & Carpenter, 1987). Syntactic analysis relies on cues in the text to indicate how words should be grouped into syntactic constituents. These cues include such things as word order, word class, function words, affixes, word meanings and punctuation. Awareness of what information is conveyed by a cue is referred to by Just and Carpenter as procedural knowledge, "a representation of the appropriate mental actions to be taken under a given set of circumstances" (p. 145).

The Transfer Of Language-Related Knowledge Across Languages.

In discussing the learning of a foreign language Beheydt (1987) pointed out that there is rarely a one-to-one match between the meanings of words in one language and the

meanings of words in another language. He concluded that in this respect, learning the vocabulary of a second language is really the acquiring of a new conceptual system along with the new verbal labels. What Beheydt's observations suggest is that the range and richness of understanding that French Immersion students have for English words does not automatically transfer when they learn the equivalent French verbal labels for words. Instead, to acquire that range and richness of meaning in their second language, French Immersion students must replicate their first language learning experiences.

Procedural knowledge, like lexical knowledge, is to a certain extent of limited transferability across languages. Because languages vary in the cues they use to signal the appropriate mental actions to be taken when reading, the expectancies set up by the reader when sampling syntactic clues in text must be related to one's knowledge of the structure of that language (Berman, 1984; Cowan, 1976; Just & Carpenter, 1987). To the extent that the languages are similar in structure, transfer of knowledge is facilitated and possible (Alderson, 1984). Confusion can occur, however, when predictions based on knowledge of the native language are used inappropriately when reading second language text. Yorio (1971) refers to this inappropriate transfer as language interference.

In the present situation where the languages of interest are fairly similar in structure there is likely to be a reasonable amount of knowledge transfer. Nevertheless, there are differences in the languages and thus to be equally fluent in their syntactic analysis of the two languages, French Immersion students must have frequent and varied exposure to both languages.

In conclusion, because there is limited transferability of lexical and procedural knowledge across languages, it is possible that French Immersion students will have unequal levels of language-related knowledge in French and English. If so, then they may have more trouble comprehending test questions presented in one language as compared to the other. It is therefore important to assess their relative levels of French and English language-related knowledge.

Assessing Language-Related Knowledge Equivalence.

The assessment of language-related knowledge is not easily achieved. Beheydt (1987) argues that the semantic values of words are only specified by their relatedness to and difference from words with adjacent meanings. What this implies is that to assess one's understanding of a word one must measure that understanding in relation to other words. In short, it is not adequate to assess language-related knowledge by having students select or provide adequate dictionary definitions of words because, according to Bussis and Chittenden (1987), simple tests of vocabulary recall cannot capture the full range and richness of meaning that a reader has for words in his or her lexicon. The process is even more complicated if it is the equivalence of language-related knowledge across languages that one is attempting to assess. The problems associated with this kind of assessment are obvious. Inequivalencies in the testing instruments or in examinees' abilities to articulate responses could confound the estimates of their actual level of knowledge in each language.

In discussing the evaluation of vocabulary understanding, Simpson (1987) noted that the issue in this type of assessment is not whether students know the words or not, but rather in what ways they know them. Simpson's comment hints at a method that can be used to infer the level of knowledge that French Immersion students have about their first and second languages: One can assess how French Immersion students have come to know what they know about their two languages. In short, one can assess their language acquisition histories. The rationale for this assumption is presented below.

The process of acquiring the procedural and lexical knowledge needed to achieve comprehension is a long and complicated one. Readers may initially know only some general features of a word, but over time they acquire a much more detailed representation of its meaning and usage. Developing a rich semantic and syntactic understanding of words requires much more than just time however. Progressive differentiation of word meaning and usage comes with frequent exposure to words in a variety of contexts (Anderson & Shifrin, 1980; Beheydt, 1987; Just & Carpenter, 1987; McKeown, Beck, Omanson, & Perfetti, 1983;

Simpson, 1987). This frequent and varied exposure is necessary because the "family of potential meanings" (Anderson & Shifrin, p. 332) that are associated with words are only articulated through contextual experience. This process is referred to as instantiation (Anderson and Shifrin, p. 334) or semantization (Beheydt, p. 55).

Because the meanings of words and the cues that they provide vary across languages, and those language specific meanings and cues are only acquired through frequent and varied contextual experiences, it follows that the quality and quantity of those experiences will have a direct relationship to the language-related knowledge that French Immersion students possess in each of their two languages. From this it can be inferred that the more equivalent is their first and second language acquisition histories, the more similar will be their levels of language-related knowledge. Consequently, one can predict their level of lexical and procedural knowledge in their two languages by assessing their experiences with those languages.

First and Second Language Acquisition Experiences.

West (1985) found that the parents of French Immersion students have an extraordinary degree of energy, enthusiasm, and commitment in regard to their children's education. They also have a higher socio-economic status, have greater confidence in their children's academic ability, and spend more time reading to their children than do parents of children in regular language programs (Carey, 1984). These characteristics are similar to those associated with high academic and reading achievement in English-language unilingual children. The parents of high academic reading achievement unilingual children are described by Friesen (1987) as having high aspirations and expectations for their children's achievement, considerable verbal interaction with them including time spent reading to them, and active involvement in their children's school programs.

Based on this similarity in parental characteristics it is possible to infer that French Immersion students will have a rich and varied experience with English, their mother tongue.

This does not appear to be the case when it comes to their experiences with French. When surveyed, parents and teachers from the French Immersion program of a large urban jurisdiction in Alberta expressed concern that their children had little opportunity to use their French language skills outside of the school (Acheson, 1986). Their concern seems legitimate for two reasons. First, the first and dominant language of 88 percent of the children in this French Immersion program was English. Only 7 percent of the children came from homes where French was the language currently spoken. Second, research shows that French Immersion students are more likely to read, watch television, and communicate with peers and adults in English rather than in French, when they are out of school (Cummins, 1987; McEwen, 1984; Swain & Lapkin, 1981).

That this discrepancy in the range and variety of experiences that French Immersion students have with their two languages will result in unequal development of those languages is evident from Bain's and Yu's (1987) comments about the Francophone experience with language in Western Canada. They note that even when Francophone parents speak French to their children "by preschool age the *lingua communis* . . . has become so dominant that it is but sentimental fiction to consider the language first spoken as the 'mother tongue'" (p. 221). If Francophones cannot maintain balance in their bilingualism, one wonders how French Immersion students from Anglophone backgrounds can be expected to.

In conclusion, it appears that in terms of their out-of-school experiences, French Immersion students are predominantly developing first language knowledge. The question then is whether their experience in a French Immersion classroom is sufficient to equalize their levels of L1 and L2 language-related knowledge. The answer, in the opinions of Carey (1987) and Swain (1974) is no; when second language learning is limited to school experiences, students rarely achieve a native-like command of that language. Part of the reason that classroom experience alone is insufficient to equalize French Immersion students' knowledge of French and English appears to be related to the quality of language children have in that setting. Different research studies have concluded that there is a considerably

higher proportion of teacher led lessons and much less small group work in Immersion programs than in regular English language programs (Cummins, 1987). As a result, students have little opportunity to use French in the classroom. Moreover, Chaudron (1983) has found that teachers faced with non-native speakers make greater efforts to simplify language than they would in a regular language classroom. This linguistic simplicity involves "less varied, more common and structurally more elemental or regularized material" (p. 128).

Children's experiences with written materials in French Immersion classrooms do not appear to be optimal either, according to the results of a survey conducted by Acheson (1986). He found that the limited availability of curricular materials in French was a paramount concern of teachers and principals. The existence of a shortage of French language resources was confirmed in an Alberta Education publication (Language Services Branch, 1985). In that document it was stated that "It will not surprise anyone that the French edition of approved English resources in not always available. Other appropriate French resources must consequently be identified to ensure that program objectives are met. At times, such resources cannot be found" (p. 13). Statements such as these strongly argue that French Immersion students are either using English resources or few resources at all. As a result, it is unlikely that students will have acquired as rich an understanding of the French language as they will have of English, their mother tongue and the language of the community.

Conclusion.

Vorhaus (1984) argues that the daily use of first language readers' own language across all communicative situations provides them with the advantage of being able to concentrate on comprehending related ideas and concepts represented by the words they are reading. In her view, however, readers in a second language feel constrained by their limited knowledge of vocabulary and grammatical rules and concentrate on segmenting meaning into understandable linguistic information. The effect, on comprehension, of these differences is

aptly described by Vorhaus:

> First language readers are interactors who use the author's language as a basis for developing concepts and an understanding of the author's idea while second language readers are mostly receivers who are constantly trying to develop more linguistic knowledge and insights about that particular author's language. . . . The first language reader has the linguistic resources that allow enough mental flexibility to understand what the author is *conveying*, while the second language reader can only use the available linguistic information to understand what the author is *saying*. (p. 413)

Given the previous description of French Immersion students' first and second language acquisition histories, it seems apparent that their communicative abilities in these languages will parallel those described by Vorhaus (1984). Simply put, French Immersion students will be less able to comprehend text presented in French than in English. Because comprehension is assessed each time students respond to multiple-choice questions it seems apparent that their unequal levels of comprehension will have an effect on how well they answer those questions. In short, there appears to be a compelling reason for arguing that their responses will vary, depending on the language of testing.

Skill Proficiency in First and Second Language Reading

In discussing the notion of language proficiency, Ingram (1985) argued that *knowledge* and *proficiency* are not the same thing. He noted that one can have considerable knowledge about a language including awareness of its grammatical rules and cues and yet not be proficient in the sense of being able to utilize that knowledge readily for practical communication purposes. A parallel distinction exists when it comes to reading. Even given an excellent command of the language, a reader will not achieve comprehension of text in the absence of proficient use of the skills that underlie the reading process (Just & Carpenter, 1987; Perfetti, 1985).

Reading skill proficiency is an issue of concern to this study because it has been consistently shown that foreign language readers perform more slowly in their second language than in their first language, for reasons not related to their knowledge of that language (Alderson, 1984; Favreau, Komoda, & Segalowitz, 1980; Favreau & Segalowitz, 1983; Woytak, 1984;). This slower second language reading rate suggests, at the very least, that when French

Immersion students write tests in their second language, they will require more time to process the test questions than if they had written the test in English. Thus some students who would be able to complete the English form of a test within the required time limits may be unable to complete the French form. This could artificially depress their scores on the test.

Their slower speed of reading may have a more deleterious effect on their test-taking ability than just an increase in testing time however. It may also affect how well they are able to comprehend the test questions. MacNamara (1967) found that the Irish-English bilingual students he tested were not only reading in their second language at a slower rate but also with lower comprehension. This lowered comprehension was not directly related to their level of language-related knowledge. Even when they understood the words and structures of the text under study (their understanding of the words and structures were tested separately) they were still less able to comprehend what they had read in their second language.

MacNamara (1967) assumed that his subjects' comprehension difficulties occurred because they required greater time and attention when decoding the semantic value of words in their second language. His hypothesis was that this increased time and attention added a burden to short-term working memory, thereby making it difficult for them to recall other parts of the message they were reading.

MacNamara's (1967) assumption that slower second language reading is a function of less efficient, more attention demanding, lower level text processing is supported by the findings of Favreau and Segalowitz (1983). They conducted a study with bilingual readers that was concerned with the use of automatic and controlled processing in a lexical decision task. Their results showed that bilinguals with *equal* first and second language reading rates responded in ways that suggested automatic processing in both languages. Bilinguals with *slower* second language reading rates showed a pattern of reaction times that suggested automatic processing in their first language but controlled or attention demanding processing in their second language. That this slower, attention demanding type of processing could cause second language readers such as MacNamara's to achieve poor comprehension fits with what is known about reading processes.

The relationship between automatic and controlled processing and skilled reading is a well documented one. Skilled reading involves the interactive processing of information from a number of information sources (Frederiksen, 1981; Perfetti, 1985). A fundamental component of this interactive processing system is the short-term or working memory (Masson & Miller, 1983). The working memory is where all of the information from the various sources is combined as evidence for or against hypotheses about meaning (Levy, 1981). Working memory is a limited capacity processor in terms of the amount of information it can process at any one time (Fletcher, 1981; Spiro, 1980). It is also limited in terms of the duration of time that traces can be held without active rehearsal (Lesgold & Perfetti, 1981).

Because of the limited capacity of short-term working memory, efficient comprehension can only be achieved by reducing competition for attentional resources among the component processes of reading. Competition is reduced by automatizing as many of the component processes as possible (Laberge & Samuels, 1974; Perfetti, 1985). The measure of automaticity is the extent to which an activity can be performed at the same time as a second activity to which attention must be directed (Underwood, 1985).

Not all of the component processes of reading are subject to automaticity of execution. For example, attention is demanded continuously if one is to integrate the meanings of individual words into a structure that corresponds to the underlying meaning of the text being read (Underwood, 1985). Similarly, a reader must consciously retain at the end of a segment of text what he or she read at the beginning for adequate comprehension to occur (Conrad, 1972; Curtis & Glaser, 1983; Masson & Miller, 1983). What is subject to automatic execution are the lower level activities of reading such as letter and word encoding and lexical access (Just & Carpenter, 1987). However, not all individuals are equal in their ability to perform these lower level activities automatically (Fredericksen, 1981; Levy, 1981; Samuels, 1987). In such cases, where increased attention must be allocated to specific lexical operations such as decoding, the higher-order processing of extended textual segments is jeopardized, resulting in poorer comprehension (Frederiksen, 1981; Laberge & Samuels, 1974;

Perfetti, 1985).

While it is clear from this description of the reading process that MacNamara's (1967) hypothesis is a valid one, one point requires clarification before it can be concluded that French Immersion students will be similarly disadvantaged when reading in their second language. That point relates to the fact that Favreau's and Segalowitz's (1983) subjects did not show uneven comprehension in their two languages even when their reading rates in those two languages differed. In other words, unlike MacNamara's subjects, their relatively inefficient lower order text processing did not appear to affect their understanding of what they read.

Favreau's and Segalowitz's (1983) subjects differed from MacNamara's (1967) in a very important way. Their subjects were fluent adult bilinguals who "read in each language at rates well within the range of normal monolingual readers" (p. 573). MacNamara's subjects were school aged children who were found to be "weaker than monolinguals in the monolinguals' language which was . . . the language of instruction" (p. 122). These descriptions suggest two reasons why the two groups of subjects differed in their abilities to comprehend what they had read. First, while Favreau's and Segalowitz's unequal reading rate subjects showed less automaticity in their second as compared to their first language, their overall efficiency may have been sufficient to permit adequate attention to what they were reading in their second language. In other words, their lower order abilities may have been sufficiently automatized as to free working memory capacity for the execution of the required higher-order processes.

The second point is that Favreau's and Segalowitz's (1983) subjects, unlike those of MacNamara (1967), were fluent in their second language, a fact which indicates that their lexical and procedural knowledge was well developed. This point is significant because it suggests that their higher level text processing abilities may have been sufficient to compensate for deficiencies at lower levels. According to Stanovich (1980) the processes of reading are not only interactive but also compensatory. This compensatory aspect of reading leaves open the possibility that higher level processes can actually compensate for deficiencies in lower

level processing. If, as Levy (1981) suggests, there is a threshold strength for determining when comprehension is achieved, then this threshold could have been achieved for Favreau's and Segalowitz's subjects through strong top-down support. MacNamara's subjects, on the other hand, may have had insufficient language-related knowledge to permit top-down compensation for weak lower level processing, with the result that their comprehension was incomplete in their second language.

Given their language acquisition histories, French Immersion students are more likely to have achieved the level of language-related knowledge acquired by MacNamara's (1967) subjects than that achieved by Favreau's and Segalowitz's (1983) subjects. Thus, it seems reasonable to conclude that, like MacNamara's subjects, French Immersion students will pay the price, in terms of comprehension, for their slower second language reading rate. That is, their comprehension of French text will be poorer than their understanding of English text. As a consequence, their ability to respond to French forms of a test will in all likelihood be more constrained by the need to read the test questions than it would be if they were taking the test in English. The predicted result is that their scores will be depressed, relative to their scores on an English version, when they respond to a test in French.

Summary and Conclusions

The purpose of this chapter was to examine support for the hypothesis that there will be no difference in French Immersion students' scores when they are tested in French as compared to English.

Two factors that affect how well students respond to test questions were examined: (a) the nature of the test questions and (b) the reading ability of the examinees. These factors were assumed to be relevant because if either of them were to vary with a change in the language of testing, then test scores could be affected.

In the case of the test questions, it was hypothesized that variables in the English items that affect item difficulty could be altered in the process of translating the questions into French. As a consequence of this variation, the probability of selecting a correct answer

could differ, depending on the language of testing.

From the literature it is apparent that the task of translating test items is a complex and difficult one; the translator must not only interpret the source message faithfully, but must also capture and then convey the linguistic and extralinguistic contexts that the text of an item calls to mind (Oller, 1979). These linguistic and extralinguistic contexts are fundamental to the nature of an item because they contain cues that affect examinee behavior. Similarly, the translator must also preserve the linguistic style established by the original test developer because examinee behavior is also determined by the readability of the items.

There has been very little research reported in the literature concerning how well translators are able to maintain the cues in, and readability levels of, source items when they translate them. Because of this absence of reported results, the possibility that the difficulty of test items will be altered through translation cannot be ruled out. Nevertheless, what little evidence is available suggests that the difficulty of test items will be affected by the translation process in a random rather than a systematic way.

In terms of the second factor, reading ability, it was hypothesized that French Immersion students could be unequal in their ability to read text presented in French and English. This unequal level of reading ability in their two languages could affect how examinees respond to test questions because reading comprehension is one of the factors being measured by paper and pencil tests of social studies achievement.

It was assumed that to achieve equal levels of comprehension of test questions presented in French and English, French Immersion students would have to have equivalent levels of language-related knowledge and skill proficiency in those two languages. Studies have shown that to have equal levels of language-related knowledge and skill proficiency in their two languages, readers must have had comparable experiences with them. Evidence suggests that Grade 6 French Immersion students in Alberta are unlikely to have had equivalent contextual experiences with French and English. This implies that their ability to comprehend test questions presented in French and English will differ. From this it is

possible to infer that their test scores could differ, depending on the language of testing.

CHAPTER III

Questions and Hypotheses

The major purpose of this study was to determine if French Immersion students performed in the same way when responding to French and English forms of a standardized test of social studies achievement. Two experiments were carried out. The research questions and hypotheses for each experiment are presented below.

Experiment 1

Questions

In Experiment 1 two research questions were studied. These were:

1.1 Do French Immersion students achieve similar scores when they respond to French and English forms of a standardized test of social studies achievement?

1.2 Are French Immersion students equally able to complete French and English forms of a standardized test of social studies achievement within the time limit established for the test?

These two questions were addressed by randomly assigning French and English forms of a standardized test of social studies achievement to Grade 6 French Immersion students. All of the standardized conditions of administration established for this test by its developers were followed, including the time limit. The responses of the group who wrote in English (E group) and the group who wrote in French (F group) were compared to determine if their performances were similar in terms of both their scores and their rates of completion.

Hypotheses

Three hypotheses were tested in Experiment 1. The first two pertained to Question 1.1. The last hypothesis was related to Question 1.2. These hypotheses and the rationale for each are presented, separately, below.

Question 1.1.

Research indicated that an examinee's performance on paper and pencil tests is shaped by his or her level of reading comprehension (Horst, 1968; Nunnally, 1967). This implied that the performances of French Immersion students, on tests presented in French and English, would be shaped by their abilities to comprehend text presented in those two languages. The research cited in Chapter II also indicated that the French Immersion students who would be tested in this study were likely to have less language-related knowledge and skill proficiency in their second (French) as compared to their first language (English). From this it was possible to predict that they would have unequal levels of comprehension when they read French and English text and thus that they would perform differently when responding to French and English forms of a test. Based on this prediction, the following hypothesis was put forth:

Hypothesis 1.1.1: The scores of examinees who write the English form of the test will be greater than those achieved by students who write the French version.

Research suggested that, as text difficulty increases, less skilled readers pay a proportionately higher price in processing efficiency than do skilled readers (Frederiksen, 1981). Since it it has been shown that reading comprehension is directly related to the efficiency of text processing in short-term memory (Laberge & Samuels, 1974; Perfetti, 1985), it was possible to infer that the greater the difficulty of the text and the less skilled the reader, the poorer would be his or her comprehension of that text. From this it was inferred that the greater the complexity of the text to be read on a test, and the more limited the language-related knowledge and skill of the reader, the poorer would be his or her comprehension of the test questions and the more likely he or she would be to answer the questions incorrectly.

It was concluded in the previous chapter that the French Immersion students in this study would have less language-related knowledge and skill in their second language (French) as compared to their first (English). As a result, they would be more like unskilled than skilled readers when reading in French. Given the aforementioned relationship between reader ability and text difficulty, it was possible to infer that as text difficulty increased, the students in the study would pay a proportionately higher price when reading text in their second language as compared to their first.

The questions on the social studies test that would be used in this study were of two different types: those that assessed recall and comprehension of previously learned information (knowledge-based items) and those that assessed the ability to process text-based information (skill-based items). The skill-based items differed from the knowledge-based items in two specific ways. First, unlike the knowledge-based items which "stood alone", the skill-based items were accompanied by graphic and/or textual data that needed to be read and interpreted for the questions to be answered. Second, the skill- but not the knowledge-based items contained information or content that was novel to the examinee. In short, skill-based items were more complex. They were, therefore, more difficult to comprehend than were knowledge-based items.

What this suggested, given the assumptions made above, was that (a) the processing efficiency of all examinees would be more taxed when reading skill-based as compared to knowledge-based items, and (b) that less skilled readers would have relatively more difficulty comprehending data-based questions than would those who were more skilled. In terms of the French Immersion students being tested in this study, this suggested that all examinees would have more difficulty reading data-based than discrete items, but that F group examinees would have relatively more difficulty than would E group students. Since performance on test questions would be related to one's ability to read and comprehend the questions, it was possible to infer that the performance of examinees in this study would be affected by the differing levels of readability of data-based and discrete items. Based on this assumption, the following hypothesis was put forth:

Hypothesis 1.1.2: Group differences in scores will be greater on data-based items than on discrete items.

Question 1.2.

The Grade 6 social studies achievement test used in this study was a power test which was designed and developed to assess the achievement of English language program students in Alberta (Student Evaluation Branch, 1984). Since the majority of students in this population were native speakers of English, it could be argued that the time limit established for the test was that which was appropriate for first language speakers of English.

Research indicated that foreign language readers typically require more time to read text presented in their second language as compared to their first (Alderson, 1984; Favreau, Komoda, & Segalowitz, 1980; Woytak, 1984). This suggested that the French Immersion students participating in this study would read French text more slowly than they would read text presented in English. It also suggested that the speed with which they would be able to read tests presented in French would be less than that in English. This suggested that they could be less able to complete French as compared to English forms of the test. Based on this assumption, the following hypothesis was put forth:

Hypothesis 1.2.1: The completion rate will be greater for examinees in the E group than in the F group.

Experiment 2

Questions

One must read and reflect on test questions to answer them correctly. This process requires time. If F group students in Experiment 1 ran out of time because of their slower reading speed, then they would have been unable to attempt questions that they may have been able to respond to correctly had they done the test in English. As a result, their scores could have been depressed relative to what they would have been had they written the test in English or had they had time to read and respond to all of the questions. This implied that all or a part of the language of testing effect predicted in Experiment 1 could have been the

result of F group students' pressure or inability to complete the test within the established time limit.

Experiment 2 was undertaken to examine this hypothesis. All of the conditions of Experiment 1 were replicated in Experiment 2 except that examinees were given unlimited time to complete the test. The following research questions were addressed:

2.1 Given unlimited writing time, do French Immersion students achieve similar scores when they respond to French and English forms of a standardized test of social studies achievement?

2.2 How do the scores of French Immersion students who were given unlimited time to write compare to those achieved by the groups who write with time limits?

Hypotheses

Three hypotheses were tested in this experiment. The first two pertained to Question 2.1. The last hypothesis was related to Question 2.2. These hypotheses and the rationale for each are presented, separately, below.

Question 2.1.

It was assumed that the French Immersion students who participated in the study would read and comprehend test questions less well in their second as compared to their first language. This deficiency in their reading ability was predicted to be sufficiently great as to cause F group students to achieve lower scores on the test than they would have obtained had they written the test in English. It was reasoned that if F group students were given unlimited time to write the test, this time would compensate for their relatively slow reading rate by allowing them to attempt all test questions. The provision of extra time would be unlikely to do anything, however, to compensate for their incomplete reading comprehension of the test questions, given that limited second language knowledge and skill proficiency was the assumed source of this deficiency. In this respect, the addition of extra writing time would have little or no effect on F group test outcomes.

Research indicated that reading comprehension is related to processing efficiency; the slower and less automatic the processing, the poorer the quality of comprehension (Laberge & Samuels, 1974; Perfetti, 1985). This implied that any F group student who was unable to complete the test because of his or her slower speed could also have had the poorest level of comprehension. This suggested that anyone who ran out of time might not have been able to understand the test questions completely even if he or she had had time to read them. From this it was inferred that the gain in scores that would accrue from having additional time to write the test would be minimal. Given this prediction, the following hypothesis was put forth:

Hypothesis 2.1.1: The scores of examinees who write the English form of the test will be greater than those achieved by students who write the French version.

The scores of F group students in Experiment 1 were expected to be more depressed, relative to E group scores, on data-based than on discrete items. As explained in the rationale for Hypothesis 2.1.1, the provision of extra writing time would not have a significant effect on the quality of F group students' reading comprehension. This implied that comprehension differences predicted to occur for F group students on discrete and data-based questions would be unaffected by the provision of more time. Unlimited writing time would, therefore, have influenced the difference in effect sizes on discrete and data-based questions only if F group students were able to correctly answer data-based questions that they would have been unable to respond to if they had a shortage of time. Given this prediction, the following hypothesis was put forth:

Hypothesis 2.1.2: Group differences in scores will be greater on data-based than on discrete knowledge items.

Question 2.2.

It was hypothesized earlier that, because of their lessened ability to comprehend the test questions, F group students would achieve lower scores on the test than would E group students. It was also argued that providing unlimited time to write the test would not alleviate this difference in scores because time, as such, would have had little or no effect on

the ability of F group students to comprehend what they had read. This implied that the scores of F group students writing with no time limit would not be significantly different from those of F group students who wrote with time limits.

E group students writing under standardized timed conditions would be unlikely to experience difficulty completing the test because the time limit set for the test was that which was appropriate for native speakers of English, and E group students in this study were native speakers of English. This implied that their scores would be unaffected by the time limit. Thus, all other things being equal, their scores and those of E group students in Experiment 2 should have been the same. Given these assumptions about the effect that unlimited time will have on the scores of E and F group students in Experiment 2 the following hypothesis was put forth:

Hypothesis 2.2.1: There will be no significant difference in the main effects for experiments 1 and 2.

CHAPTER IV

Methods and Results

The two experiments that were run and the results that were achieved in each, are presented and discussed in this chapter.

Experiment 1

Subjects

Six urban elementary schools in central Alberta provided the setting for Experiment 1. Permission to carry out this research was obtained from each school's respective central office administration. The schools that were selected for the study drew children from families of similar middle to upper middle class socio-economic backgrounds. None of these schools could be considered to have had children in their French Immersion programs who were from disadvantaged families.

Most, if not all, of the children in the classes under study had followed the usual pattern of early total immersion in which Kindergarten and Grade 1 were totally taught in French, followed by the introduction of English language arts in grades 2 or 3. At the time of the study, at least 60% of all of their school subjects were being taught in French, including social studies.

Instruments

Two tests were administered to determine if social studies achievement as measured in French differed from social studies achievement as measured in English: *The Grade 6 Social Studies Achievement Test Part A: Multiple Choice* (Student Evaluation Branch, 1985a) and its French translation, *Test de Rendement Etudes sociales 6e année Partie A: Choix multiples*

(Student Evaluation Branch, 1985b). These tests are described below.

Grade 6 Social Studies Achievement Test Part A: Multiple Choice.

The *Grade 6 Social Studies Achievement Test Part A: Multiple Choice* was developed by this author under the auspices of Alberta Education. Its purpose was to provide educators, trustees, and others with information about levels of social studies achievement at local and provincial levels. The test measures student knowledge and skills in relation to social studies program objectives (Student Evaluation Branch, 1984). Its content emphasis is derived from the *Grade 6 Social Studies Curriculum Specifications* (Curriculum Branch, 1984). The test has 50 items with an administration time of 50 minutes. Examinees are required to use separate machine-scorable answer sheets.

Although the design mean of the test was 62.5%, the provincial average, when administered in 1985, was 29.9 out of 50 or 59.8%. The standard deviation was 8.5. Items ranged in difficulty (p-value) from .36 to .88 and had discrimination values no lower than .200 (Student Evaluation Branch, 1985c). For reporting purposes, the test items were grouped into the following categories (subtests):

1. Topic A: All questions related to how people in ancient times met their physical, psychological, and social needs.
2. Topic B: All questions related to how people in Eastern societies meet their physical, psychological, and social needs.
3. Topic C: All questions related to meeting physical, psychological, and social needs through local, provincial, and federal government.
4. Recall & Comp. A: Recalls and understands facts, concepts, and generalizations related to how people in ancient times met their needs.
5. Recall & Comp. B: Recalls and understands facts, concepts, and generalizations related to how people in Eastern societies meet their needs today.
6. Recall & Comp. C: Recalls and understands facts, concepts, and generalizations related to meeting needs through local, provincial, and federal government.
7. Values: Recalls and understands competing values and uses skills to analyse competing value positions.
8. Inquiry I: Uses skills related to identifying elements of an issue, formulating research questions and procedures, and gathering data.

9. Inquiry II: Uses skills related to analysing, evaluating, and synthesizing data.
10. Inquiry III: Uses skills related to resolving issues, planning courses of action, and evaluating decisions and courses of action.

The distribution of questions, by category, is provided, along with the tests, in the Appendix.

Prior to its administration, the test was reviewed for content validity, accuracy, and technical merit by Grade 6 social studies teachers from all parts of the province and by a test review committee. No evidence of the construct validity of the reporting categories has been provided.

Test de Rendement Etudes sociales 6e année Partie A: Choix multiples.

Following final review by the Test Review Committee, the *Grade 6 Social Studies Achievement Test Part A: Multiple Choice* was professionally translated by Alberta Education as a service to school jurisdictions offering Grade 6 social studies in French. Students who were taught social studies in French were exempted from writing the provincial achievement test in that subject in 1985 (Student Evaluation Branch, 1984). However, a number of schools offering it in French opted to have their students write the achievement test in French. The scores of the French Immersion students who wrote the *Test de Rendement Etudes sociales 6e année Partie A: Choix multiples* in 1985 are not available.

Data Collection Procedures

The *Grade 6 Social Studies Achievement Test Part A: Multiple Choice* and its French translation *Test de Rendement Etudes sociales 6e année Partie A: Choix multiples* were administered to French Immersion students by their social studies teachers during the first two weeks of June, 1986. Each version of the test was randomly distributed to half of the students in each of the eight classrooms tested. In all, 95 students wrote the English version and 84 students wrote the test in French. Teachers were instructed to follow the standardized administration procedures developed for these tests, including the 50 minute time limit. The only variation from the original administration procedures was that student instructions and sample questions were presented in both French and English. Consistent with the provincial

administration of these tests, the students were told that their marks would not count toward their final grades but that it was important that they do their very best. It was indicated that the purpose of the study was to determine if the scores of those who wrote the French version would be the same as those achieved by the students who wrote the English form.

Data Analysis

In order to test the mean differences in student scores on French and English versions of this social studies achievement test, *t* tests for independent samples were performed on the total test and reporting category mean scores of the E and F groups. The .05 level of significance was used in testing the hypotheses. Effect sizes were used as a means of comparing group differences in performance across reporting categories. These effect sizes were calculated using either the standard deviations from the 1985 provincial administration of the achievement test or the pooled standard deviations of the E and F groups. Any difference in effect size greater than .25 was treated as important.

Results

Two questions were examined in this experiment. The results pertaining to each question are addressed, separately, below.

Question 1.1.

Do French Immersion students achieve similar scores when they respond to French and English forms of a standardized test of social studies achievement?

The means and standard deviations of the E and F groups on the total test and the subtest reporting categories used in reporting the 1985 provincial achievement testing results are presented in Table 4-1. The results show that, as predicted in Hypothesis 1.1.1, examinees who wrote the English form of the test achieved significantly higher scores on the total test and on all subtest reporting categories than did those who wrote the French version.

What is apparent from the data is that these differences in scores are systematic (i.e., unidirectional), [illegible]ose who wrote in English, across all reporting categories. This

Table 4-1
Summary Results Using Alberta Education Reporting Categories
Experiment 1

Reporting Category	Mean			Standard Deviation			Significance[1]	Effect Size[2]
	E Group	F Group	Prov.[3]	E Group	F Group	Prov.		
Total Test	31.5	24.5	29.9	6.9	7.3	8.5	***	.83
Topic A	11.0	7.8	10.7	2.5	2.9	3.2	***	.98
Topic B	10.3	8.8	9.7	2.7	3.0	3.1	***	.48
Topic C	10.3	7.8	.5	3.0	3.3	3.5	***	.70
Recall & Comp. A	4.8	4.0	5.0	1.4	1.6	1.4	***	.51
Recall & Comp. B	4.3	3.7	4.1	1.3	1.7	1.6	*	.34
Recall & Comp. C	4.3	3.2	4.1	1.8	2.1	2.0	***	.55
Values	3.9	2.6	3.2	1.3	1.1	1.5	***	.84
Inquiry I	5.4	4.0	5.0	1.6	1.4	1.7	***	.81
Inquiry II	6.0	4.5	5.4	1.6	1.9	2.0	***	.76
Inquiry III	3.3	2.8	3.1	1.4	1.4	1.5	*	.29

[1] Probability levels of t tests of significance.
[2] Effect sizes are calculated in relation to the standard deviations obtained when the Grade 6 Social Studies Achievement Test was administered provincially to English program students in Alberta in 1985. Any difference greater than .25 is treated as important.
[3] Mean scores achieved when the Grade 6 Social Studies Achievement Test was administered provincially to English program students in Alberta in 1985.
*$p<.05$. **$p<.01$. ***$p<.001$.

systematicity in the results may indicate that one or more language of testing variables had a pervasive influence on F and/or E group behavior across the whole test.

While the differences in scores across the tests are systematic in their direction, they are not constant in their magnitude; relative to provincial standard deviations, the effect sizes range from .29 to .98 across reporting categories. Since variations in effect sizes of this magnitude are unlikely to have occurred by chance, it can be inferred from these results that there is a relationship between the category of question being asked and the size of the differences in E and F group scores. This implies that variables related to the way these items were grouped affected how students responded to those questions in French and English. In short, while there may have been general factors contributing to the differences in scores across all items, factors specific to individual reporting categories may also have had varying degrees of influence on examinee behavior.

That there would be a variation in effect sizes relative to item groupings was anticipated and expressed in Hypothesis 1.1.2. Specifically, it was hypothesized that effect sizes would be greater on items that were data-based than on those that were discrete. This hypothesis was based on the assumption that reading ease would differ across these groupings of items and that F group behavior would be more significantly affected by this difference in reading ease than would E group behavior.

The test questions used in this study were grouped, for reporting purposes, as they were for the 1985 provincial administration of the test. The results in 1985 were not specifically reported in relation to item type, that is, according to whether the items were discrete or data based. To be consistent, the data from this experiment were not analysed in this manner either. Nevertheless, it is possible to infer from the results what the magnitude of the group differences in scores is on discrete and data-based items because three reporting categories (i.e., Recall & Comp. A; Recall & Comp. B; and Recall & Comp. C) are composed of discrete items only and all of the questions in reporting categories Values, Inquiry I, Inquiry II, and Inquiry III are data-based. When the results in Table 4-1 are examined in relation to these item groupings, it can be seen that there is some support for

Hypothesis 1.1.2: Except for the reporting category Inquiry Skills III, the effect sizes on reporting categories composed of data-based items are larger than are those for categories composed of discrete items. This may indicate that E and F group performances on the questions were related to item type.

Another trend in the data suggests that item type may not have been the only variable related to the differences in E and F group performance. The results show that there are considerable variations in the size of the language of testing effects among the three topic-specific reporting categories: The effect size is greatest on those questions covering material from Topic A (ancient civilizations), followed by those from Topic C (governments in Canada), and finally those from Topic B (Southeast Asian societies). These results seem to suggest that the ease with which students were able to answer the questions in one language as compared to the other was related to the conceptual content of the question, as defined by its curriculum topic of study. In other words, the magnitude of the language of testing effect seems to have varied in relation to what the questions were about.

One other trend in the figures in Table 4-1 is worth noting. The data reveal that the pattern of effect sizes across the topic-specific comprehension reporting categories is not consistent with that which is present across the three reporting categories that reflect all questions (i.e., recall and comprehension, value, and skill items) in each topic. In particular, the effect sizes for Recall & Comp. A and Recall & Comp. C are more similar to each other than are those for Topic A and Topic C in their entirety.

This discrepancy is of interest because reporting categories Recall and Comp. A and Recall and Comp. C are subsets of Topics A and C, respectively. This means that for the effect sizes on topics A and C to have differed from each other as much as they do, then the language of testing effect must have been greater on the Topic A data-based questions than on either the Topic A knowledge-based or the Topic C data-based questions. In other words, effect sizes must have varied among the items within topics as well as among items within item types. This implies that either item topic and item type variables interacted to affect E and F group behavior differentially across reporting categories or that the apparent

relationships between examinee behavior and item topic and between examinee behavior and item type were in fact spurious and that some other variable or group of variables produced the results seen in Table 4-1.

So as to investigate these apparent relationships between effect size and item topic and item type further, the items were regrouped according to topic and type and new average scores were calculated for the E and F groups. These figures are presented in Table 4-2. Unlike those presented in Table 4-1, the effect sizes for these new groupings were calculated in relation to the pooled standard deviations of the E and F groups, and not the provincial standard deviations, as these figures were not available.

The composition of reporting categories Topic A, Topic B , and Topic C are identical in tables 4-1 and 4-2. This means that the same pattern of effect sizes are present for the reporting categories in both tables. Similarly, reporting categories Discrete A, Discrete B, and Discrete C in Table 4-2 are simply the reporting categories Recall & Comp A, Recall & Comp. B, and Recall & Comp. C, from Table 4-1, renamed. This renaming has been done to emphasize in what way the items in these categories differ from those in the data-based reporting categories. Because these reporting categories are identical to those in Table 4-1, there is no new information to be gained from these portions of Table 4-2. Instead, this data is presented as a way of providing a context for that information which is new in Table 4-2. What is unique in Table 4-2 is the grouping of all discrete items into one reporting category (Discrete), the grouping of all data-based items into another reporting category (Data), and the regrouping of items from the value and skill reporting categories in Table 4-1 into data-based categories that are topic specific (Data A, Data B, and Data C).

The data in Table 4-2 indicate that when items are pooled according to whether they are discrete or data-based, the resulting effect sizes are considerably different. These data seem to indicate support for Hypothesis 1.1.2 because it is apparent that group differences are greater on data-based than on discrete items. Closer scrutiny reveals, however, that the relationship between item type and effect size may not be as simple as that assumed by Hypothesis 1.1.2. When the items in reporting categories Discrete and Data are further

Table 4-2

Summary Results Using Reconstructed Reporting Categories

Experiment 1

Reporting Category[1]	Mean		Standard Deviation		Effect Size[2]
	E Group	F Group	E Group	F Group	
Topic A	11.0	7.8	2.5	2.9	1.19
Topic B	10.3	8.8	2.7	3.0	.53
Topic C	10.3	7.8	3.0	3.3	.80
Discrete	13.4	10.9	3.4	3.9	.69
Data	18.2	13.5	4.2	4.1	1.13
Discrete A	4.8	4.0	1.4	1.6	.54
Discrete B	4.3	3.7	1.3	1.7	.40
Discrete C	4.3	3.2	1.8	2.1	.57
Data A	6.2	3.8	1.7	1.8	1.37
Data B	6.0	5.1	1.9	2.0	.46
Data C	6.0	4.6	1.9	2.0	.72

[1]Discrete A; Discrete B; & Discrete C are the reporting categories Recall & Comp. A; Recall & Comp. B; and Recall & Comp. C from Table 4-1, renamed.

[2] Effect sizes are calculated in relation to the pooled standard deviations. Any difference greater than .25 is treated as important.

subdivided by topic, the resulting patterns of effect sizes are not consistent with the overall finding. In particular, there are no real differences in the effect sizes for Discrete B and Data B, and only a difference of .15 in the effect sizes for Discrete C and Data C. This indicates that the relationship specified by Hypothesis 1.1.2 only holds for items from Topic A.

It appears from the data in Table 4-2 that the situation may be somewhat similar in regard to the apparent relationship between item topics and effect sizes. There are important differences in effect sizes across the data-based reporting categories that are topic specific. However, no significant differences are present among the topic-specific reporting categories made up of discrete items. This suggests that variables related to item topics may be related to differences in E and F group scores on data-based but not on discrete items.

In summary, the scores of the two groups of French Immersion students participating in this study are consistently different from each other, with the E group achieving higher scores than the F group. While the differences in scores across all reporting categories are significant, the size of those differences varies. In particular, the effect sizes on topic specific data-based questions are all larger than the effect sizes on the same topic discrete item reporting categories. The most notable differences in E and F group scores seem to have occurred in relation to the data-based items from Topic A.

Question 1.2.

Are French Immersion students equally able to complete French and English forms of a standardized test of social studies achievement within the time limit established for the test?

The item analyses indicates that, while all of the E group students were able to complete the test in the time given, nine out of the 84 F group students who wrote were not. This finding suggests that it took examinees more time to read the French form of the test than the English version. However, no conclusion can be drawn about the cause of this apparent difference in reading rates. It may be that examinees had slower reading rates in their second language and therefore required more time to process the test questions in French as compared to English. On the other hand, it may be that the French form had more text to read than the English one. In other words, the French test may have had more and/or longer

text as a result of the translation process or simply as a function of natural differences in the two languages.

This unequal rate of completion across groups may explain why students had lower scores when they wrote the tests in French as compared to English. Students writing the French form may have felt a time pressure and may, therefore, have rushed to get through the test. This could have caused even those students who completed the test to score less well than they would have, had they written the test in English.

In order to determine whether or not the ability to finish the test in the time given was a crucial variable underlying the language of testing effect, it was decided to perform a second experiment. All of the conditions of Experiment 1 were replicated in this second experiment, except that examinees were given unlimited time to complete the test. The design of Experiment 2 and the results that were achieved are described below.

Experiment 2

Subjects

Six urban elementary schools in central Alberta provided the setting for Experiment 2. A process similar to that used to obtain permission to carry out Experiment 1 was undertaken. The schools that were selected for the study drew children who had the same or similar characteristics as the children who participated in Experiment 1. That is, the students came from families of middle to upper middle class socio-economic backgrounds. None of the schools had children in their French Immersion programs who were from disadvantaged families.

As with Experiment 1, most, if not all, of the children in the classes under study had followed the usual pattern of early total immersion in which Kindergarten and Grade 1 were totally taught in French, followed by the introduction of English language arts in grades 2 or 3. At the time of the study, at least 60% of all of their school subjects were being taught in French, including social studies.

Instruments

The same two tests that were administered in Experiment 1 were administered in Experiment 2. These were: *Grade 6 Social Studies Achievement Test Part A: Multiple Choice* and its French translation, *Test de Rendement Etudes sociales 6e année Partie A: Choix multiples*.

Data Collection Procedures

The tests were administered to French Immersion students by their social studies teachers during the first two weeks of June, 1987. Each version of the test was randomly distributed to half of the students in each classroom. In all 72 children wrote the English form and 75 wrote the French version. Teachers were instructed to follow the same standardized administration procedures as were used in Experiment 1. The only variation from the original procedures was that students were given unlimited time to complete the test. Consistent with the provincial administration of these tests, the students were told that their marks would not count toward their final grades but that it was important that they do their very best. It was indicated that the purpose of their writing the test was to determine if the students who wrote the French version would achieved the same scores as those who wrote in English.

Data Analysis

In order to test the mean differences in student scores on French and English versions of this social studies achievement test, t tests for independent samples were performed on the total test and reporting category mean scores of the E and F groups. The .05 level of significance was used in testing the hypotheses. Effect sizes were used as a means of comparing group differences in performance across reporting categories. Effect sizes were calculated using either the standard deviations from the 1985 provincial administration of the achievement test or the pooled standard deviations of the E and F groups. Any difference in effect size greater than .25 was treated as important.

Results

Two questions were examined in this experiment. The results pertaining to each question are examined, separately, below.

Question 2.1.

Given unlimited writing time, do French Immersion students achieve similar scores when they respond to French and English forms of a standardized test of social studies achievement?

The means and standard deviations of the E and F groups on the total test and the subtest reporting categories are presented in Table 4-3. These results show that in all cases the F group means are significantly lower than are those achieved by the E group. This indicates that, in spite of having unlimited time to complete the test, examinees who wrote the French form achieved significantly poorer performances on the total test and on all subtest reporting categories than did those who wrote the English version.

There is a broad range in the effect sizes across the test, indicating that the scores of the F group are more depressed, relative to those of the E group, on some categories of questions than on others. In particular, there seems to be a significantly greater depression of scores on Topic A items than on those from either topics B or C. This trend is present on both the total topic and the discrete item levels of reporting, suggesting that the effect size on Topic A data-based questions may also have been significantly different. These trends suggest that there is a relationship between the topic of the items and the magnitude of the discrepancy in E and F group scores.

The effect size on Inquiry III items is considerably smaller than are those on Inquiry I or Inquiry II items. These reporting categories consist of items involving increasing more complex data interpretation processes, with Inquiry I items requiring the least, and Inquiry III items the most, complex interpretions. These results may, therefore, indicate that effect sizes are related to the cognitive complexity of the items.

For the sake of clarification and ease of comparison with the results from Experiment 1, the items were regrouped according to their topic and type and new mean scores were calculated. Effect sizes for these new groupings were calculated using the pooled standard

Table 4-3
Summary Results Using Alberta Education Reporting Categories
Experiment 2

Reporting Category	Mean			Standard Deviation			Significance[1]	Effect Size[2]
	E Group	F Group	Prov.[3]	E Group	F Group	Prov.		
Total Test	32.1	24.8	29.9	7.4	7.8	8.5	***	.86
Topic A	10.9	7.6	10.7	2.9	3.3	3.2	***	1.03
Topic B	10.5	8.5	9.7	2.6	3.3	3.1	***	.65
Topic C	10.7	8.8	.5	3.1	2.7	3.5	***	.54
Recall & Comp. A	4.8	3.7	5.0	1.5	1.6	1.4	***	.79
Recall & Comp. B	4.2	3.6	4.1	1.5	1.7	1.6	*	.38
Recall & Comp. C	4.5	4.0	4.1	1.9	1.8	2.0		.25
Values	3.8	2.7	3.2	1.4	1.3	1.5	***	.73
Inquiry I	5.5	3.9	5.0	1.5	1.5	1.7	***	.94
Inquiry II	6.2	4.5	5.4	1.5	1.8	2.0	***	.85
Inquiry III	3.4	2.5	3.1	1.5	1.6	1.5	***	.60

[1] Probability levels of t tests of significance.
[2] Effect sizes are calculated in relation to the standard deviations obtained when the Grade 6 Social Studies Achievement Test was administered provincially to English program students in Alberta in 1985. Any difference greater than .25 is treated as important.
[3] Mean scores achieved when the Grade 6 Social Studies Achievement Test was administered provincially to English program students in Alberta in 1985.
*$p<.05$. **$p<.01$. ***$p<.001$.

deviations of the E and F groups, as was done in Experiment 1. These figures are presented in Table 4-4.

Two trends are noticeable in the data. First, when items are grouped by type, the effect size on the discrete items is considerably different from that on the items that are data-based. This trend is present across all three topics, thereby providing strong support for Hypothesis 2.1.2. Second, when items are grouped by topic, the effect size on those items pertaining to the study of ancient civiliations (Topic A) is significantly greater than are those for topics B or C. This trend holds regardless of whether the items are discrete or data-based.

In summary, there are significant differences in group scores across all reporting categories, in favor of those who wrote in English. The magnitude of these differences is greater on data-based items than on those that are discrete, regardless of which topics of study those items reflect. The effect sizes on Topic A discrete and data-based items are considerably greater than are those on the same type items from the other topics. Finally, when items are grouped by level of skill complexity, the effect size is smallest on those items judged by the test developers to be the most cognitively complex.

It is clear from these results that the French Immersion students participating in this study did not achieve the same scores when responding to French and English forms of the test, even though they had unlimited time to complete those forms.

Question 2.2.

How do the scores of French Immersion students who were given unlimited time to write compare to those achieved by the groups who wrote with time limits?

The total test mean scores of the E and F groups from experiments 1 and 2 (see tables 4-1 and 4-3, respectively) were compared using ANOVA. There is no significant difference in overall effect sizes across the two experiments. A glance back to the earlier tables reveals other similarities in the data across years. First, and most importantly, in both experiments F group scores are significantly and consistently lower than E group scores across all reporting categories. The amount that they differ varies across reporting categories with a similar range

Table 4-4

Summary Results Using Reconstructed Reporting Categories

Experiment 2

Reporting Category[1]	Mean		Standard Deviation		Effect Size[2]
	E Group	F Group	E Group	F Group	
Topic A	10.9	7.6	2.9	3.3	1.06
Topic B	10.5	8.5	2.6	3.3	.67
Topic C	10.7	8.8	3.1	2.7	.65
Discrete	13.4	11.4	3.7	4.2	.51
Data	18.7	13.5	4.3	4.4	1.20
Discrete A	4.8	3.7	1.5	1.6	.71
Discrete B	4.2	3.6	1.5	1.7	.37
Discrete C	4.5	4.0	1.9	1.8	.27
Data A	6.2	3.8	1.9	2.2	1.17
Data B	6.3	4.9	1.6	2.0	.77
Data C	6.2	4.7	1.7	1.6	.91

[1]Discrete A; Discrete B; & Discrete C are the reporting categories Recall & Comp. A; Recall & Comp. B; and Recall & Comp. C from Table 4-3, renamed.

[2] Effect sizes are calculated in relation to the pooled standard deviations. Any difference greater than .25 is treated as important.

in effect sizes across experimental conditions. Second, when skill-based items are grouped by their level of cognitive complexity, the effect size is smallest on those items that are most complex. Third, the effect sizes on topic-specific data-based questions are consistently greater than the effect sizes on the same topic discrete items. Fourth, the rank order of the effect sizes on the topic-specific data-based questions is the same across experiments. Finally, when items are grouped by topic and item type, the largest effect size is on data-based items from Topic A.

Given these similarities in the results across years, a number of tentative conclusions are possible. The first of these conclusions is that the performance of the E and F groups in Experiment 2 did not differ substantially from that of their counterparts in Experiment 1. The second conclusion is that, in both experiments, the magnitude of the effect sizes was related to the complexity of the cognitive processing requirements of the items. Finally, if one accepts the assumption of group comparability across experiments, then it can be concluded from these findings that the provision of extra writing time had no significant effect on the pattern of results in the second experiment. This implies that the differences in E and F group scores across experiments were not related to the ability to complete the test in the time given.

While there are many important similarities in the data from the two experiments there are also some notable differences. These differences need to be considered before it can reasonably be concluded that the provision of extra time had little or no effect on examinee behavior. The data show that the effect sizes for Discrete A and Discrete C reporting categories differ significantly in Experiment 2 but not in Experiment 1. At the same time, there are significant differences in the effect sizes for reporting categories Data B and Data C in Experiment 1 but not in Experiment 2. These trends are indicative of some of the remarkable changes that occurred in the effect sizes for specific reporting categories across experiments. Of particular note are the increases in effect size on reporting categories Inquiry III and Data B, and the decrease in effect size on Recall & Comp. C (Discrete C).

Increases in the effect size for any given reporting category, from Experiment 1 to Experiment 2, indicate that the discrepancy in E and F group scores was greater in Experiment 2 than it was in Experiment 1. This implies that F group scores decreased, or E group scores increased, or both. The opposite is true of a decrease in effect size: This implies an increase in F scores, a decrease in E scores, or both.

The results show that, in Experiment 2, the F group mean score on Inquiry III items is .3 lower than that achieved by the F group in Experiment 1; the E group mean is .1 higher than that of its counterpart. Similarly, the F group mean on Data B items is .2 lower than the F group mean on those items in Experiment 1, but the E group means in Experiment 2 increased by .3 over those in Experiment 1. In other words, in both of these situations, the Experiment 2 F group means decreased while the E group means increased. This implies that the increase in effect sizes across experiments, on these two reporting categories, can be attributed more to the relatively better performance of E group students than to the relatively poorer performance of F group students.

For two reasons, these results do not support the notion that time was responsible for any differences in E and F group performance. First, all E group students in Experiment 1 completed the test in the time given. Because experience has shown that French Immersion students have better than average first language reading abilities and that examinees who have this level of reading proficiency have no difficulty completing the test in the time given, it is assumed that this result indicates not only that they were able to complete the test but also that they were under no pressure to do so. This implies that the increase in E group performance in Experiment 2, relative to that in Experiment 1, was unrelated to the increase in writing time. Second, logic suggests that if time had had any effect on F group behavior in Experiment 2, that effect would have been to increase examinees' possibilities of selecting correct answers, not to decrease them. This means that their scores should have increased in Experiment 2. Given this premise, one can only account for the pattern of F group scores across experiments by assuming that time had no effect, or that that effect was masked by F group differences across years. This hypothesis seems unlikely, given that E group scores in

the second experiment were as good as or better than E group scores in Experiment 1. Thus, aside from the possibility of measurement error, the only reasonable conclusion is that, in Experiment 2, F group performance on these items was unaffected by the provision of extra time.

A somewhat different story emerges in relation to the effect size decrease on Recall & Comp. C (Discrete C) items. The data show that the mean scores of the F and E groups in Experiment 2 are higher than are those in Experiment 1 by .8 and .2, respectively. This pattern suggests that the decrease in effect size in Experiment 2 on this reporting category is, in large measure, attributable to the increase in the Experiment 2 F group mean on these items. The question is, was this increase related to the provision of unlimited writing time? It was noted earlier that eight out of the final nine questions on the test used in this study were recall and comprehension questions from Topic C. If F group students were pressured for time in Experiment 1, and the results say that at least some of them were, then the most obvious effect of this pressure would have been an inability to complete the final items on the test. Providing unlimited writing time in Experiment 2 would have meant that unlike those in Experiment 1, all F group students would have had the opportunity to complete all of these questions on the test, including those recall and comprehension questions from Topic C. This suggests that time may have been a factor contributing to the differences in F group scores on this reporting category across experiments.

In summary, there are a number of important similarities and consistencies in the data across the two experiments. These include the significant depression of F group scores, relative to those of the E group, across all reporting categories, as well as a similar pattern in the magnitude of group differences relative to the composition of reporting categories. These commonalities in the data suggest that the same factors were responsible for the depression of F group scores in both experiments. It appears that time pressures are not a *major* contributing factor in the depression of F group scores, because the provision of unlimited time does not seem to have made any significant difference in the scores.

There were, however, some notable differences in the data. While some of these differences could be attributed to group differences across experimental conditions, the lower F group mean in Experiment 1 compared to Experiment 2 on Discrete C items may have been the result of students' inability to complete all of the questions. This suggests that the provision of extra time may have had some impact on the responses of F group students in Experiment 2.

CHAPTER V

Conclusions and Implications

This chapter provides a summary of the conclusions that were reached in terms of the research questions that were posed. Theoretical and practical implications are also presented as are suggestions for further research.

Conclusions

The major finding of this study is that there are significant differences in the scores of Grade 6 French Immersion students who write a standardized test of social studies achievement in French and English, respectively. This difference is systematic across all reporting categories, in favor of those who write in English. The general conclusion reached from this finding is that the performance of French Immersion students, on a test of social studies achievement, is affected by the language of testing. In this respect, there is support for Carey's (1980) argument that language of testing differences could account for all or a portion of any observed discrepancy in the performances of French Immersion and English program students.

While the differences in E and F group scores are systematic across the test in terms of their direction, their magnitude varies across reporting categories. Because items were grouped into reporting categories according to apparent similarities in what they measured, and effect sizes varied across these categories, it is assumed that this indicates a relationship between what is being measured and students' abilities to respond to those items in French and English.

There are considerable differences in the effect sizes on discrete and data-based items. Data-based items differ from discrete items in that they require examinees to read and interpret novel information to answer them correctly. Discrete items typically assess an examinee's ability to recall previously learned information. This means that data-based items

measure more complex cognitive processes than do discrete items. Because data-based and discrete items differ in the complexity of what they measure, and effect sizes vary in relation to these items types, it can be concluded that the amount that F group scores are depressed, relative to those of the E group, is related to the cognitive processing demands of the item.

The results showed that while all students who wrote the English form of the test were able to finish in the time given, some of those who wrote the French form were not. From this it can be concluded that it takes French Immersion examinees more time to read the French form of the test than to read the English version.

When students were given unlimited time to write the test, their scores did not differ significantly or in any apparently important way from those of E and F group students who wrote under timed conditions. This implies that test-taking speed (i.e., the ability to complete the test in the time given) is not a significant factor underlying the language of testing effect. It must be noted, however, that some F group students in Experiment 1 were unable to complete the test in the time given. From this it can be concluded that, while the provision of extra writing time has no significant impact on the group score of those writing the French test, it can have a substantial effect for any given individual. At the same time, however, it must be acknowledged that anyone who is unable to finish the test because of slower reading speed, is also likely to have poor second language reading comprehension. As a consequence, he or she may not be able to increase his or her score significantly by being given more time to complete the test.

Implications

The contribution that this study can make to educational research in general is that it illustrates how extraneous variables can influence test outcomes thereby confounding data interpretations. It also demonstrates why it is so important to conceptualize and then control or account for the effects of such variables when conducting educational research. The E group scores in both experiments were equal to or higher than the averages obtained when the test was administered, provincially, to English program students in 1985. Based on these E

group scores, one might conclude that through instruction in a French Immersion program, it is possible to develop proficiency in French while at the same time achieving levels of performance that are equal to provincial levels.

One's conclusions would be quite different if one only had access to the F group scores and had no awareness of the apparent underestimation of social studies knowledge and skill that they provide. Given the differences in F group and provincial averages, one might assume that the students in this study had levels of performance that were considerably lower than were those of Grade 6 children in English language programs in 1985. An interpretation of these results might lead to the conclusion that, while French Immersion students are acquiring proficiency in French, they are doing so at the expense of their academic performance in social studies. Clearly this conclusion about the efficacy of French Immersion programs and the former one are widely discrepant, even though both are based on the performance of the same group of students.

In addition to showing how important it is to conceptualize and control confounding variables, the results of this study have a number of theoretical and practical implications that are more directly related to the issue of French Immersion testing. These implications are discussed below.

Theoretical Implications

While it is important to know that there is a language of testing effect on scores, it is as or more important to know what caused French Immersion students to achieve lower scores when they wrote in French compared to English. This understanding of the causal agents is important because it is only through knowing what the source of that depression is that one is able (a) to control or eliminate that element as a confounding factor or (b) to account for its impact on test outcomes.

Two factors that could be causally related to effect sizes were examined in Chapter II. These factors include the equivalence of the French and English forms of the test and the first and second language reading abilities of the examinees. Test equivalence was defined in

terms of item difficulty and text readability. First and second language reading abilities were considered in terms of levels of language-related knowledge and skill and the likelihood, given these levels, of achieving full comprehension of the textual material that was read.

Because of technical limitations, the equivalence of the two tests, in terms of their item difficulties and/or levels of readability could not be established. Similarly, the first and second language reading abilities of the students participating in this study were not determined. Consequently, it was not possible to determine what impact each of these variables had on E and F group performance. Nevertheless, given some of the trends in the data from the two studies, it is possible to develop some hunches about the effect of these variables on test outcomes.

The results showed that effect sizes were significantly larger on data-based skill items than they were on discrete knowledge-based items. This implies that F group scores were more depressed on the former type of question than the latter. Similarly, there seems to be some relationship between the topic of the items and the amount that F group scores were depressed relative to those of the E group. These results suggest that the discrepancy in E and F group scores can be attributed to test inequivalencies if, and only if, one can explain how these inquivalencies could be greater on some types and topics of items then on others. Conversely, to promote the argument that it is reading ability differences that account for the depression of scores, one has to explain how there could be greater reading ability differences on one type and topic of items then on others.

One could argue for the possibility of item difficulty and/or item readability differences across question types if one assumed that it was the amount of textual material and not the level of cognitive complexity that was the salient feature distinquishing performance on discrete knowledge and data-based skill items. Simply put, it could be that the more text associated with an item the harder it is for a translator to maintain the meaning and/or linguistic style of the original writer. Since meaning is related to item difficulty and linguistic style is a determiner of reading ease, then it could be assumed that the greater the difference in the quality of these variables across the French and English forms, the greater

the difference in E and F group scores.

A similar argument could be presented to explain the tentative relationship between effect size and item topic. That is, it could be that the extent to which a translator is able to remain faithful to the original item, in terms of its meaning and style, is related to the content that is being translated. Perhaps, for example, it is easier to retain meanings or levels of readability across languages (English and French) when dealing with abstractions about government (Topic C) than with concepts such as psychological needs or social equality (Topic A). As a result, the item difficulties and/or readability levels of the item from one topic may have been more similar to that of the originals than were those in another topic.

The limitation of both of these explanations is that they assume that the effect of the argued translation difficulties on item difficulty or readability would be systematic. In other words, they assume that all of the items of a certain type or from a certain topic would have been made either systematically more or systematically less difficult or readable than the English originals. The research cited in Chapter II did not support such an assumption.

The alternate explanation for the differences in E and F group scores put forth in this thesis was that French Immersion students have unequal levels of first and second language reading abilities and these differences systematically affect their comprehension of the questions and, therefore, their ability to answer those questions. This means that to argue for this explanation in a way that is consistent with the trends in the data, it must be shown that variables related to item type and item topic could be causally related to E and F group students' relative abilities to read and comprehend the test questions.

The rationale for an assumed relationship among the variables reading ability, item type, and effect size was presented in Chapter III. Briefly, it was argued that as text difficulty increases F group students would pay a proportionately greater price in processing efficiency then would E group students with the result that their level of text comprehension would be negatively correlated with the complexity of the question. E group students would not pay the same "price" as F group students when processing increasing complex text because relative to F group students they are skilled readers. According to Frederiksen (1981), the

price one pays is related to skill level.

To be consistent with the theory that it was reading ability differences rather than test differences that accounted for the discrepancies in E and F group scores, it must also be shown that the reading ability of either the E or the F group could vary according to the conceptual content of items. Such an explanation is possible and rests on the premise that reading comprehension is dependent among other things, on the reader's ability to achieve lexical access, and that this in turn depends on one's familiarity with the language. Since lexical access is related to the range and richness of one's understanding of the language (Just & Carpenter, 1987; Perfetti, 1985) an understanding that comes from a broad and varied exposure to that language, then it could be argued that French Immersion students in this study had different levels of language related knowledge about one topic of study than another. This could result from different exposures to the language of each topic because of differences in the text books and other resource materials, or because of differences in the teachers' familiarity with the terminology from one topic compared to the other.

In summary, arguments that are consistent with the data trends (i.e., a relationship between effect size and item topic and/or item type) can be made to support the notion that test inequivalencies accounted for group differences in performance. Arguments can also be made to show that it is reading ability differences that underlie the language of testing effect, and that the variation in effect sizes relative to item topics and types is a function of differences in E and F group levels of language related knowledge and skill in these areas. The weight of these arguments is not equal however. Instead the hypothesized relationship between reading ability and effect size variations seems more compelling for two reasons.

First, the reading ability arguments are based on assumptions that are supported in the literature. None of the assumptions underlying the test inequivalencies arguments are supported. For example, there is no evidence that linguistic style is easier to maintain with some topics than with others. The second point arguing in favor of the notion that it is reader differences that accounts for the apparent topic related variations in effect size is that some F group students were unable to finish the test in the time given. This finding is

meaningful because it suggests that the French Immersion students in this study had unequal first and second language reading speeds. Since reading speed has been correlated with processing efficiency and processing efficiency is fundamental to text comprehension, then this pattern implies that the F group students may have achieved incomplete comprehension of what they read.

Practical Implications

The results of this study indicate that it does matter in which language French Immersion students are tested; the scores of the students participating in this study were significantly lower when they wrote in French as compared to English. What is implied by these results is a need for educators to make a decision about the language in which French Immersion students should be tested.

Alberta Education exempts from its Achievement Testing Program those students for whom the test is inappropriate (Student Evaluation and Records Branch, in press). Included in the list of students who are eligble for exemption are those whose language of instruction is other than English and/or those students who are enrolled in an English as a Second Language program. Implicit in these categories of exemption is the principle that for participation to be appropriate, the language of instruction and students' language of fluency must match the language of testing.

On the basis of the evidence presented in Chapter II and the patterns of results obtained in experiments 1 and 2, it seems reasonable to conclude that a probable cause of the depression of F group scores in this study was reading ability differences. If this is the case, then it is not possible when testing French Immersion students to achieve the principles for appropriate participation as set out by Alberta Education (Student Evaluation and Records Branch, in press) because the language of instruction and the language of fluency are not the same thing.

This situation poses a dilemma in choosing a language of testing. It is clear from the number of jurisdictions choosing to have their French Immersion students write these optional

tests (Student Evaluation and Records Branch, in press), that educators want to know how well these students are achieving the goals and objectives of the programs of study. However, regardless of which language of testing is chosen there is a price to be paid (Carey, 1980). It appears from this study that choosing to test in the language of instruction when it is not the language of fluency will cause scores to be artifically low. Such low scores could have negative political and pedagogical implicatons if they are interpreted to mean that the level of French Immersion achievement is lower than expected. On the other hand, choosing to test in the language of fluency rather than the language of instruction suggests that the gap in fluency between the dominant and nondominant language is sufficiently great as to cast doubts on the efficacy of the entire French Immersion process. Moreover, the act of testing French Immersion students in English could lead some teachers to introduce English terminology into their instruction as a way of "preping" students to write that form of the test. This could further jeopardize the integrity of the program.

If one accepts that unequal levels of first and second language reading ability was causally related to the depression of F group scores, then one can infer that French Immersion students do not achieve complete comprehension of the school based texts that they read in their second language. This inference is not startling, given Carey's (1987) finding that high school students who have received all or most of their schooling in French Immersion programs achieve unequal levels of comprehension when they read excerpts from French and English versions of a textbook.

What does seem unusual and therefore warrants some consideration, is why the students in this study had levels of achievement that were equal to or better than provincial averages if they had achieved incomplete comprehension of the materials they had read in class. In other words, if their test performance was negatively affected by their second language reading comprehension, why wasn't their classroom performance (as measured by the English version of the test) equally depressed?

Two explanations are possible. One is that their performance was negatively affected. That is, their performance was at or above provincial levels, not because they achieved at

their optimal level, but because as a self-selected group (Carey, 1984; West, 1985) they were so superior to the regular program students that even though their social studies achievement was depressed by their inability to understand what they read in class (in French) they were still able to outperform the regular English program students.

A second possibility is that because these students were from Anglophone communities and French resources were not always available to them (Acheson, 1986), they were able to acquire some understanding of the concepts from the program through their experiences, in and out of school, in English. If this is true, then it provides a strong argument for ensuring that French Immersion students continue to have at least part of their resource materials provided in English so that their academic achievement is not jeopardized.

It is common to find researchers and educators who are willing to argue that parallel tests cannot be produced through translation and that it is this lack of parallelism that is responsible for effects such as that which were present in this study. The patterns of results in this study imply that contrary to this perspective, it is *not* the use of translations, per se, that is the problem. Rather it is the assumption that L2 students can perform equivalently in their nondominant and dominant languages that is problematic. In other words, it may be inappropriate to test French Immersion students using a translation, not because it diverges from the source (English) test but, ironically, because it remains too faithful: By maintaining the linguistic style of the original, the translator produces a version of the test that is appropriate for a native language reader. This is an important distinction and one worth pointing out to those who may have erroneously concluded from their findings (e.g., Scoon, 1974) that instrumen. differences and not unequal levels of language ability accounted for variations in group scores. In short, the use of transations may have been condemned in the past for the wrong reasons.

This inference has major consequences for test construction and administration. Given the increasingly multicultural nature of North American society, as well as the increased emphasis being placed on international assessments the demand for standardized tests in multiple languages is constantly growing. If equivalent tests in multiple languages can

be produced through translation then a significant economic burden is eliminated from such testing initiatives because test construction is a very costly and time consuming enterprise. Equally significant are the problems that would be eliminated in trying to equate parallel tests produced in separate languages.

Suggestions for Further Research

The ability to establish the parallelism of test forms in two languages is fundamentally related to the ability to answer an underlying question in this study which was "To what extent do test differences rather than reader differences account for the variation in scores across languages of testing?" The parallelism or equivalence of the two forms can only be established if one knows what text-related factors affect examinee behavior and then one is able to compare and contrast the tests in relation to these variables. This requires a technically and theoretically adequate enumeration and definition of the factors that contribute to or diminish test parallelism either within or across languages.

The variation in effect sizes in relation to item topic and type suggests that the amount that French Immersion students' scores will be depressed when they write a test in French rather than English will be inconstant across subject areas or even across test forms within a subject area. Similarly, the suspected relationship between first and second language reading abilities and effect sizes implies that the language of testing effect could vary according to the number of years students have spent in a French Immersion program. For these reasons it is suggested that this study be replicated with French Immersion students in different grade levels and in different subject areas.

V. REFERENCES

Acheson, J. B. (1986). *French immersion program survey*. Edmonton: Department of Program Services, Edmonton Catholic School District.

Adams, M. (1980). Failure to comprehend and levels of processing in reading. In R. J. Spiro, B. C. Bruce, & W. Brewer, (Eds.), *Theoretical issues in reading comprehension* (pp. 7-9). Hillsdale, NJ: Erlbaum.

Alderson, J. (1984). Reading in a foreign language: a reading problem or a language problem? In J. C. Alderson & A. H. Urquhart (Eds.), *Reading in a foreign language* pp. 1-24). New York: Longman.

Anderson, R. & Shifrin, Z. (1980). The meaning of words in context. In R. J. Spiro, B. C. Bruce, & W. Brewer, (Eds.), *Theoretical issues in reading comprehension* (pp. 331-348). Hillsdale, NJ: Erlbaum.

Atkins, B., Duval, A., & Milne, R. (1988). *Collins-Robert French-English English-French Dictionary* (2nd ed.). London: Collins.

Bain, B., and Yu, A. (1987). Issues in second-language education in Canada. In L. Stewin, & S. McCann (Eds.). *Contemporary educational issues - The Canadian mosaic* (pp. 215-225). Toronto: Copp Clark Pitman.

Barnett, M. (1986). Syntactic and lexical/semantic skill in foreign language reading: Importance and interaction. *The Modern Language Journal*, 70, 343-349.

Beheydt, L. (1987). The semantization of vocabulary in foreign language learning. *System*, 15, 55-67.

Berman, R. (1984). Syntactic components of the foreign language reading process. In J. C. Alderson & A. H. Urquhart (Eds.), *Reading in a foreign language* (pp. 139-159). New York: Longman.

Block, E. (1986). The comprehension strategies of second language readers. *TESOL Quarterly*, 20, 463-494.

Bloom, B., Madaus, G., & Hastings, J. (1981). *Evaluation to improve learning*. New York:

McGraw-Hill.

Bussis, A.M. and Chittenden, E. (1987). Research currents: What the reading tests neglect. *Language Arts*, 64, 302-308.

Capell, F. & de Porcel, A. (1979). Assessment of reading achievement in two languages: New methods for studying bias. In R. Silverstein, (Ed.), *Proceedings of the Third International Conference on Frontiers in Language Proficiency and Dominance Testing* (pp. 96-112). Carbondale: Southern Illinois University.

Carey, S. (1980). *Student evaluation in French programs in Alberta*. Edmonton, AB: Alberta Education.

Carey, S. (1984). Reflections on a decade of French immersion. *Canadian Modern Language Review*, 41, 246-259.

Carey, S. (1987). Reading comprehension in first and second languages of Immersion and Francophone students. *Canadian Journal for Exceptional Children*. 3, 103-108.

Chaudron, C. (1983). Foreigner talk in the classroom. An aid to learning? In H. Seliger & M. Long (Eds.), *Classroom oriented research in second language acquisition* (pp. 127-145). London: Newbury House.

Conrad, C. (1972). Cognitive economy in semantic memory. *Journal of Experimental Psychology*, 92, 149-154.

Cooper, M. (1984). Linguistic competence of practised and unpractised non-native readers of English. In J. C. Alderson & A. H. Urquhart (Eds.), *Reading in a foreign language* (pp. 122-138). New York: Longman.

Cowan, J.R. (1976). Reading, perceptual strategies and contrastive analysis. *Language Learning*. 26, 95-109.

Cronbach, L. (1971). Test validation. In R. Thorndike (Ed.), *Educational measurement* (2nd ed.) (pp. 443-507). Washington, D.C.: American Council on Education.

Cummins, J. (1983). Language proficiency, biliteracy and French immersion. *Canadian Journal of Education*, 8(2), 117-137.

Cummins, J. (1987). Immersion programs: Current issues and future directions. In L.

Stewin, & S. McCann (Eds.). *Contemporary educational issues - The Canadian mosaic* (pp. 192-206). Toronto: Copp Clark Pitman.

Curriculum Branch, (1984). *Grade 6 social studies curriculum specifications*. Edmonton: Alberta Education.

Curtis, M. & Glaser, R. (1983). Reading theory and the assessment of reading achievement. *Journal of Educational Measurement*. 20, 133-147.

Duncan, R. (1986). *Theoretically based test item readability: An approach to estimating the degree to which an item can be understood and answered correctly.* Unpublished doctoral dissertation, University of Texas, Austin.

Duff, A. (1981). *The third language: Recurrent problems of translation into English.* Toronto: Pergamon Press.

Dye, O. (1971). The effects of translation on readability. *Language and Speech*, 14, 392-397.

Favreau, M., Komoda, M., & Segalowitz, N. (1980). Second language reading: Implications of the word superiority effect in skilled bilinguals. *Canadian Journal of Psychology*, 34, 370-380.

Favreau, M. & Segalowitz, N. (1983). Automatic and [illegible] processes in the first and second language reading of fluent bilinguals. *Memory a[illegible]nition*, 11, 565-57[illegible].

Flesch, R. F. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32, 221-233.

Fletcher, C. (1981). Short-term memory processes in text comprehension. *Journal of Verbal Learning and Verbal Behavior*, 20, 564-574.

Freebody, R. & Anderson, R. (1983). Effects on text comprehension of differing proportions and locations of difficult vocabulary. *Journal of Reading Behavior*, 15(3), 10-39.

Frederiksen, J. (1981). Sources of process interactions in reading. In A. M. Lesgold & C. A. Perfetti (Eds.), *Interactive processes in reading* (pp. 361-386). Hillsdale, NJ: Erlbaum.

Friesen, J. (1987). Family and school. An uneasy partnership. In L. Stewin, & S. McCann (Eds.). *Contemporary educational issues - The Canadian mosaic* (pp. 304-312). Toronto: Copp Clark Pitman.

Graham, J. (1985). *Differences in translation.* Ithaca, NY:Cornell University Press.

Horst, P. (1968). *Psychological measurement and prediction.* Belmont, CA: Wadsworth.

Ingram, D. (1985). Assessing proficiency: An overview on some aspects of testing. In K. Hyltenstam & A. Pienemann (Eds.), *Modelling and assessing second language acquisition* (pp. 215276). Clevedon, Avon: Multilingual Matters.

Jackson, J. & McClelland, J. (1981). Exploring the nature of a basic visual-processing component of reading ability. In O. Tzeng, & H. Singer (Eds.), *Perception of print: Reading research in experimental psychology* (pp 29-63). Hillsdale, NJ: Erlbaum.

Johnson, N. (1981). Integration processes in word recognition. O. Tzeng, & H. Singer (Eds.), *Perception of print: Reading research in experimental psychology* (pp 29-63). Hillsdale, NJ: Erlbaum.

Jones, J. (1984). Past, present, and future needs in immersion. *Canadian Modern Language Review*, 41, 260-267.

Jones, L. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.) (pp. 335-355). Washington, D.C.: American Council on Education.

Just, M. & Carpenter, P. (1987). *The psychology of reading and language comprehension.* Newton, MA: Allyn & Bacon.

Katz, J. (1972). *Semantic theory.* New York: Harper & Row.

Klare, G. (1963). *The measurement of readability.* Ames: Iowa State University Press.

Klare, G. (1974-75). Assessing readability. *Reading Research Quarterly*, 10, 62-102.

Klare, G. (1984). Readability. In P.D. Pearson, R. Barr, M. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 681-744). New York: Longman.

Koenke, K. (1987). Readability formulas: Use and misuse. *The Reading Teacher*, 40, 672-674.

Kolers, P. A. (1968). Bilingualism and information processing. *Scientific American*, 218,

78-86.

LaBerge, D. & Samuels, S. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6, 293-323.

Language Services Branch. (1985). *French Immersion in Alberta*. Edmonton: Alberta Education.

Lapkin, S. (1985). Pedagogical implications of direct second language testing: A Canadian example. In K. Hyltenstam & A. Pienemann (Eds.), *Modelling and assessing second language acquisition* (pp. 333-347). Clevedon, Avon: Multilingual Matters.

Lebauer, R. (1985). Nonnative English speakers problems in content and English classes: Are they thinking or reading problems? *Journal of Reading*, 30, 136-142.

Lesgold, A., & Perfetti, C. (1981). Interactive processes in reading: Where do we stand? In A. M. Lesgold, & C. A. Perfetti (Eds.), *Interactive processes in reading* (pp. 387-407). Hillsdale, NJ: Erlbaum.

Levy, B. (1981). Interactive processing during reading. In A. M. Lesgold, & C. A. Perfetti (Eds.), *Interactive processes in reading* (pp. 1-35). Hillsdale, NJ: Erlbaum.

Lipson, M.J. (1982). Learning new information from text. The role of prior knowledge and reading ability. *Journal of Reading Behavior*, 14, 243-261.

Long, M. (1983). Linguistic and conversational adjustments to non-native speakers. *Studies in Second Language Acquisition*, 6, 177-193.

MacNamara, J. (1967). The effect of instruction in a weaker language. *Journal of Social Issues*, 23, 121-135.

MacNamara, J. (1970). Comparative studies of reading and problem solving in two languages. *TESOL Quarterly*, 4, 107-116.

Masson, M. & Miller, J. (1983). Working memory and individual differences in comprehension and memory of text. *Journal of Educational Psychology*, 75, 314-318.

McConnell, C. (1983). Readability: Blind faith in numbers? *Journal of Economic Education*, 14, 65-71.

McEwen, N. (1984). Francophone students in northeastern Alberta. *Canadian Modern*

Language Review, 41, 353-364.

McKeown, M., Beck, I., Omanson, R., & Perfetti, C. (1983). The effects of long-term vocabulary instruction on reading comprehension: A replication. *Journal of Reading Behavior*, 15, 3-17.

Nida, E. (1964). *Toward a science of translating*. Leiden: E.J. Brill.

Nunnally, J. (1967). *Psychometric theory*. Toronto: McGraw Hill.

Oller, J. (1979). *Language tests at school*. London: Longman.

Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.

Pergnier, M. (1978). Language-meaning and message-meaning: Towards a sociolinguistic approach to translation. In D. Gerver & H. W. Sinaiko (Eds.), *Language interpretation and communiation* (pp. 199-204). New York: Plenum.

Rayner, K. (1981). Eye movements and the perceptual span in reading. In F. Pirozzolo, & M. Wittrock (Eds.), *Neuropsychological and cognitive processes in reading*, (pp. 145-165). Toronto: Academic Press.

Reckase, M. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement*, 9, 401-412.

Richards, J. (1976). The role of vocabulary teaching. *TESOL Quarterly*, 10, 77-89.

Sapir, E. (1961). Culture, language and personality. In D.G. Mandlebaum (Ed.), *Selected Essays*. Berkeley: University of California.

Samuels, S. (1987). Information processing abilities and reading. *Journal of Learning Disabilities*, 20, 18-22.

Sax, G. (1974). *Principles of education measurement and evaluation*. Belmont, CA: Wadsworth.

Scoon, A. (1974). *The feasibility of test translation - English to Navajo*. Unpublished doctoral dissertation. Albuquerque: University of New Mexico.

Shoham, M., Peretz, A., & Vorhaus, R. (1987). Reading comprehension tests: General or subject-specific. *System*, 15, 81-88.

Simpson, M. (1987). Alternative formats for evaluating content area vocabulary

understanding. *Journal of Reading*, 31, 20-27.

Spiro, R. (1980). Constructive processes in prose comprehension and recall. In R. J. Spiro, B. C. Bruce, & W. Brewer, (Eds.), *Theoretical issues in reading comprehension* (pp. 245-278). Hillsdale, NJ: Erlbaum.

Stahl, S. (1983). Differential word knowledge and reading comprehension. *Journal of Reading Behavior*, 15(4), 33-50.

Stanovich, K. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, 16, 32-71.

Stanovich K. (1981). Attentional and automatic context effects in reading. In A. M. Lesgold & C. A. Perfetti (Eds.), *Interactive processes in reading* (pp. 241-267). Hillsdale, NJ: Erlbaum.

Stanovich, K. & West, R. F. (1979). Mechanisms of sentence context effects in reading: Automatic activation and conscious attention. *Memory & Cognition*, 7, 77-85.

Student Evaluation Branch (1984). *Student achievement testing program bulletin: Grade 6 social studies*. Edmonton: Alberta Education.

Student Evaluation Branch (1985a). *Grade 6 social studies achievement test part A: Multiple choice*. Edmonton: Alberta Education.

Student Evaluation Branch (1985b). *Test de rendement etudes sociales 6e année partie A: Choix multiples*. Edmonton: Alberta Education.

Student Evaluation Branch (1985c). *Provincial report: Grade 6 social studies achievement test*. Edmonton: Alberta Education.

Student Evaluation Branch (1987). *Grade 9 social studies achievement test part A: Multiple choice*. Edmonton: Alberta Education.

Student Evaluation & Records Branch (in press). *Provincial assessment of students in French Immersion programs*. Edmonton: Alberta Education.

Swain, M. (1974). Child bilingual language learning and linguistic interdependence. In S. Carey (Ed.), *Bilingualism, biculturalism and education* (pp. 75-81). Proceedings from the Conference at College Universitaire Saint-Jean, University of Alberta.

Swain, M. (1978). School reform through bilingual education: problems and some solutions in evaluating programs. *Comparative Education Review*, 22, 420-433.

Swain, M., & Lapkin, S. (1981). *Bilingual education in Ontario: A decade of research.* Toronto: Ministry of Education.

Thorndike, R., & Hagen, E. (1977). *Measurement and evaluation in psychology* (4th ed.). Toronto: Wiley & Sons.

Tinker, M.A. (1963). *Legibility of Print.* Ames: Iowa State University.

Tighe, V., & Hadaway, N. (1986). Old issues and helpful hints: Content area reading for second language learners. *Reading Improvement*, 23, 293-298.

Underwood, G. (1985). Information processing in skilled readers. In G. MacKinnon, & T. Waller (Eds.), *Reading research: Advances in theory and practice* (Vol. 4) (pp. 139-181). Toronto: Academic Press.

Valdes, G., Barrera, R., & Cardenas, M. (1984). Constructing matching text in two languages: The application of propositional analysis. *Journal of the National Association for Bilingual Education*, 9, 3-19.

Vorhaus, R. (1984). Strategies for reading in a second language. *Journal of Reading*, 27, 412-416.

Webster's Ninth New Collegiate Dictionary. (1984). Markham Ontario: Thomas Allen.

West, A. (1985). Post secondary bilingual education at the University of Calgary. *Canadian Modern Language Review*, 41, 246-259.

Willig. A. (1985). A meta-analysis of selected studies on the effectiveness of bilingual education. *Review of Educational Research*, 55, 269-317.

Woytak, L. (1984). Reading proficiency and a psycholinguistic approach to second language reading. *Foreign Language Annals*, 17, 509-517.

Yorio, C. A. (1971). Some sources of reading problems in foreign language learners. *Language Learning*, 21, 107-115.

APPENDIX

Distribution of Questions by Alberta Education Reporting Category

Category	Questions	Number
Topic A	1 - 17	17
Topic B	18 - 33	16
Topic C	34 - 50	17
Recall & Comp. A	2,3,4,5,6,7,12	7
Recall & Comp. B	19,20,21,22,23,24,29	7
Recall & Comp. C	42,43,44,46,47,48,49,50	8
Values	10,11,26,27,34,45	6
Inquiry I	1,8,9,18,25,35,37,38	8
Inquiry II	13,14,28,30,31,32,36,39	8
Inquiry III	15,16,17,33,40,41	6

Distribution of Questions by Reconstructed Reporting Category

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

Grade 6 Social Studies

ACHIEVEMENT TEST

Part A: Multiple Choice

June 1985

Alberta
EDUCATION

GRADE 6 SOCIAL STUDIES ACHIEVEMENT TEST

PART A: MULTIPLE CHOICE

GENERAL INSTRUCTIONS

1. Please be sure that you have put your name and other information on your answer sheet before you begin this part of the test.

2. This test has 50 questions. You will not be finished until you see the STOP sign on page 28. You have 50 minutes to complete this part of the test.

3. Read carefully. Choose the CORRECT or BEST answer for each question.

4. Mark your answer on the answer sheet with an HB pencil only. Be sure that the number on the answer sheet is the same as the question number in the test booklet.

Example:

Test Booklet

8. Edmonton is the capital city of

A. British Columbia
B. Alberta
C. Saskatchewan
D. Manitoba

Answer Sheet

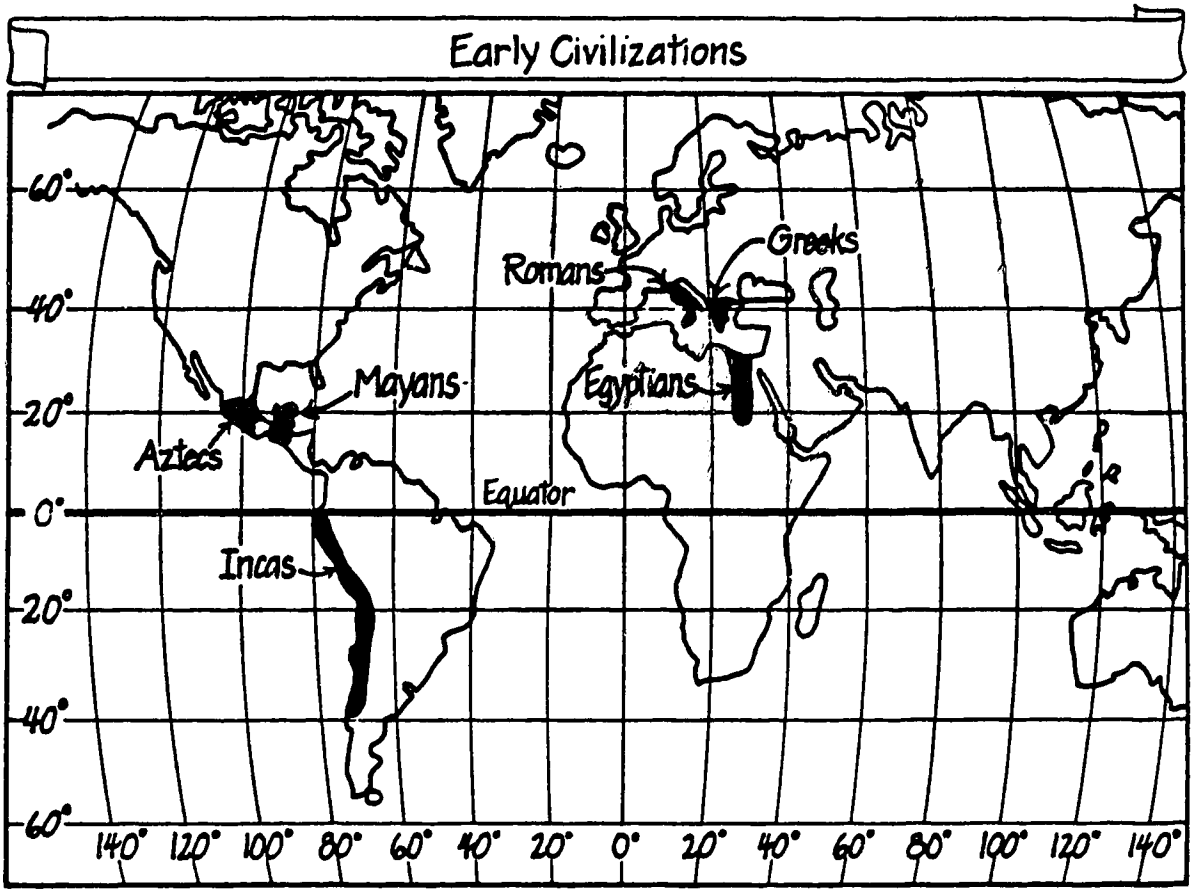
	A	B	C	D
8.	○	●	○	○

5. Fill in the entire circle for each answer that you choose. If you make a mistake, erase your first mark COMPLETELY and fill in the correct circle.

6. Try to answer every question.

DO NOT WRITE ON THE TEST BOOKLET

THIS PART OF THE TEST IS ABOUT HOW PEOPLE IN EARLIER TIMES MET THEIR NEEDS.



Questions 1 to 4 refer to the civilizations shown on the map on page 2.

1. Archeologists found artifacts at latitude 39° north and longitude 23° east. To which early civilization did the artifacts belong?

   A. The Aztecs
   B. The Greeks
   C. The Mayans
   D. The Egyptians

2. Climate would have had the GREATEST effect in determining how the people of an area met their need for

   A. shelter
   B. social order
   C. transportation
   D. communication

3. Most early civilizations developed in areas that had

   A. dense forests
   B. sandy deserts
   C. mild climates
   D. rugged mountains

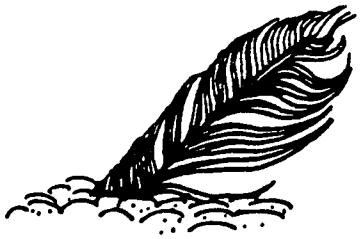
4. When people build shelters, the basic need they are trying to meet is

   A. social
   B. cultural
   C. physical
   D. psychological

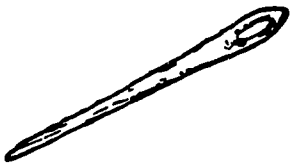
5. Which object found by archeologists would teach us the most about how people met basic needs?

A. A piece of volcanic rock

B. A rib from a buffalo



C. An eagle feather buried in sand



D. A needle made from a bone

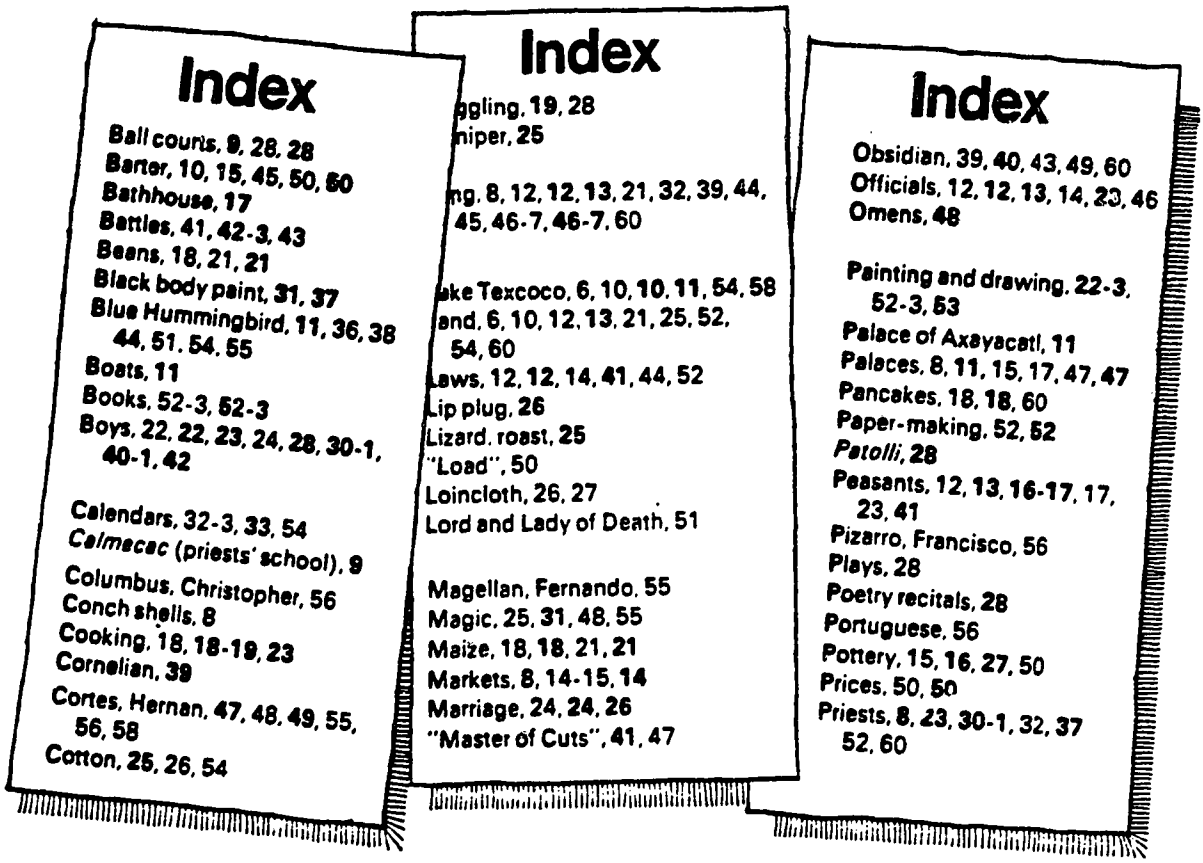
6. To live together successfully, people in early civilizations had to develop

A. laws
B. towns
C. armies
D. temples

7. Many early civilizations experienced rapid, major changes to their cultures as a result of

A. being conquered
B. building bigger shelters
C. keeping historical records
D. practising religious beliefs

Question 8 is based on the following index pages taken from a book about Aztecs.



Index

Ball courts, 9, 28, 28
Barter, 10, 15, 45, 50, 50
Bathhouse, 17
Battles, 41, 42-3, 43
Beans, 18, 21, 21
Black body paint, 31, 37
Blue Hummingbird, 11, 36, 38 44, 51, 54, 55
Boats, 11
Books, 52-3, 52-3
Boys, 22, 22, 23, 24, 28, 30-1, 40-1, 42

Calendars, 32-3, 33, 54
Calmecac (priests' school), 9
Columbus, Christopher, 56
Conch shells, 8
Cooking, 18, 18-19, 23
Cornelian, 39
Cortes, Hernan, 47, 48, 49, 55, 56, 58
Cotton, 25, 26, 54

Index

ggling, 19, 28
niper, 25

ng, 8, 12, 12, 13, 21, 32, 39, 44, 45, 46-7, 46-7, 60

ake Texcoco, 6, 10, 10, 11, 54, 58
and, 6, 10, 12, 13, 21, 25, 52, 54, 60
aws, 12, 12, 14, 41, 44, 52
Lip plug, 26
Lizard, roast, 25
"Load", 50
Loincloth, 26, 27
Lord and Lady of Death, 51

Magellan, Fernando, 55
Magic, 25, 31, 48, 55
Maize, 18, 18, 21, 21
Markets, 8, 14-15, 14
Marriage, 24, 24, 26
"Master of Cuts", 41, 47

Index

Obsidian, 39, 40, 43, 49, 60
Officials, 12, 12, 13, 14, 23, 46
Omens, 48

Painting and drawing, 22-3, 52-3, 53
Palace of Axayacatl, 11
Palaces, 8, 11, 15, 17, 47, 47
Pancakes, 18, 18, 60
Paper-making, 52, 52
Patolli, 28
Peasants, 12, 13, 16-17, 17, 23, 41
Pizarro, Francisco, 56
Plays, 28
Poetry recitals, 28
Portuguese, 56
Pottery, 15, 16, 27, 50
Prices, 50, 50
Priests, 8, 23, 30-1, 32, 37 52, 60

8. Suppose you want to know if the Aztecs met their need for food by trading with other communities. Which heading in the index would most likely lead you to the information you need?

   A. Beans
   B. Cooking
   C. Markets
   D. Prices

Use the information below to answer questions 9 to 11.

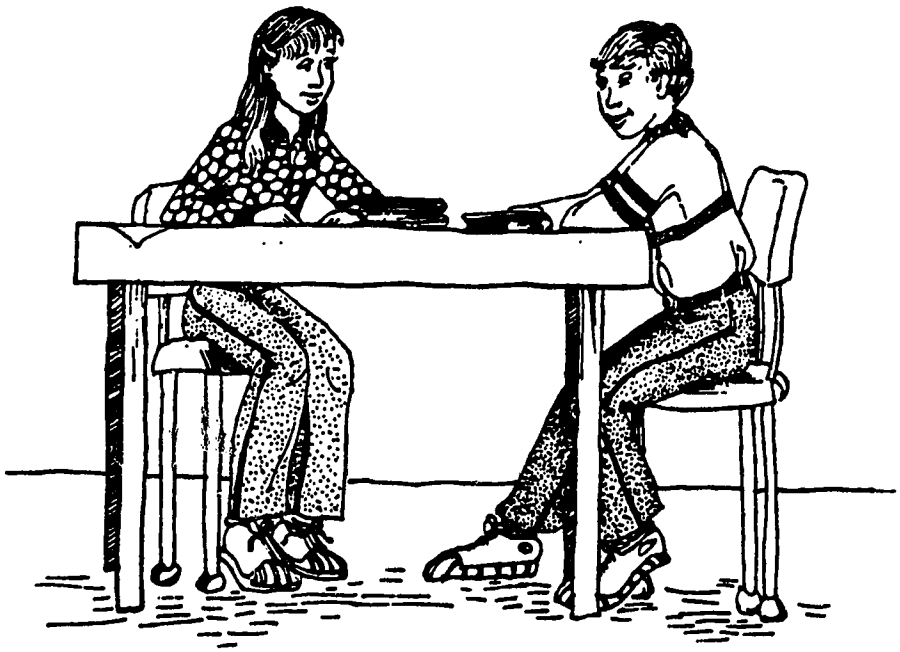
Cathy and Blair had a debate at school. Cathy chose to talk about the Mayan civilization. Blair spoke about his society.

Cathy said:

I like the Mayan way of making decisions. Although their system was harsh, at least the Mayans knew what the rules were. The *Halach Uinic* (chief ruler) and high priests lived in the palaces. Noblemen were expected to live near the city centre, while farmers and peasants lived in small huts at the edge of the city.

Blair said:

Yes, but we have more freedom in our society. Some people don't think it's important to spend most of their money on housing. Others want large homes so they work hard to earn extra money. These people can afford to rent or own bigger homes. I think our system is fair.



9. Cathy and Blair have different ideas about how people in a society should make a decision about

   A. how much to pay for housing
   B. how much to charge for housing
   C. who gets the best kind of housing
   D. who gets a house of his own

10. According to Cathy, an advantage of the Mayan system of housing was that people

   A. did not have to go into debt
   B. knew what was expected of them
   C. got housing without having to work
   D. worked hard to get what they wanted

11. Blair believes it is important for people to

   A. be prepared to put in more effort
   B. buy what they can afford
   C. be able to make choices
   D. rent the best houses

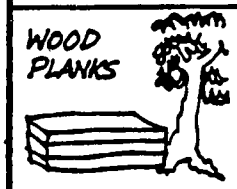
---

12. In MOST early civilizations, wealth and power were

   A. held mainly by the merchants and traders
   B. held mainly by the nobles and priests
   C. shared equally by the warriors
   D. shared equally by all citizens

Use the chart below to answer questions 13 to 15.

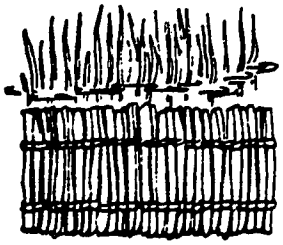
Building Materials Used by Early Civilizations

MATERIAL	QUALITY OF MATERIAL	AVAILABILITY OF MATERIAL	ABILITY TO LAST	SKILL TO MAKE AND USE MATERIAL	LABOR FORCE NEEDED
REEDS JOINED TOGETHER WITH MUD 	poor quality	easy to find	must be repaired often	easy to use	one or two workers
CLAY BRICKS 	good quality	easy to find	will last many years	some skill needed	one or many workers
WOOD PLANKS 	good quality	only found in certain places	will last many years	many skills needed	many workers
ROCK CUT FROM QUARRIES 	excellent quality	only found in certain places	will outlast civilization	great skill needed	great number of workers

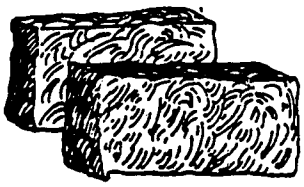
13. A disadvantage of using reeds as a building material was related to

    A. their availability
    B. their ability to last
    C. the skill needed to use them
    D. the labor force needed to build with them

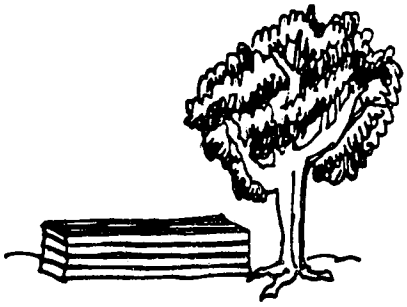
14. The work of laborers had to be MOST organized when they were building with



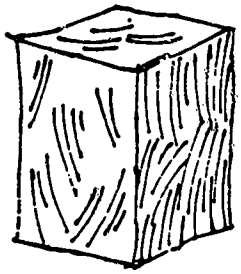
A. reeds



B. clay bricks



C. wood planks



D. rock

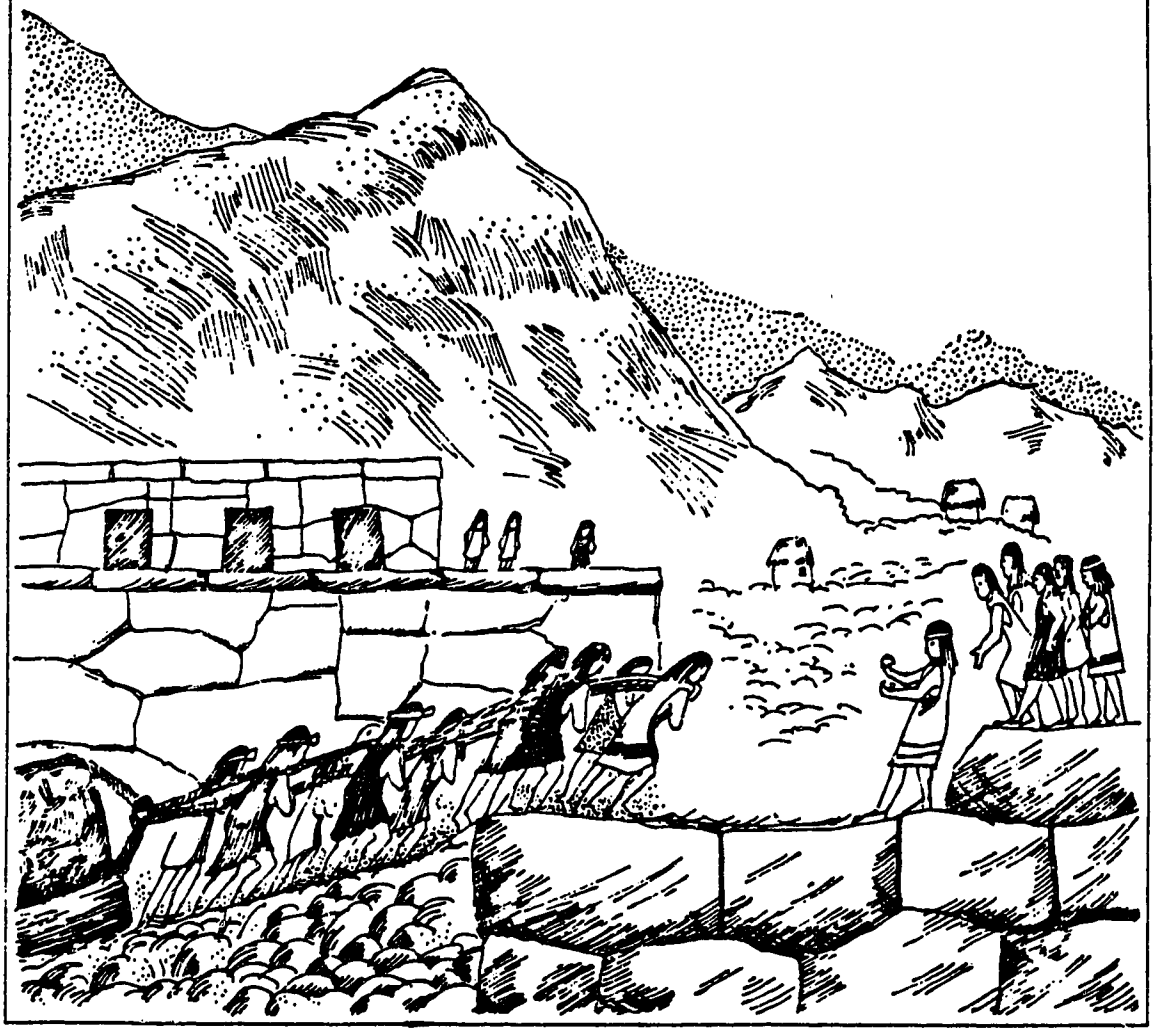
15. "Only very wealthy people in early civilizations built homes made from rock."

What information from the chart gives the BEST support for the statement above?

A. Building with rock takes a lot of highly skilled labor.
B. Small homes are not easily built with rock.
C. Houses built from rock last a long time.
D. Rock is only found in quarries.

Use the information below to answer questions 16 and 17.

Important buildings such as the palace of the *Inca* (ruler) and the temples were constructed by peasants paying their *mita*. The *mita* was a tax that was paid by working for the *Inca*. Once a year, each male peasant spent two or three weeks working on building projects. The workers were fed, but not paid, while working for the *Inca*.



16. How would a public arena be built in Alberta today if the *mita* system were followed exactly?

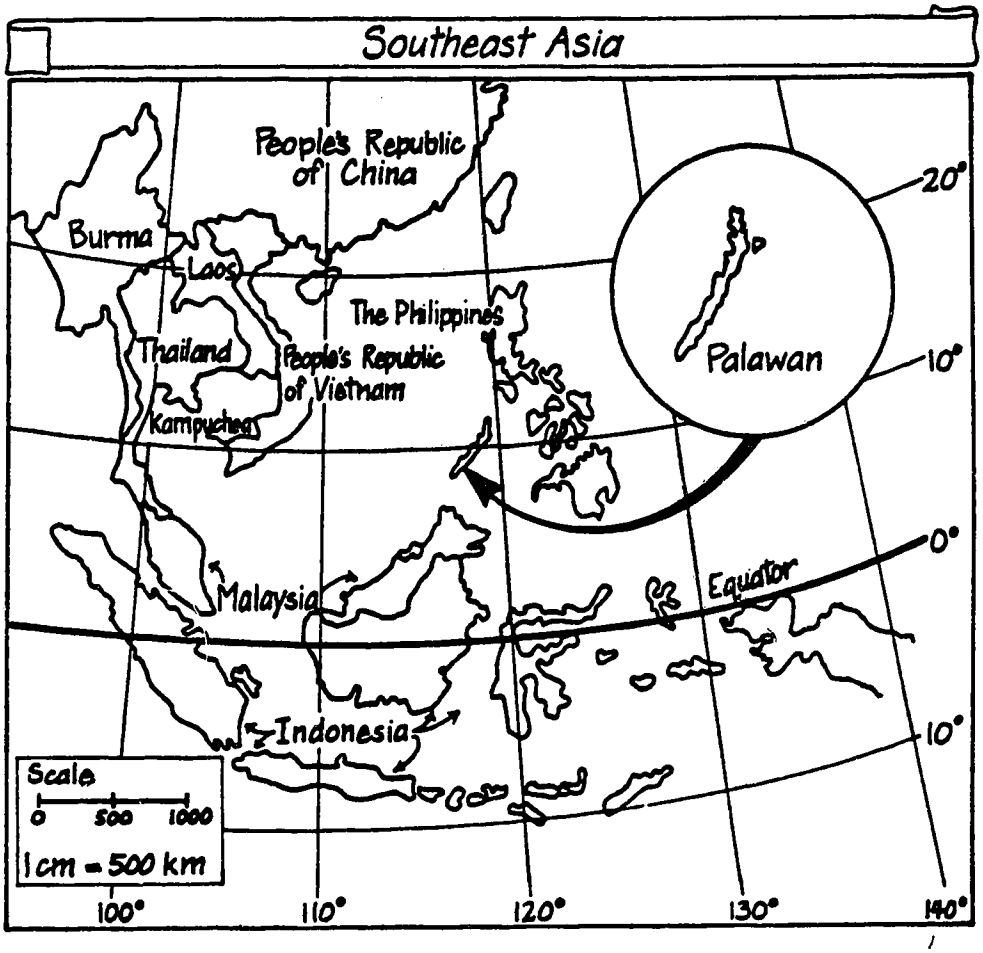
A. Male laborers would build it without pay.
B. Foreign laborers would be paid to build it.
C. A private company would be paid to build it.
D. Government employees would build it without pay.

17. Why might a person who valued equal treatment for all consider the *mita* system to be unjust?

A. Projects would take longer to complete.
B. Taxes would no longer have to be paid.
C. Some workers would be paid more than others.
D. Some people would have to work but not others.

THIS PART OF THE TEST IS ABOUT HOW PEOPLE IN EASTERN SOCIETIES MEET THEIR NEEDS TODAY.

Use the map below to answer question 18.



18. What information on the map supports the claim that the temperature in Palawan varies little between winter and summer?

A. Palawan is a large island.
B. Palawan is near the equator.
C. Palawan has mountain ridges.
D. Palawan has a monsoon season.

19. Which of the following basic needs is LEAST affected by the climate of a country?

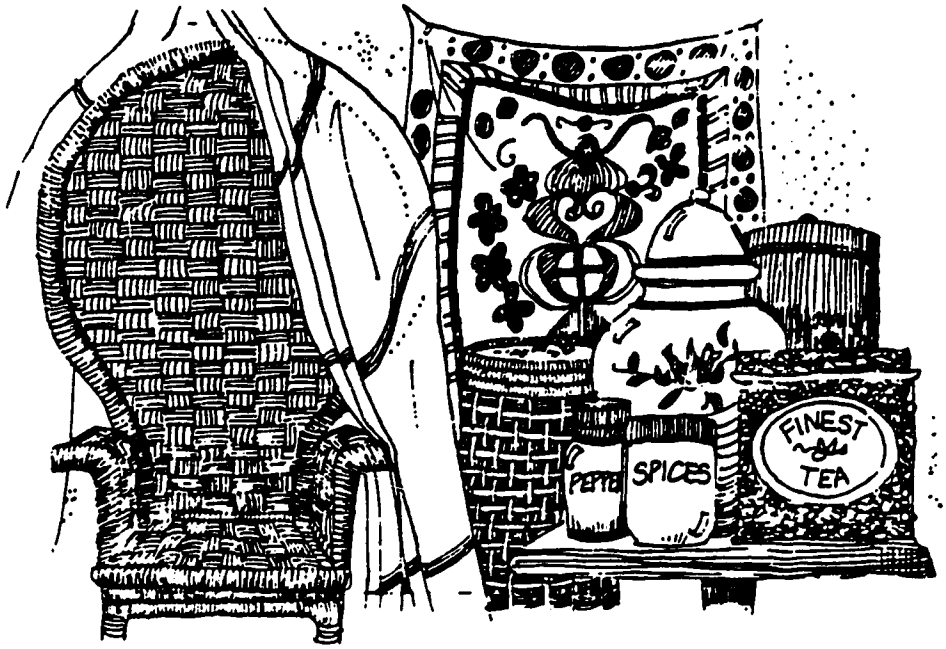
A. Food
B. Housing
C. Clothing
D. Education

20. Which statement BEST describes how most people react to changes within a society?

A. People change their traditions easily.
B. People do not change their traditions.
C. People in all societies try to keep their traditions from changing.
D. People only change their traditions when they get ideas from other societies.

21. Which statement shows that tradition affects the way people meet their need for food?

A. Some people do not eat meat because of their religious beliefs.
B. Some people eat fruit often because it contains vitamins.
C. People make bread from many different kinds of flour.
D. People do not grow rice in dry climates.



22. A MAJOR reason for importing goods from Southeast Asia is to

A. learn about other ways of life
B. save our own goods for future use
C. give aid to Southeast Asian people
D. have things not grown or made in Canada

23. The aim of organizations such as the Canadian International Development Agency (CIDA) is to improve living conditions in Southeast Asian countries by

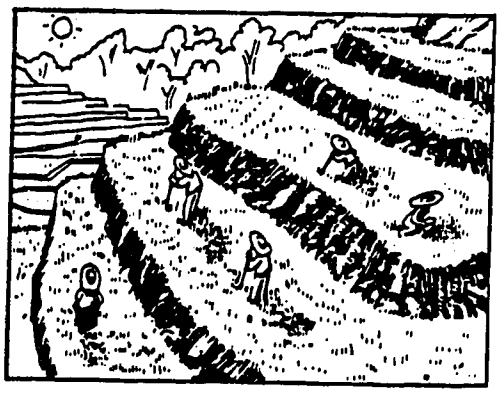
A. supporting local self-help programs
B. selling wheat to the people at low prices
C. collecting used clothing to ship overseas
D. encouraging Southeast Asians to move to Canada

24. Some Southeast Asian countries have many people to feed but very little farmland for growing crops. One way they are trying to solve this problem is by

A. sending people to work in the fields

B. dividing up the land so every family has some

C. moving people so they can farm in other countries



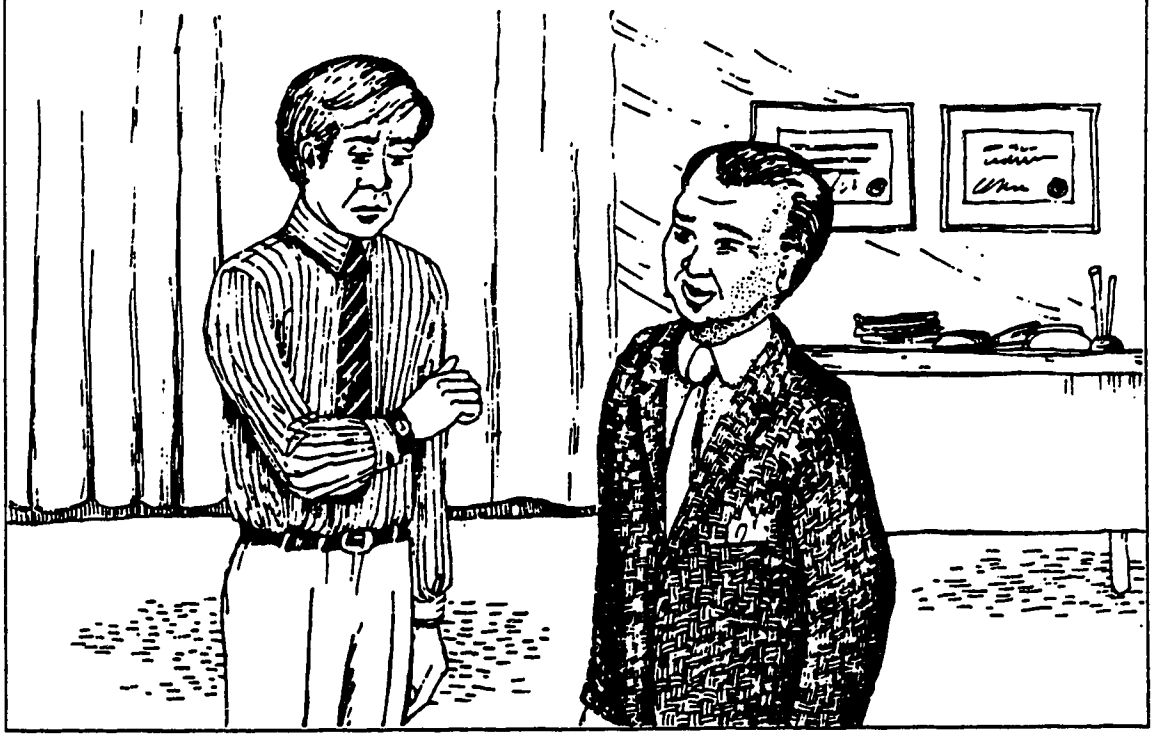
D. increasing the amount of farmland by terracing hillsides

Use the information below to answer questions 25 to 28.

Mr. Duncan, the manager of a Canadian engineering company, is working in a Southeast Asian country. He is surprised when Mr. Ochoa, a local businessman, arrives 30 minutes late for an appointment. The following comments are part of a conversation they had.

Mr. Duncan: "You are late. I expected you half an hour ago. We have to follow schedules and meet tight deadlines if we are going to do this job properly. In the engineering business, time is money. My staff is well-trained and that means well-paid. We can't afford to have them waiting for meetings."

Mr. Ochoa: "You have misunderstood our values. I never intended to come at the appointed time; my countrymen would not have expected me until later. To arrive on the dot suggests that I am not very important, that I have nothing better to do than keep this one appointment. A busy and successful man has so many responsibilities that he has the right to be late."



25. What issue about working in a country that has a different culture is being discussed by Mr. Duncan and Mr. Ochoa?

A. Who should work in other countries?
B. How should people keep business appointments?
C. Who should be allowed to attend meetings?
D. How should workers be paid?

26. The conflict is between one person's desire to be treated with dignity and the other person's need to

A. do a job on time
B. be well-respected
C. use modern technology
D. hire well-trained workers

27. In running his business, Mr. Duncan places the MOST importance on

A. following other people's traditions
B. doing work in other countries
C. using well-trained workers
D. being very efficient

28. The misunderstanding that has occurred between the two men results from the

A. languages they speak
B. customs they follow
C. technology they use
D. work they do

---

29. A major reason that Canadians are hired to work on some Southeast Asian building projects is that

A. Canadians want to learn about Southeast Asian ways
B. Canadians need help with their unemployment problems
C. specially-trained engineers are needed in Southeast Asia
D. it is too expensive to hire Southeast Asian engineers

Use the comments below to answer questions 30 and 31.

I use my telephone a lot. I travel a great deal in my job, so I have friends all over the country. I can't visit often, so I need to keep in touch by phone. I often use the phone for my work, too, because I must speak with computer programmers all over Canada.

— Christie Drapeau,
Canadian consultant

I don't need to use a telephone. After selling my catch of fish each morning, I go to the coffee shop for breakfast. I meet my friends there, and we discuss news from the village. Sometimes truck drivers stop in on their way to Kuala Lumpur and bring news from other villages.

— Abraham B. Hasan,
Malaysian fisherman

30. Christie and Abraham have different ideas about the need to use telephones because Christie has

A. more friends
B. a different lifestyle
C. different religious beliefs
D. a greater need to talk with friends

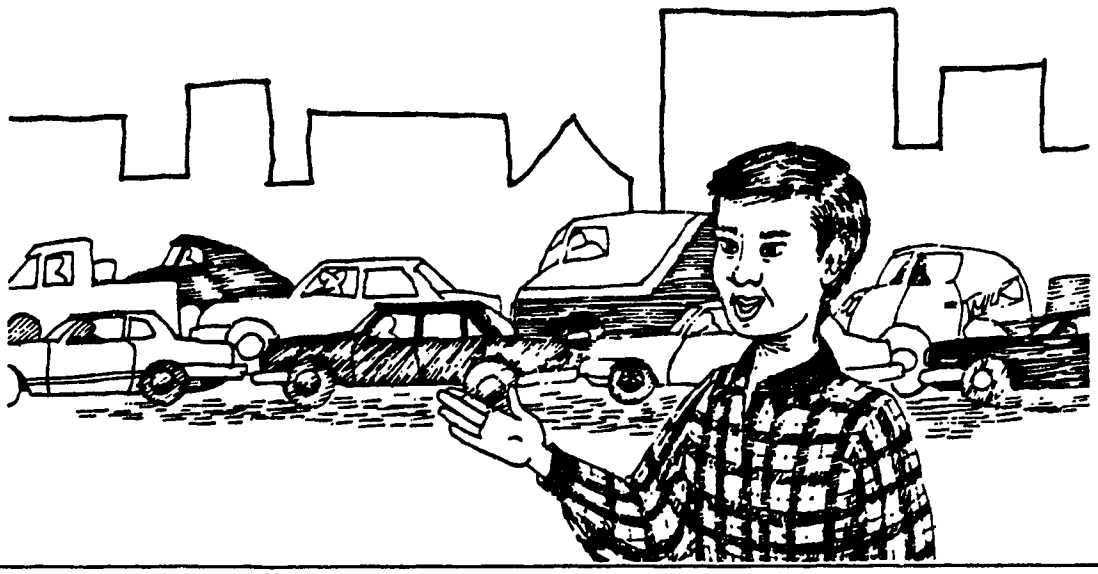
31. A general statement about all societies that can be supported by the two comments is that people

A. like to communicate with friends
B. have to use telephones to do business
C. need to talk to their friends every day
D. want to have friends in far-away communities

Use the information below to answer questions 32 and 33.

Cars are expensive to own and operate, yet every Canadian feels the need to own one. This is a terrible waste. It is common to see one person driving to work in a huge car that could carry six people. It's no wonder Canadians have trouble with traffic jams in large cities.

— Somchai Meesiri, a student from Thailand



32. Somchai's views about car ownership differ from those of many Canadians. Which of the following statements BEST explains this difference?

A. Canadians can get a driver's licence more easily than can Thais.
B. Canadians can afford to buy cars more easily than can Thais.
C. The climate in Canada is colder than in Thailand.
D. The roads are better in Canada than in Thailand.

33. A solution to traffic jams in Canada that would encourage more CO-OPERATION would be to

A. walk to work
B. buy cheaper cars
C. travel by car pool
D. drive smaller cars

THIS PART OF THE TEST IS ABOUT MEETING NEEDS THROUGH LOCAL, PROVINCIAL, AND FEDERAL GOVERNMENTS.

Read the information below, then answer questions 34 to 36.

The law says that people riding motorcycles must wear helmets. Some people do not like this law and want to see it repealed (removed). Other people support the law.

These are some comments that citizens have made on this issue.

Use the speakers' opinions on page 20 to answer questions 34 to 36.

34. With which speaker does Steve DISAGREE MOST?

A. Ms. Maguire

B. Mr. Gibeau

C. Mr. Brandon

D. Miss Seldon

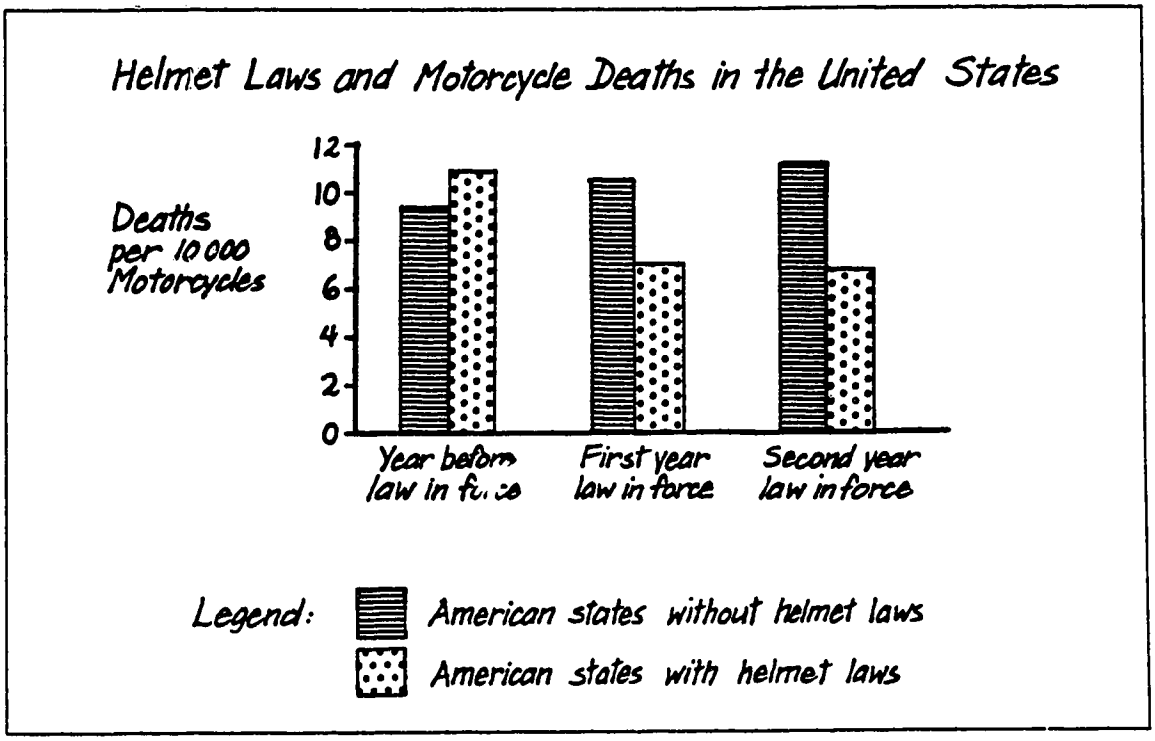
35. What is the MAIN issue being discussed by the citizens?

A. Should taxes be used to pay for motorcycle injuries?
B. Should there be a law requiring people to wear motorcycle helmets?
C. Should riders who wear helmets be allowed to pay less for medicare?
D. Should the law requiring helmets be limited to children who are passengers?

36. If all of the speakers' comments were being put on a chart, what would be the best title?

A. Opinions About the Helmet Law
B. Effects of Having the Helmet Law
C. Reasons for Keeping the Helmet Law
D. Persons Who Voted for the Helmet Law

Use the graph below to answer questions 37 to 39.



37. Which of the following questions could be answered using the graph?

A. When do most motorcycle accidents happen?
B. Why does the wearing of helmets save lives?
C. How many Canadians die in motorcycle accidents?
D. What might happen if our helmet laws were removed?

38. The graph shows that for every 10 000 motorcycles, American states WITHOUT helmet laws have

A. more motorcycle deaths than do states with helmet laws
B. more motorcycle accidents than do states with helmet laws
C. fewer people wearing helmets than do states with helmet laws
D. fewer people riding motorcycles than do states with helmet laws

39. What is the MAIN point presented in the graph?

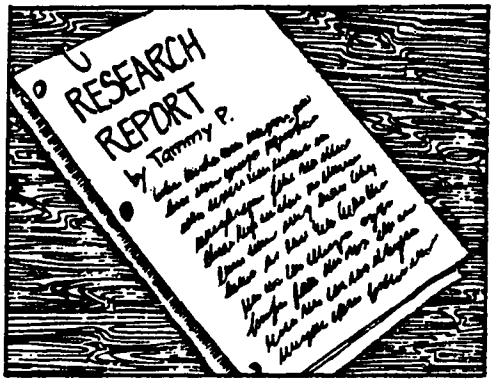
A. Many states removed or weakened their helmet laws.
B. Many states have laws requiring the wearing of helmets.
C. Fewer motorcycle deaths occur when people wear helmets.
D. Fewer motorcycle deaths occurred in the second year.

40. After learning that the helmet law could be removed, a Grade 6 class decided to take action to let people know how they felt about this issue.

Which action would make the students' point of view known to the greatest number of people?

A. Survey students in other schools.

B. Speak to the principal at recess.

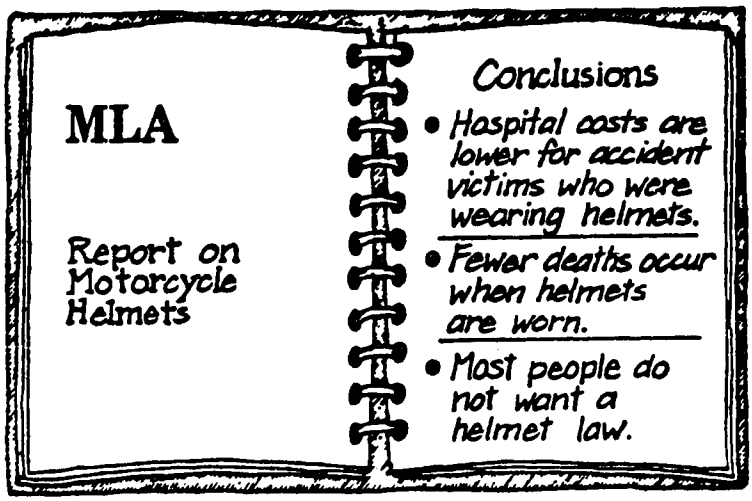


C. Write a research report.

D. Make a presentation to parents.

Use the information below to answer question 41.

A number of MLA's presented the following report to the legislature.



After the report was discussed, a vote was taken. The helmet law was repealed (removed).

41. Based on the principle of majority rule, the removal of the helmet law was

A. democratic, because most citizens didn't want the law
B. democratic, because some MLA's did research on the law
C. undemocratic, because hospital costs would be higher for all
D. undemocratic, because more Albertans would now die

---

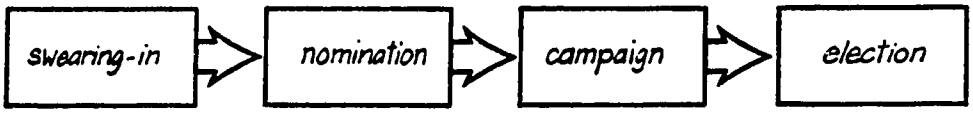
42. One of the reasons that Canada is considered to be a democracy is that we have

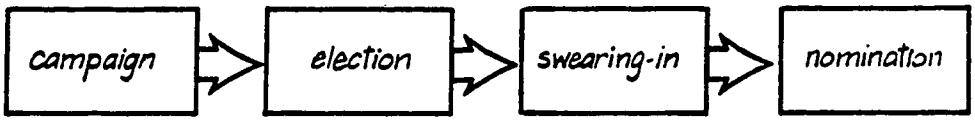
A. premiers in every province
B. leaders who are elected
C. a federal government
D. a constitution

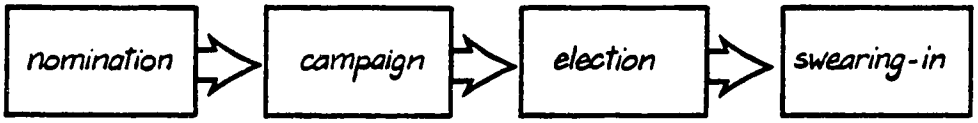
43. The discussion held by members of the legislative assembly on a proposed law is called a

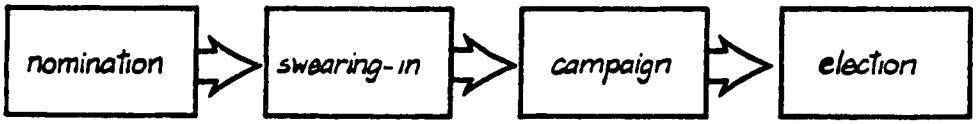
A. lobby
B. survey
C. debate
D. campaign

44. The steps, in CORRECT ORDER, that a person must go through to become a member of the legislative assembly are

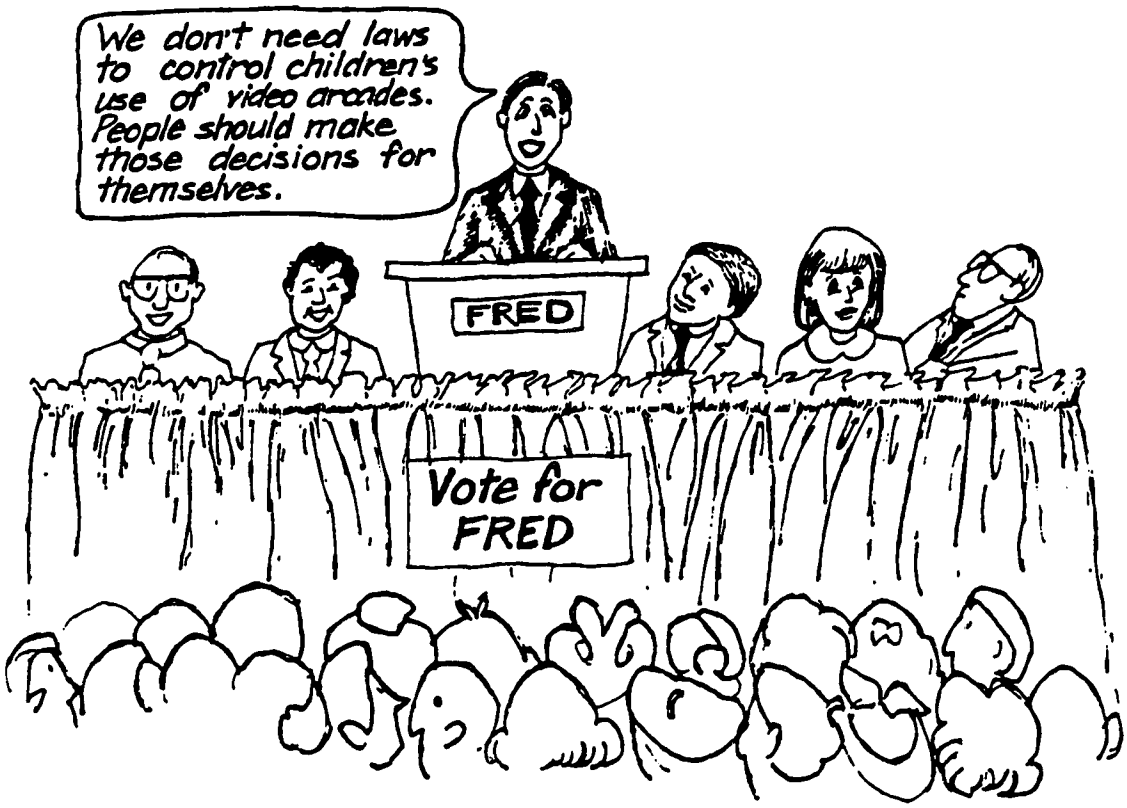
A. 

B. 

C. 

D. 

Use the information below to answer question 45.



45. This candidate would MOST LIKELY be supported by a voter who believed that

A. children shouldn't have to obey laws
B. children shouldn't have a say in making laws
C. people should work together to meet needs
D. people should look after their own needs

---

46. Normally, the leader of the party that has the greatest number of elected members in the House of Commons is called the

A. Leader of the Opposition
B. Speaker of the House
C. Governor General
D. Prime Minister

47. One purpose of political parties in Canada is to

A. bring together people who share similar views on government
B. collect taxes to pay the salaries of elected officials
C. appoint members to town or city councils
D. choose members to sit in the Senate

48. The main role of the opposition in the federal government is to

A. research bills for the government
B. protect the rights of provinces
C. present other points of view
D. support the party-in-power

49. Governments in Canada do not usually pay for the building of

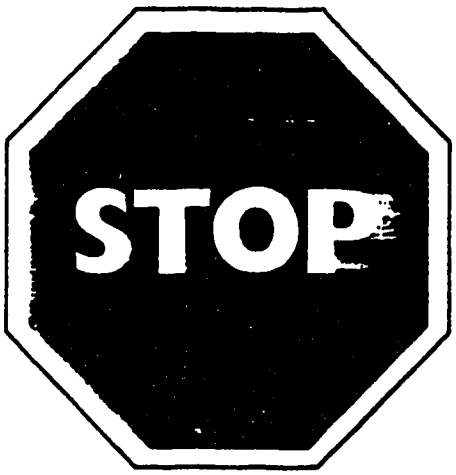
A. schools
B. churches
C. hospitals
D. fire stations

50. Provincial governments can pass laws regarding

A. national defence
B. postal services
C. citizenship
D. education

THIS IS THE END OF THE TEST.

You may go back and check your answers.



Thank you for writing the Grade 6 Social Studies Achievement Test.

CREDITS

Sources 25-29 "Businessmen from the States. . . he is entitled to be late." Adapted and abridged from *THE LAND AND PEOPLE OF THE PHILIPPINES* by John Nance (J. B. Lippincott Company).

TEST DE RENDEMENT

Etudes sociales 6e année

Partie A: Choix multiples

Juin 1985

Alberta
EDUCATION

TEST DE RENDEMENT D'ETUDES SOCIALES - 6° ANNEE

PARTIE A: CHOIX MULTIPLES

INSTRUCTIONS GENERALES

1. Assure-toi que tu as mis ton nom et tout autre renseignement sur la feuille de réponses avant de commencer ce test.

2. Ce test comprend 50 questions. Tu n'auras pas fini avant de voir le signe ARRÊT à la page 28. Tu as 50 minutes pour faire ce test.

3. Lis attentivement. Choisis la BONNE ou la MEILLEURE réponse pour chaque question.

4. Indique ta réponse sur la feuille de réponses seulement avec un crayon HB. Assure-toi que le numéro sur la feuille de réponses est le même que celui de la question dans le livret de questions.

Exemple:

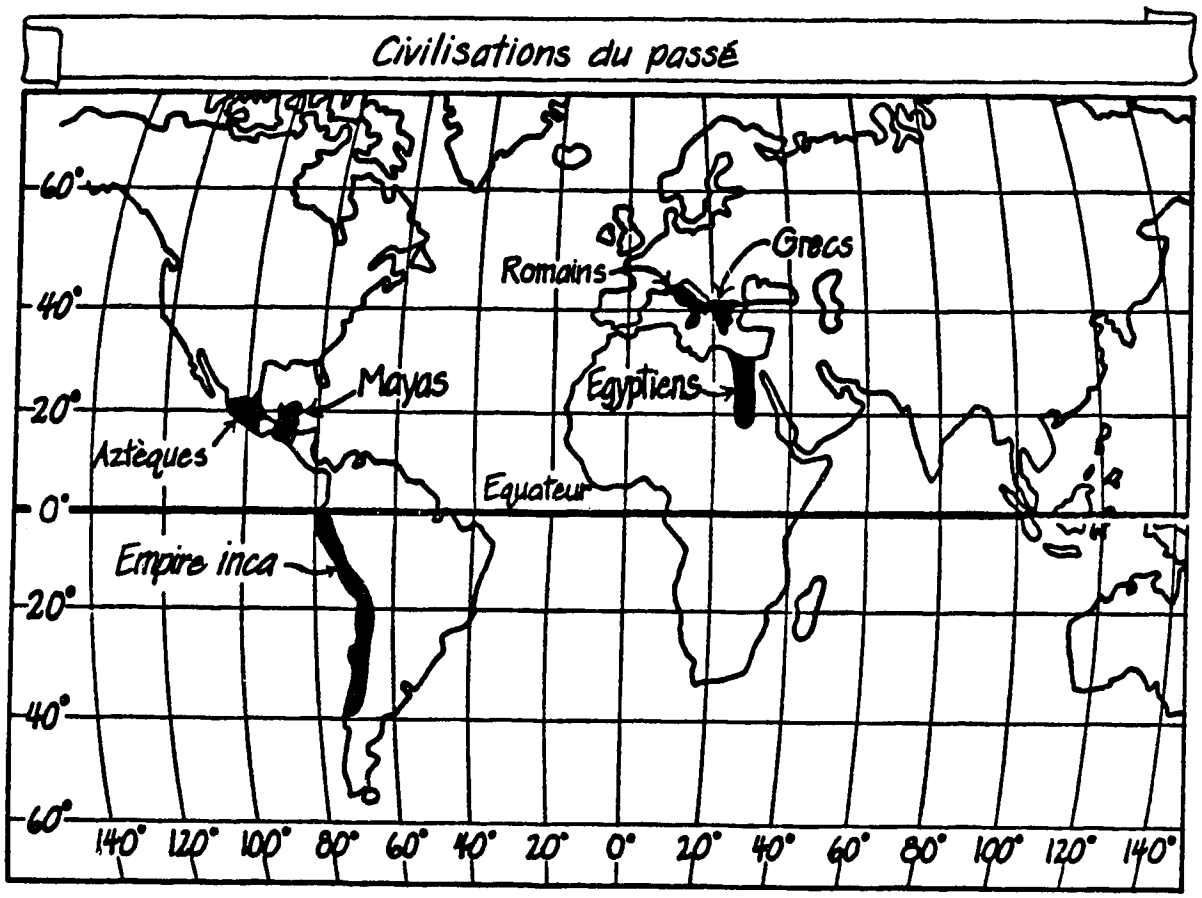
Livret de questions	Feuille de réponses				
		A	B	C	D
8. Edmonton est la capitale A. de la Colombie-Britannique B. de l'Alberta C. de la Saskatchewan D. du Manitoba	8.	O	●	O	O

5. Emplis le cercle en entier pour chaque réponse que tu choisis. Si tu fais une erreur, efface la première marque COMPLETEMENT et emplis en entier le bon cercle.

6. Essaie de répondre à chaque question.

N'ECRIS PAS SUR LE LIVRET DE QUESTIONS

CETTE PARTIE DU TEST PORTE SUR LA FAÇON DONT L'HOMME SATISFAISAIT SES BESOINS A DES EPOQUES RECULEES



Les questions 1 à 4 se rapportent aux civilisations indiquées sur la carte de la page 2.

1. Les archéologues ont trouvé des objets fabriqués à 39° de latitude nord et 23° de longitude est. A quelle civilisation du passé ces objets appartenaient-ils?

   A. Les Aztèques
   B. Les Grecs
   C. Les Mayas
   D. Les Egyptiens

2. Le climat aurait eu l'effet LE PLUS GRAND quand il s'agissait de déterminer la façon dont les habitants d'une région répondaient à leurs besoins

   A. d'abri
   B. d'ordre social
   C. de transport
   D. de communication

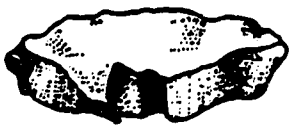
3. La plupart des civilisations du passé se sont développées dans des régions qui avaient

   A. des forêts denses
   B. des déserts de sable
   C. des climats doux
   D. des montagnes rocailleuses

4. Quand on construit un abri, le besoin élémentaire que l'on essaie de satisfaire est

   A. social
   B. culturel
   C. physique
   D. psychologique

5. Quel objet trouvé par les archéologues nous apprendrait le plus de choses sur la façon dont les habitants répondaient à leurs besoins essentiels?



A. Un morceau de roche volcanique

B. Une côte de bison

C. Une plume d'aigle enterrée dans le sable

D. Une aiguille en os

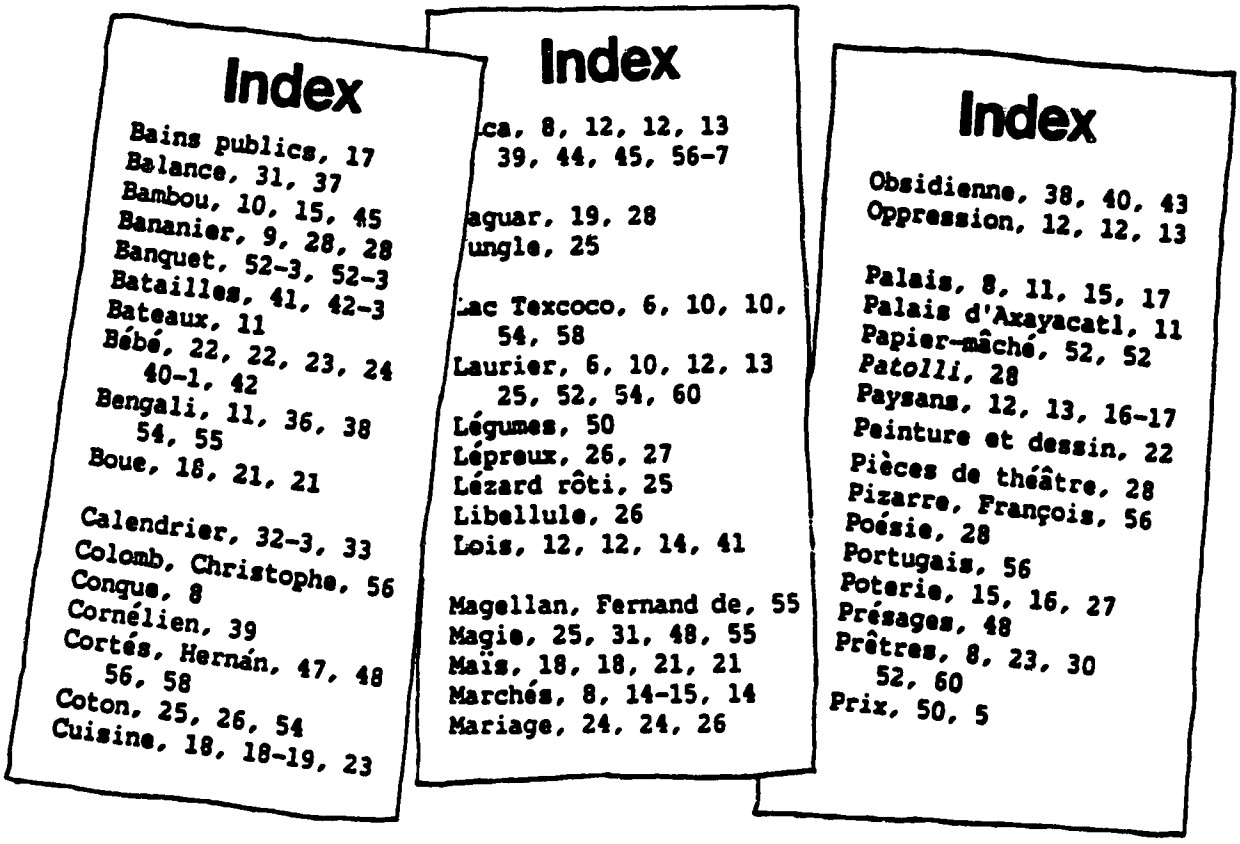
6. Pour vivre harmonieusement ensemble, les gens des civilisations du passé ont dû se faire des

A. lois
B. villes
C. armées
D. temples

7. Beaucoup de civilisations du passé ont subi des changements majeurs et rapides dans leurs cultures à cause de

A. conquêtes
B. la construction d'abris plus grands
C. la rédaction de documents historiques
D. la pratique de croyances religieuses

La question 8 est basée sur les pages d'index suivantes tirées d'un livre sur les Aztèques.



Index

Bains publics, 17
Balance, 31, 37
Bambou, 10, 15, 45
Bananier, 9, 28, 28
Banquet, 52-3, 52-3
Batailles, 41, 42-3
Bateaux, 11
Bébé, 22, 22, 23, 24
40-1, 42
Bengali, 11, 36, 38
54, 55
Boue, 18, 21, 21

Calendrier, 32-3, 33
Colomb, Christophe, 56
Conque, 8
Cornélien, 39
Cortés, Hernán, 47, 48
56, 58
Coton, 25, 26, 54
Cuisine, 18, 18-19, 23

Index

ca, 8, 12, 12, 13
39, 44, 45, 56-7

aguar, 19, 28
ungle, 25

ac Texcoco, 6, 10, 10,
54, 58
Laurier, 6, 10, 12, 13
25, 52, 54, 60
Légumes, 50
Lépreux, 26, 27
Lézard rôti, 25
Libellule, 26
Lois, 12, 12, 14, 41

Magellan, Fernand de, 55
Magie, 25, 31, 48, 55
Maïs, 18, 18, 21, 21
Marchés, 8, 14-15, 14
Mariage, 24, 24, 26

Index

Obsidienne, 38, 40, 43
Oppression, 12, 12, 13

Palais, 8, 11, 15, 17
Palais d'Axayacatl, 11
Papier-mâché, 52, 52
Patolli, 28
Paysans, 12, 13, 16-17
Peinture et dessin, 22
Pièces de théâtre, 28
Pizarre, François, 56
Poésie, 28
Portugais, 56
Poterie, 15, 16, 27
Présages, 48
Prêtres, 8, 23, 30
52, 60
Prix, 50, 5

8. Tu veux savoir si les Aztèques répondaient à leurs besoins de nourriture en faisant du commerce avec d'autres communautés. Quel mot de l'index te conduirait le plus probablement à l'information dont tu as besoin?

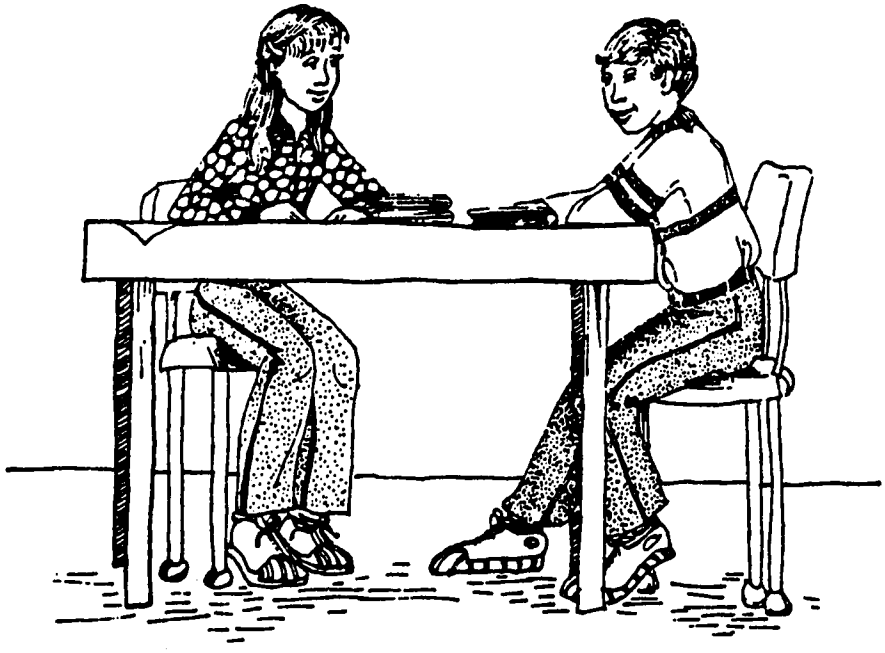
A. Bambou
B. Cuisine
C. Marchés
D. Prix

Utilise l'information suivante pour répondre aux questions 9 à 11.

Catherine et Laurier ont participé à un débat en classe. Catherine a choisi de parler de la civilisation des Mayas. Laurier a parlé des gens de son entourage.

Catherine a dit:
J'aime la façon de prendre des décisions chez les Mayas. Leur système était dur, mais les Mayas au moins connaissaient les règles. Le *Halach Uinic* (l'autorité suprême) et les grands prêtres vivaient dans des palais. Les nobles devaient vivre près du centre-ville, tandis que les fermiers et les paysans habitaient de petites huttes au bord de la ville.

Laurier a dit:
Oui, mais nous avons plus de liberté dans notre société. Certaines personnes ne pensent pas qu'il soit important de dépenser la plus grande partie de leur argent pour se loger. D'autres aiment avoir de belles maisons et travaillent dur pour gagner plus d'argent. Ces gens-là peuvent se permettre de louer ou de posséder de plus grandes maisons. Je pense que notre système est très bien.



9. Catherine et Laurier ont différentes idées sur la façon dont, dans une société, les gens devraient prendre une décision sur

   A. ce qu'on devrait payer pour se loger
   B. ce qu'on devrait demander pour un logement
   C. qui a le meilleur type de maison
   D. qui a sa propre maison

10. Selon Catherine, un avantage du système de logement des Mayas était que les gens

   A. n'avaient pas à s'endetter
   B. savaient ce qu'on attendait d'eux
   C. obtenaient un logement sans avoir à travailler
   D. travaillaient dur pour obtenir ce qu'ils voulaient

11. Laurier pense qu'il est important pour les gens

   A. d'être prêts à faire plus d'efforts
   B. d'acheter ce qu'ils peuvent s'offrir
   C. de pouvoir faire des choix
   D. de louer les meilleures maisons

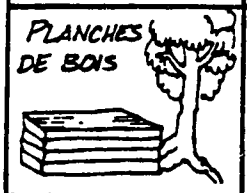
---

12. Dans LA PLUPART des civilisations du passé, la richesse et la puissance étaient

   A. détenues principalement par les marchands et les commerçants
   B. détenues principalement par les nobles et les prêtres
   C. partagées également par les guerriers
   D. partagées également par tous les citoyens

Utilise le tableau ci-dessous pour répondre aux questions 13 à 15.

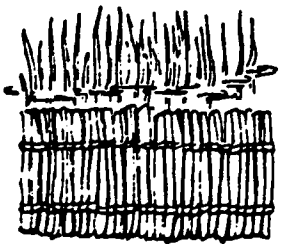
Matériaux de construction utilisés par les civilisations du passé

MATERIAUX	QUALITE DU MATERIAU	DISPONIBILITE DU MATERIAU	RESISTANCE	CAPACITES NECESSAIRES POUR FABRIQUER ET UTILISER LE MATERIAU	MAIN D'OEUVRE NECESSAIRE
ROSEAUX COLLÉS ENSEMBLE AVEC DE LA BOUE 	mauvaise qualité	facile à trouver	doit souvent être réparé	facile à utiliser	un ou deux travailleurs
BRIQUES D'ARGILE 	bonne qualité	facile à trouver	durera de nombreuses années	certaines capacités	un ou beaucoup de travailleurs
PLANCHES DE BOIS 	bonne qualité	ne se trouve qu'à certains endroits	durera de nombreuses années	beaucoup de capacités	beaucoup de travailleurs
BLOCS DE PIERRE TAILLÉS DANS UNE CARRIERE 	excellente qualité	ne se trouve qu'à certains endroits	survivra à la civilisation	de grandes capacités	un grand nombre de travailleurs

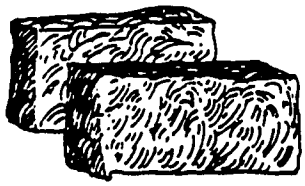
13. Un désavantage d'utiliser les roseaux comme matériaux de construction était lié

A. à leur disponibilité
B. à leur résistance
C. aux capacités nécessaires pour les utiliser
D. à la main d'oeuvre nécessaire pour s'en servir dans la construction

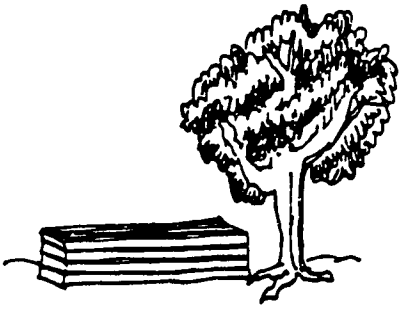
14. Le travail des ouvriers devait être LE PLUS organisé quand ils construisaient avec



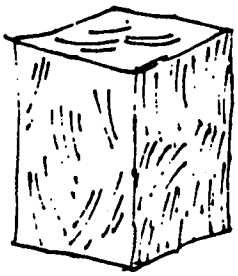
A. des roseaux



B. des briques d'argile



C. des planches de bois



D. de la pierre taillée

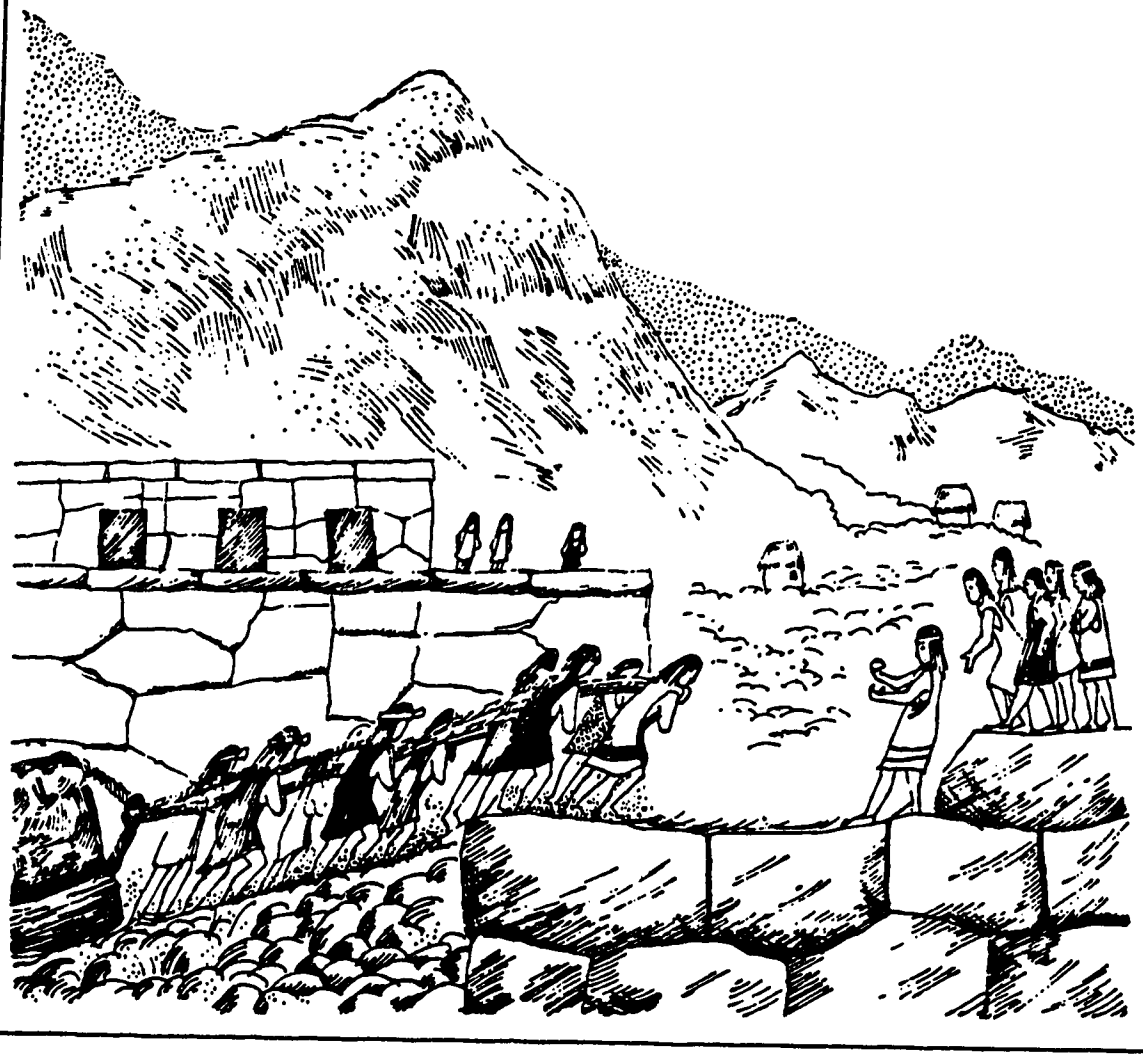
15. "Dans les civilisations du passé, seuls les gens très riches construisaient des maisons de pierre taillée."

Quelle information du tableau constitue LE MEILLEUR argument en faveur de l'affirmation ci-dessus?

A. Il faut beaucoup de main d'oeuvre très spécialisée pour construire avec de la pierre taillée.
B. Les petites maisons ne se construisent pas facilement en pierre taillée.
C. Les maisons construites en pierre taillée durent longtemps.
D. La pierre taillée ne se trouve que dans les carrières.

Utilise l'information ci-dessous pour répondre aux questions 16 et 17.

Les constructions importantes comme le palais de l'*Inca* (le souverain) et les temples étaient bâties par les paysans payant leur *mita*. La *mita* était une taxe qui était payée en travaillant pour l'*Inca*. Une fois par an, chaque paysan passait deux ou trois semaines à travailler à des projets de construction. Les ouvriers étaient nourris, mais pas payés quand ils travaillaient pour l'*Inca*.



16. Comment une patinoire publique serait-elle construite en Alberta aujourd'hui si le système de la *mita* était suivi exactement?

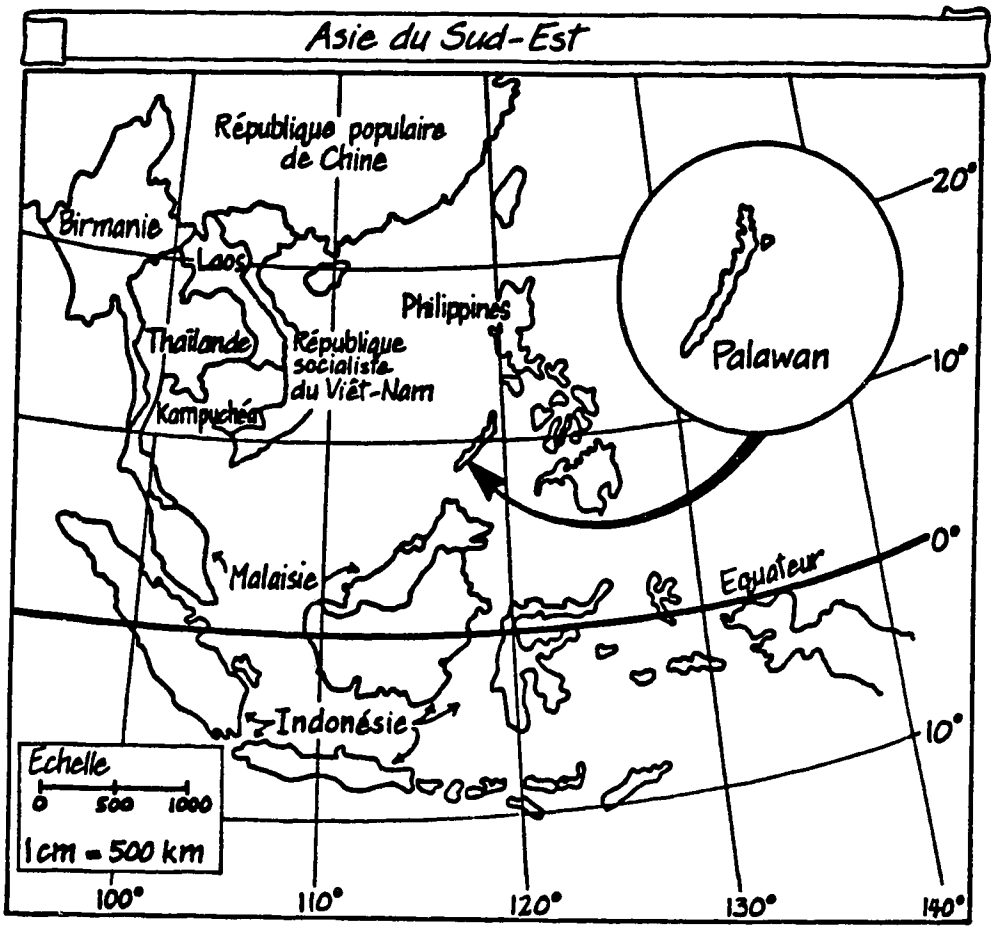
A. De la main d'oeuvre masculine la construirait sans être payée.
B. Des travailleurs étrangers seraient payés pour la construire.
C. Une compagnie privée serait payée pour la construire.
D. Des employés du gouvernement la construiraient sans être payés.

17. Pourquoi quelqu'un qui apprécie un traitement égal pour tous considérerait-il le système de la *mita* injuste?

A. Cela mettrait plus longtemps à terminer les projets.
B. Les taxes n'auraient plus à être payées.
C. Certains ouvriers seraient plus payés que d'autres.
D. Certains devraient travailler et pas les autres.

CETTE PARTIE DU TEST PORTE SUR LA FAÇON DONT LES SOCIETES ORIENTALES SATISFONT LEURS BESOINS ACTUELS.

Utilise la carte ci-dessous pour répondre à la question 18.



18. Quelle information sur la carte est en faveur de l'argument que la température à Palawan varie peu entre l'hiver et l'été?

A. Palawan est une grande île.
B. Palawan est près de l'équateur.
C. Palawan a des chaînes de montagnes.
D. Palawan a une mousson d'été.

19. Lequel des besoins essentiels suivants est LE MOINS affecté par le climat d'un pays?

A. L'alimentation
B. L'habitation
C. Le vêtement
D. L'éducation

20. Quelle affirmation décrit LE MIEUX la façon dont les gens réagissent aux changements dans une société?

A. Les gens changent facilement leurs traditions.
B. Les gens ne changent pas leurs traditions.
C. Dans toutes les sociétés, les gens essaient d'empêcher leurs traditions de changer.
D. Les gens changent seulement leurs traditions quand ils ont des idées qui viennent d'autres sociétés.

21. Quelle affirmation montre que la tradition affecte la façon dont les gens satisfont leurs besoins alimentaires?

A. Certaines personnes ne mangent pas de viande à cause de leurs convictions religieuses.
B. Certaines personnes mangent souvent des fruits parce qu'ils contiennent des vitamines.
C. On fait du pain avec toutes sortes de farines.
D. On ne cultive pas de riz dans les climats secs.

22. Une raison CAPITALE d'importer des marchandises d'Asie du Sud-Est est

A. d'apprendre à connaître d'autres modes de vie
B. d'épargner nos propres marchandises pour plus tard
C. de venir en aide aux populations d'Asie du Sud-Est
D. d'avoir des choses qui ne sont pas cultivées ou fabriquées au Canada

23. Le but d'organismes tels que l'Agence canadienne de développement international (ACDI) est d'améliorer les conditions de vie dans les pays d'Asie du Sud-Est en

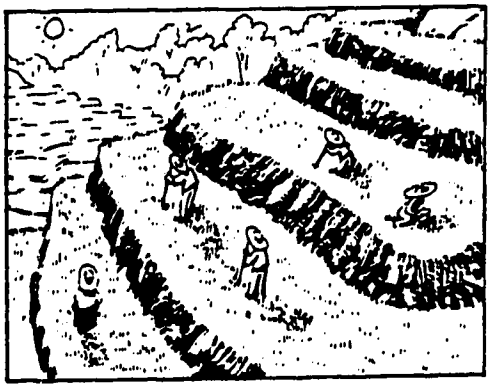
A. finançant des programmes d'aide locaux
B. vendant aux gens du blé à bas prix
C. ramassant des vêtements usagés pour les expédier outre-mer
D. encourageant les habitants de l'Asie du Sud-Est à venir s'installer au Canada

24. Certains pays d'Asie du Sud-Est ont beaucoup d'habitants à nourrir mais très peu de terres cultivables pour faire pousser leurs récoltes. Une façon dont ils essaient de résoudre ce problème est en

A. envoyant les habitants travailler dans les champs

B. divisant les terres de sorte que chaque famille en ait une partie

C. déplaçant les habitants pour qu'ils puissent travailler la terre dans d'autres pays



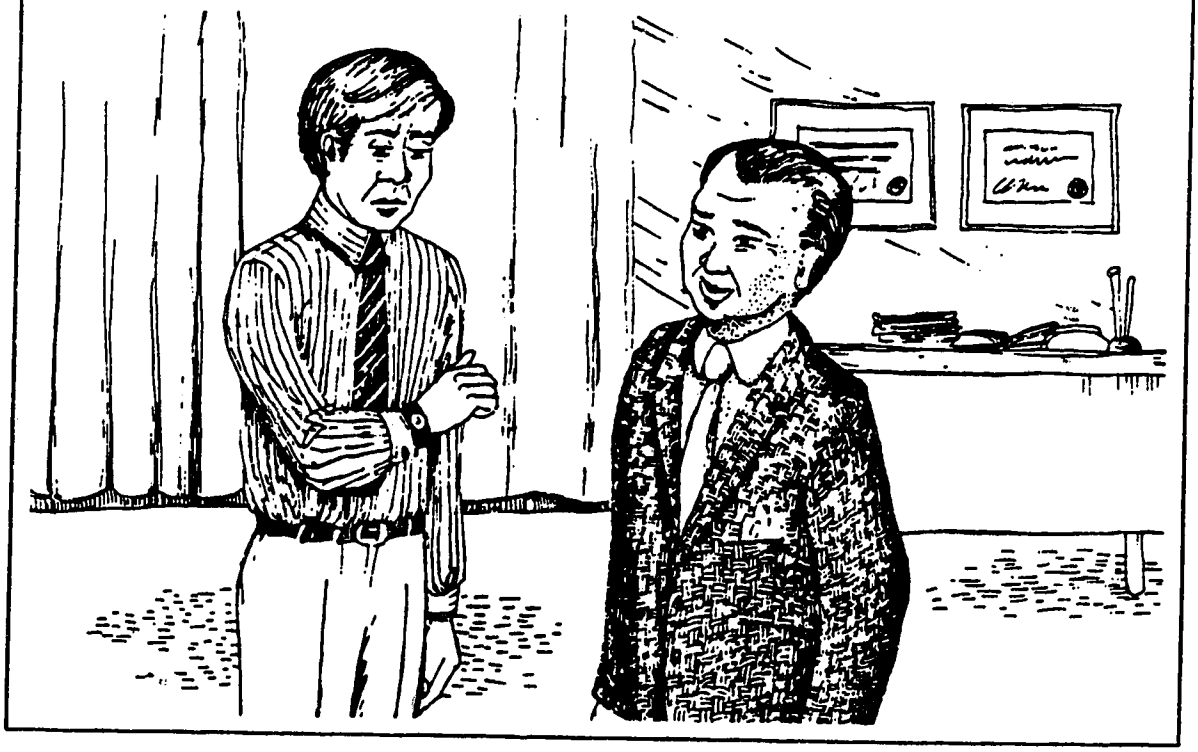
D. augmentant la quantité de terres cultivables en arrangeant des terrasses au flanc des collines

Utilise l'information suivante pour répondre aux questions 25 à 28.

M. Tremblay, gérant d'une compagnie canadienne d'ingénierie, travaille dans un pays d'Asie du Sud-Est. Il est surpris quand M. Ochoa, homme d'affaires local, arrive avec 30 minutes de retard pour son rendez-vous. L'échange qui suit fait partie de la conversation qu'ils ont eue.

M. Tremblay: "Vous êtes en retard. Je vous attendais il y a une demi-heure. Il faut des horaires et respecter des délais stricts si nous voulons bien faire ce travail. Dans le monde de l'ingénierie, le temps c'est de l'argent. Mon personnel a une excellente formation et cela veut dire qu'il est bien payé. On ne peut pas se permettre de le faire attendre."

M. Ochoa: "Vous comprenez mal nos valeurs. Je n'ai jamais eu l'intention d'arriver à l'heure fixée; mes compatriotes ne m'auraient attendu que plus tard. Arriver à temps veut dire que je ne suis pas très important, que je n'ai rien de mieux à faire que de venir à ce rendez-vous. Un homme actif et prospère a tant de responsabilités qu'il a le droit d'être en retard."



25. Quel point sur le travail dans un pays qui a une culture différente est discuté par M. Tremblay et M. Ochoa?

A. Qui devrait travailler dans d'autres pays?
B. Comment les gens devraient-ils tenir leurs rendez-vous d'affaires?
C. Qui devrait être autorisé à assister aux réunions?
D. Comment les ouvriers devraient-ils être payés?

26. Le conflit est entre le désir d'une personne d'être traitée avec dignité et le besoin d'une autre personne

A. de faire son travail à l'heure
B. d'être respectée
C. d'utiliser la technologie moderne
D. d'engager des ouvriers ayant une bonne formation

27. Dans la conduite de ses affaires, M. Tremblay attache LE PLUS d'importance à

A. suivre les traditions des autres
B. travailler dans d'autres pays
C. employer des ouvriers ayant une bonne formation
D. être très efficace

28. Le malentendu qui s'est produit entre les deux hommes résulte

A. des langues qu'ils parlent
B. des coutûmes qu'ils suivent
C. de la technologie qu'ils emploient
D. du travail qu'ils font

---

29. Une raison majeure pour laquelle des Canadiens sont engagés pour travailler à des projets de construction en Asie du Sud-Est est

A. que les Canadiens ont besoin d'apprendre les manières de l'Asie du Sud-Est
B. que les Canadiens ont besoin d'aide pour résoudre leurs problèmes de chômage
C. qu'on a besoin d'ingénieurs qui ont une formation spéciale en Asie du Sud-Est
D. que cela est trop cher d'engager des ingénieurs d'Asie du Sud-Est

Utilise les commentaires ci-dessous pour répondre aux questions 30 et 31.

Je téléphone beaucoup. Je voyage beaucoup pour mon travail, aussi j'ai beaucoup d'amis dans tout le pays. Je ne peux pas les recontrer souvent, il faut donc que je reste en contact par téléphone. J'utilise souvent le téléphone aussi pour mon travail, parce qu'il faut que je parle avec des programmeurs d'ordinateur dans tout le Canada.

-- Christine Drapeau,
expert-conseil au Canada

Je n'ai pas besoin d'utiliser le téléphone. Après avoir vendu mes poissons chaque matin, je vais au café prendre mon petit déjeuner. J'y rencontre mes amis et nous discutons les nouvelles du village. Parfois, les chauffeurs de camion s'y arrêtent en route vers Kuala Lumpur et amènent des nouvelles d'autres villages.

-- Abraham B. Hasan,
pêcheur malaysien

30. Christine et Abraham ont des idées différentes sur le besoin d'utiliser le téléphone parce que Christine a

A. plus d'amis
B. un style de vie différent
C. des croyances religieuses différentes
D. un besoin plus grand de parler avec des amis

31. Une affirmation générale sur toutes les sociétés qui peut être soutenue par les deux commentaires est que les gens

A. aiment communiquer avec leurs amis
B. doivent utiliser le téléphone pour leurs affaires
C. ont besoin de parler à leurs amis tous les jours
D. veulent avoir des amis dans des endroits éloignés

Utilise l'information ci-dessous pour répondre aux questions 32 et 33.

Cela coûte cher de posséder une voiture et de s'en servir, pourtant tous les Canadiens éprouvent le besoin d'en avoir une. C'est un gaspillage terrible. On voit souvent une personne aller au travail dans une immense voiture qui pourrait transporter six personnes. Il n'est pas étonnant que les Canadiens aient des embouteillages dans les grandes villes.

-- Somchai Meesiri, étudiant de Thaïlande

32. L'opinion de Somchai sur la possession d'une voiture diffère de celle de beaucoup de Canadiens. Laquelle des affirmations suivantes explique LE MIEUX cette différence?

A. Les Canadiens peuvent obtenir un permis de conduire plus facilement que les Thaïlandais.
B. Les Canadiens ont les moyens d'acheter une voiture plus facilement que les Thaïlandais.
C. Le climat du Canada est plus froid que celui de la Thaïlande.
D. Les routes sont meilleures au Canada qu'en Thaïlande.

33. Une solution aux embouteillages au Canada qui encouragerait une plus grande COOPERATION serait

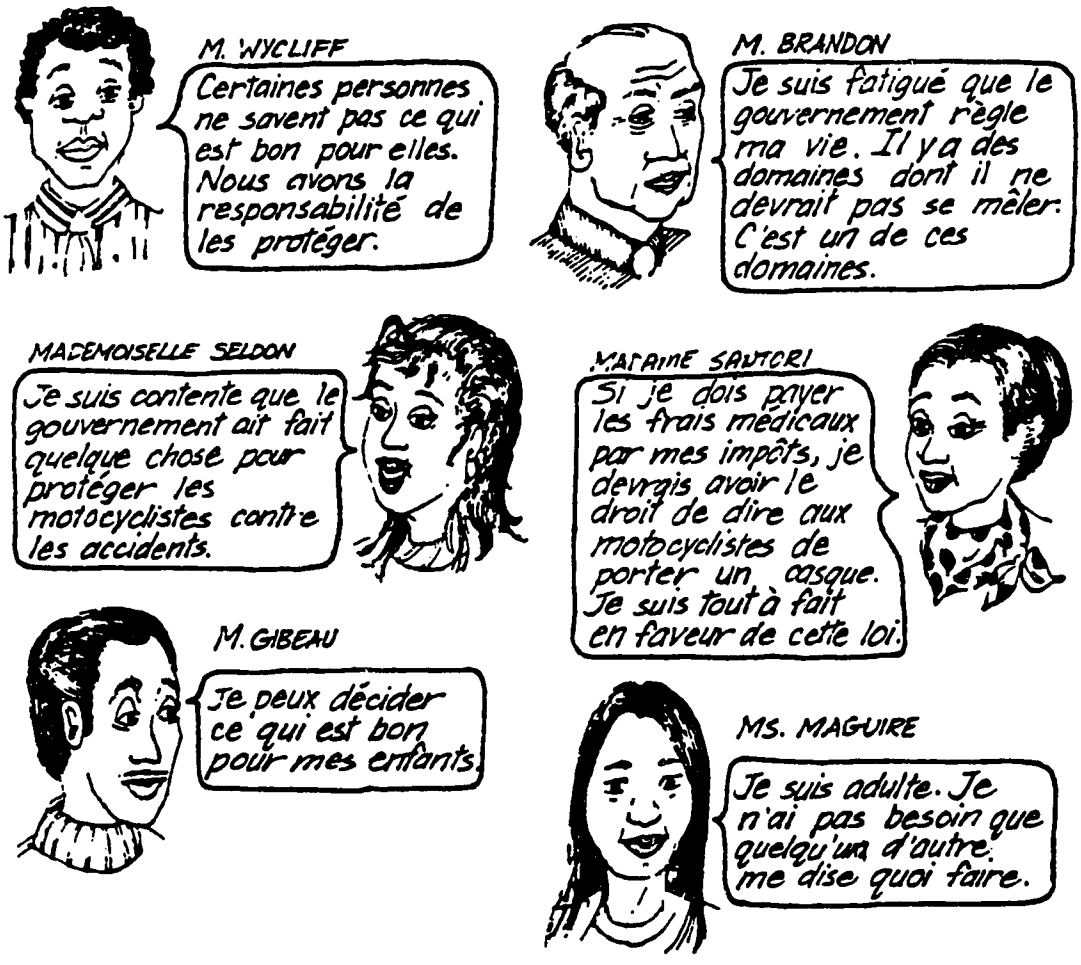
A. d'aller au travail à pied
B. d'acheter des voitures moins chères
C. de se grouper pour voyager
D. de conduire de petites voitures

CETTE PARTIE DU TEST PORTE SUR LA SATISFACTION DES BESOINS: LES GOUVERNEMENTS AU NIVEAU LOCAL, PROVINCIAL ET FEDERAL.

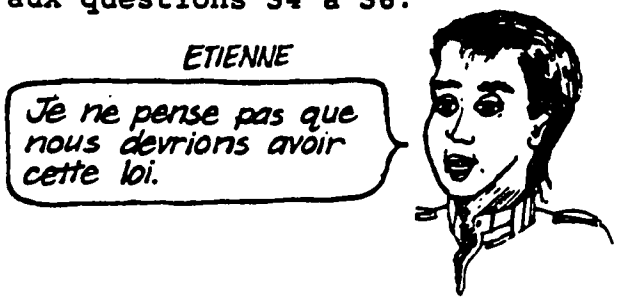
Lis l'information ci-dessous et réponds ensuite aux questions 34 à 36.

La loi dit que les gens qui font de la motocyclette doivent porter un casque. Certaines personnes n'aiment pas cette loi et veulent la voir annulée (enlevée), d'autres soutiennent la loi.

Voici des commentaires faits par des citoyens sur cette question.



Utilise les opinions exprimées à la page 10 pour répondre aux questions 34 à 36.



34. Avec quel interlocuteur Etienne est-il LE PLUS EN DESACCORD?

A. Ms. Maguire

B. M. Gibeau

C. M. Brandon

D. Mademoiselle Seldon

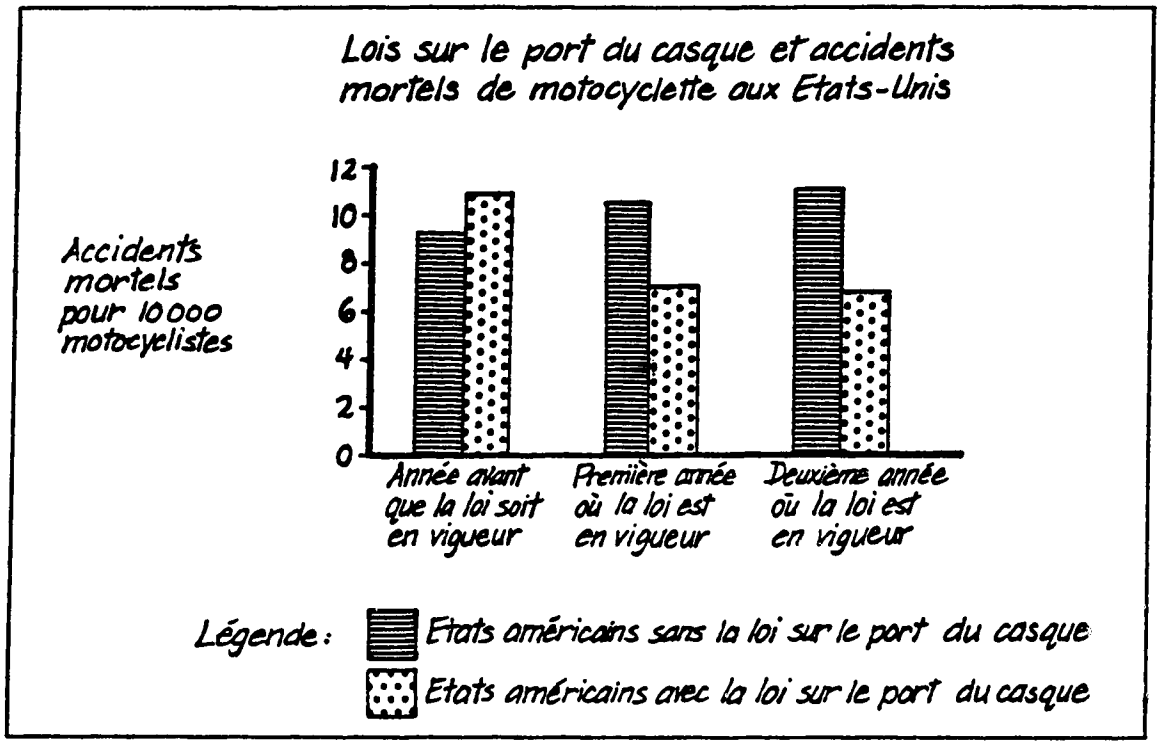
35. Quelle est la PRINCIPALE question discutée par les citoyens?

A. Les impôts devraient-ils servir à payer les accidents de motocyclette?
B. Devrait-il y avoir une loi exigeant le port du casque?
C. Les motocyclistes qui portent un casque devraient-ils être autorisés à payer moins cher pour la sécurité sociale?
D. La loi qui exige le port du casque devrait-elle être limitée aux enfants qui sont passagers?

36. Si on rassemblait en un tableau toutes les opinions exprimées, quel serait le meilleur titre?

A. Opinions sur la loi sur le port du casque.
B. Effets de la loi sur le port du casque.
C. Raisons de maintenir la loi sur le port du casque.
D. Personnes qui ont voté pour la loi sur le port du casque.

Utilise le graphique ci-dessous pour répondre aux questions 37 à 39.



37. A laquelle des questions suivantes pourrait-on répondre en utilisant le graphique?

A. Quand la plupart des accidents de motocyclette arrivent-ils?
B. Pourquoi le port du casque sauve-t-il des vies?
C. Combien de Canadiens meurent dans des accidents de motocyclette?
D. Qu'arriverait-il si nos lois sur le port du casque étaient retirées?

38. Le graphique montre que pour 10 000 motocyclistes, les Etats américains SANS loi sur le port du casque ont

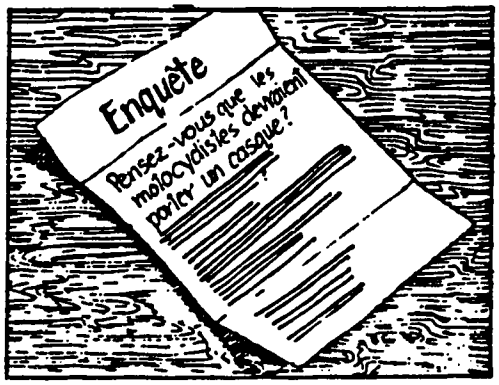
A. plus de morts dues à des accidents de motocyclette que les Etats avec des lois
B. plus d'accidents de motocyclette que les Etats avec des lois
C. moins de gens qui portent un casque que les Etats avec des lois
D. moins de gens faisant de la motocyclette que les Etats avec des lois

39. Quel est le PRINCIPAL point présenté par le graphique?

A. Beaucoup d'Etats ont supprimé ou adouci leurs lois sur le port du casque.
B. Beaucoup d'Etats ont des lois qui exigent le port du casque.
C. Il y a moins de morts en accident de motocyclette quand les gens portent un casque.
D. Il y a eu moins de morts pendant la deuxième année.

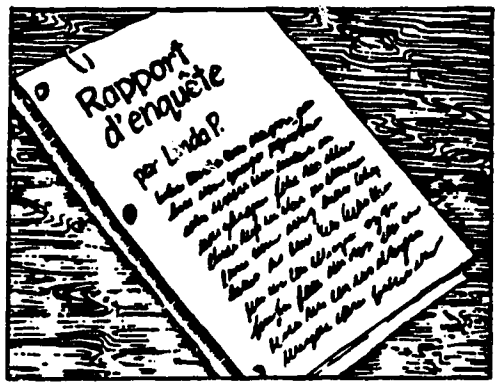
Après avoir appris qu'on pourrait supprimer la loi sur le port du casque, une classe de 6e année a décidé de faire connaître son opinion sur cette question.

40. Quelle action ferait connaître l'opinion des élèves au plus grand nombre de personnes?



A. Faire une enquête auprès des élèves dans les autres écoles

B. Parler au principal à la récréation

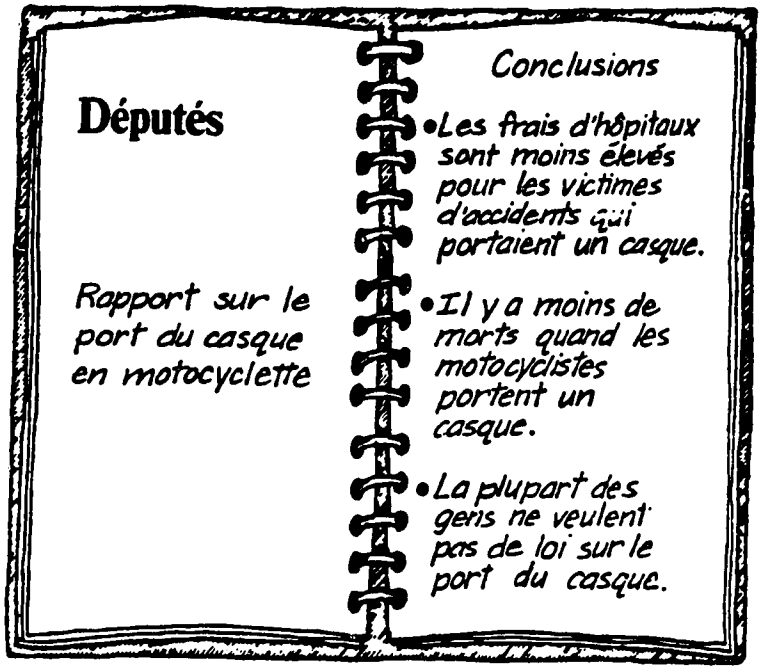


C. Rédiger un rapport d'enquête

D. Présenter la question aux parents

Utilise l'information ci-dessous pour répondre à la question 41.

Un certain nombre de députés ont présenté le rapport suivant au Parlement.



Après la discussion du rapport, on est passé au vote. La loi sur le port des casques a été supprimée (enlevée).

41. Selon le principe du gouvernement par la majorité, la suppression de la loi sur le port du casque était

A. démocratique, parce que la plupart des citoyens ne voulaient pas la loi
B. démocratique, parce que certains députés ont fait des recherches sur la loi
C. antidémocratique, parce que le coût de l'hôpital serait plus élevé pour tous
D. antidémocratique, parce que plus d'Albertains mourraient maintenant

---

42. Une des raisons pour lesquelles le Canada est considéré une démocratie est que nous avons

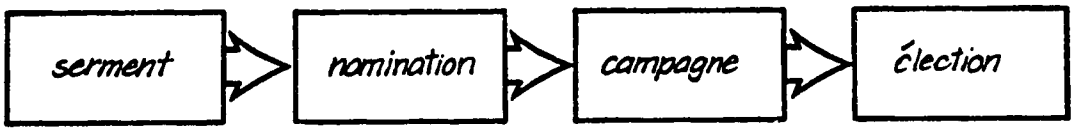
A. un premier ministre dans chaque province
B. des chefs qui sont élus
C. un gouvernement fédéral
D. une constitution

43. La discussion qui a lieu entre les députés sur un projet de loi est appelée

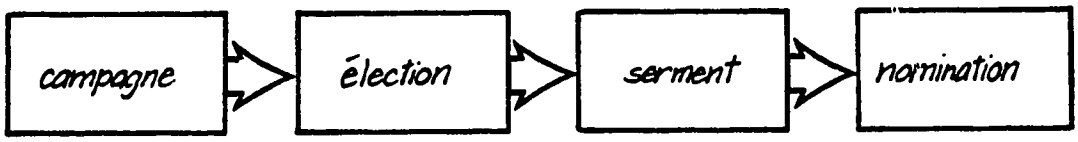
A. groupe de pression
B. enquête
C. débat
D. campagne

44. Les étapes qu'une personne doit suivre pour devenir député sont, dans le BON ORDRE,

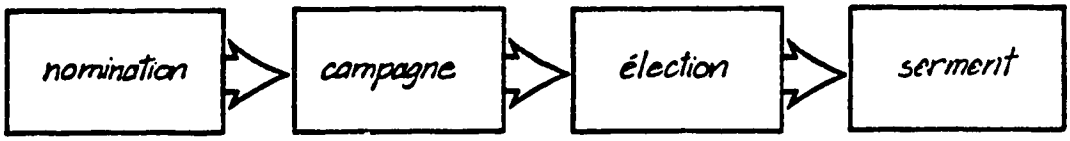
A.



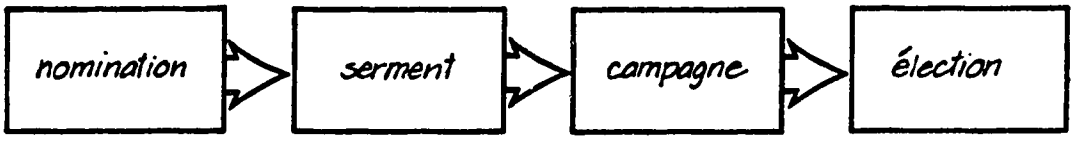
B.



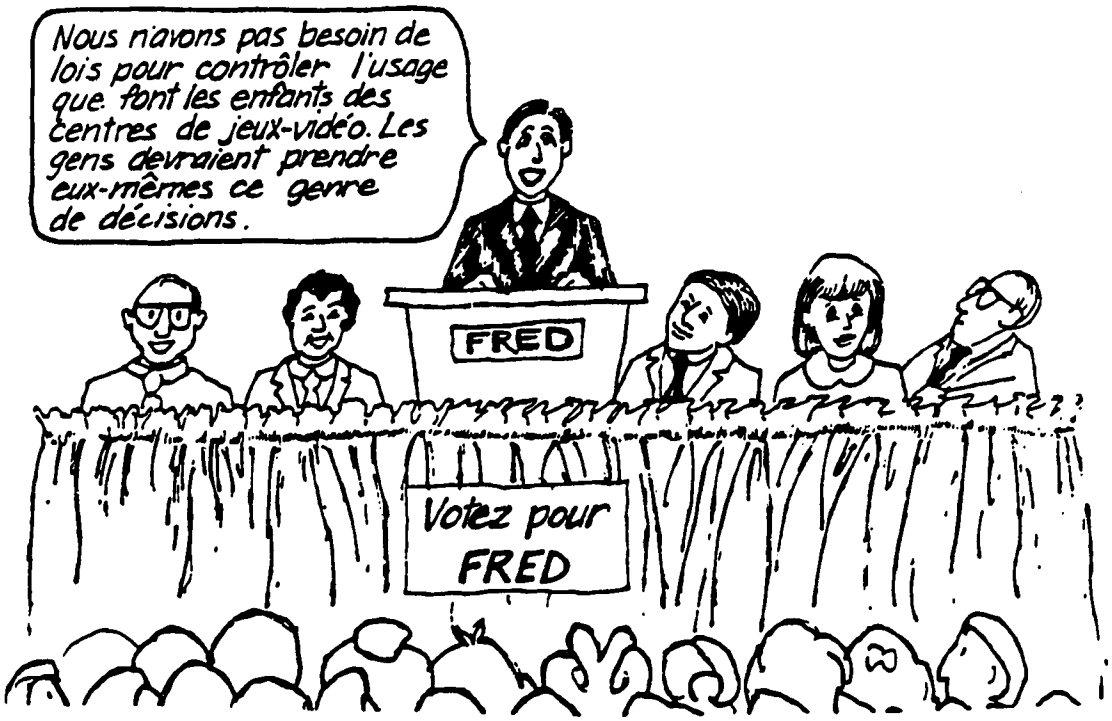
C.



D.



Utilise l'information ci-dessous pour répondre à la question 45.



45. Ce candidat serait LE PLUS PROBABLEMENT soutenu par un électeur qui croirait que

A. les enfants ne devraient pas avoir à obéir aux lois
B. les enfants ne devraient pas avoir leur mot à dire quand il s'agit de faire les lois
C. les gens devraient travailler ensemble pour satisfaire leurs besoins
D. les gens devraient s'occuper de leurs propres besoins

---

46. Normalement, le chef du parti qui a le plus grand nombre de membres élus à la Chambre des communes s'appelle le

A. chef de l'opposition
B. président de la chambre
C. Gouverneur général
D. Premier Ministre

47. Un des buts des partis politiques au Canada est de

A. rassembler des gens qui partagent les mêmes idées sur le gouvernement
B. percevoir (recueillir) les impôts pour payer les salaires des personnes élues
C. nommer les membres des conseils municipaux
D. choisir les membres qui siègent au Sénat

48. Le principal rôle de l'opposition au gouvernement fédéral est de

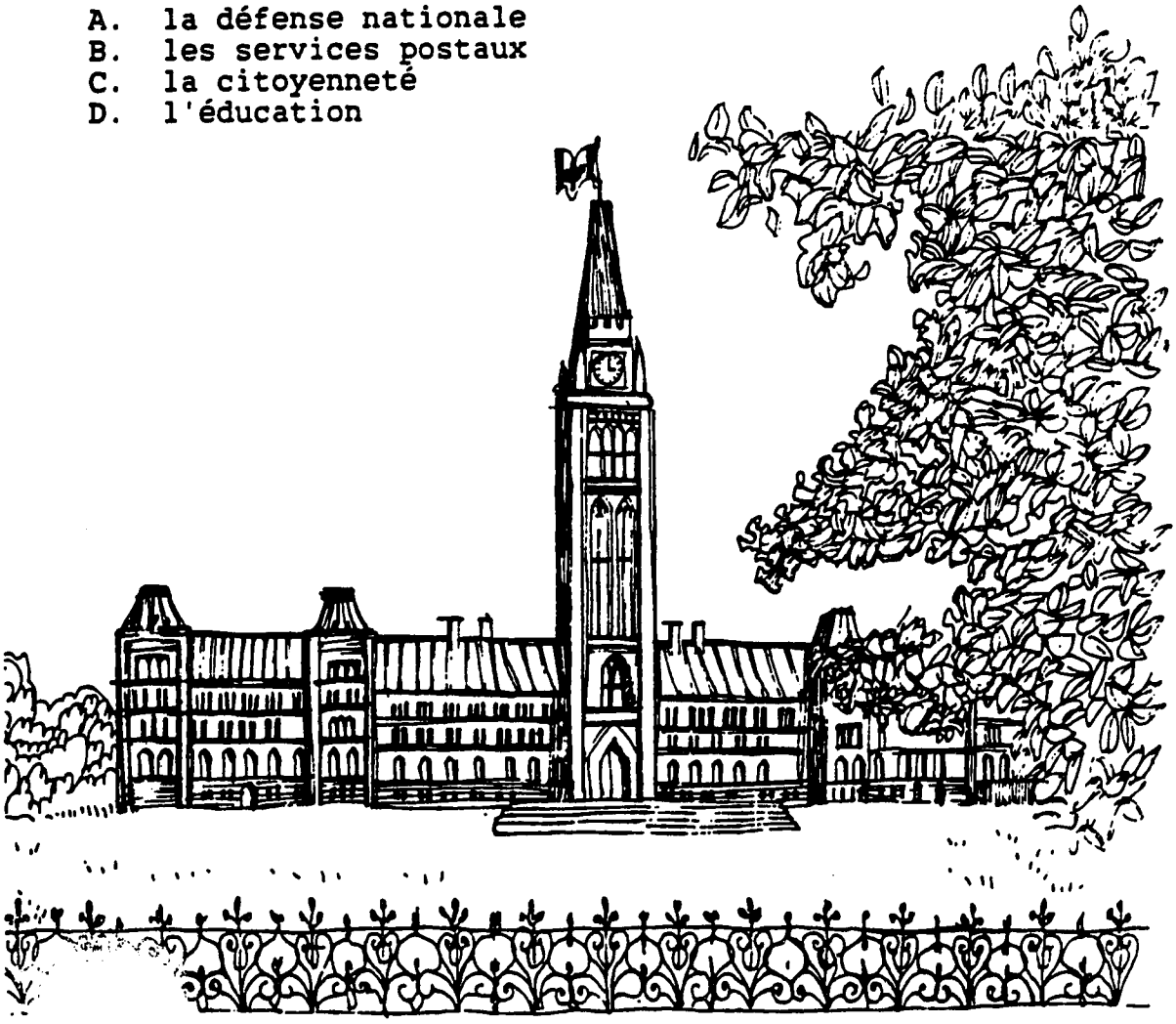
A. préparer des projets de loi pour le gouvernement
B. protéger les droits des provinces
C. présenter d'autres points de vue
D. soutenir le parti au pouvoir

49. Au Canada les gouvernements ne paient généralement pas la construction des

A. écoles
B. églises
C. hôpitaux
D. casernes de pompiers

50. Les gouvernements provinciaux peuvent passer des lois sur

A. la défense nationale
B. les services postaux
C. la citoyenneté
D. l'éducation



C ' E S T L A F I N D U T E S T

Tu peux maintenant vérifier tes réponses.

Merci d'avoir fait le test de rendement d'Etudes sociales de 6° année.

CREDITS

Sources 25-29 "Businessmen from the States. . . he is entitled to be late." Adapté et abrégé à partir de *THE LAND AND PEOPLE OF THE PHILIPPINES* par John Nance (J. B. Lippincott Company).