

DOCUMENT RESUME

ED 324 957

FL 018 943

TITLE Provincial Assessment of Students in French Immersion Programs. Special Report.
 INSTITUTION Alberta Dept. of Education, Edmonton.
 PUB DATE Jun 90
 NOTE 29p.
 PUB TYPE Reports - Evaluative/Feasibility (142) -- Statistical Data (110)

EDRS PRICE MF01/PC02 Plus Postage.
 DESCRIPTORS *Academic Achievement; Academic Standards; Elementary Education; English; Evaluation Criteria; *Evaluation Methods; Foreign Countries; *French; Grade 3; Grade 6; *Immersion Programs; Language Arts; Second Language Programs; Social Studies; State Programs; *Student Evaluation; Student Participation; Testing Programs; Test Selection
 IDENTIFIERS *Alberta

ABSTRACT

This special report on the assessment of French immersion students in Alberta focuses on critical issues and methodological problems with testing methods. After an introduction, the second section gives an overview of these issues and problems. Issues discussed include the following: whether students instructed in French immersion programs should be assessed and results reported on a provincial basis; whether participation in the achievement testing program should be mandatory for this population; in what language the students should be tested; whether all students should be tested in the same language; whether achievement should be assessed using the same tests used in the regular achievement testing program; against what standards student achievement should be measured; and against what reference group student achievement should be compared. The third section reports on a 1989 study to address the effect of the test's language on achievement. In the study, grade 3 and grade 6 French immersion students took the social studies achievement test in either English or French. Results showed that students testing in French scored significantly lower than those testing in English. Section 4 presents some statistical data concerning identification of appropriate expectations for French immersion students. The final section draws conclusions and discusses future directions for policy-making. (MSE)

 * Reproductions supplied by EDRS are the best that can be made *
 * from the original document. *

IN FRENCH IMMERSION PROGRAMS

SPECIAL REPORT

JUNE 1990

ED324957

PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

S. Wolodko

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

X This document has been reproduced as
received from the person or organization
originating it.

Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official
OERI position or policy.

Student Evaluation



CONTENTS

SECTION 1:	INTRODUCTION	1
SECTION 2:	OVERVIEW OF ISSUES RELATED TO ASSESSING STUDENTS INSTRUCTED IN FRENCH IMMERSION PROGRAMS	3
SECTION 3:	LANGUAGE OF TESTING STUDY	9
SECTION 4:	APPROPRIATE STANDARDS AND REFERENCE GROUPS FOR STUDENTS INSTRUCTED IN FRENCH IMMERSION PROGRAMS	19
SECTION 5:	CONCLUSIONS AND FUTURE DIRECTIONS	27

Copyright 1990, the Crown in Right of Alberta, as represented by the Minister of Education, Alberta Education, Student Evaluation Branch, 11160 Jasper Avenue, Edmonton, Alberta, T5K 0L2. All rights reserved. Additional copies may be obtained from the Student Evaluation Branch.

SECTION 1

INTRODUCTION

Since the inception of the Achievement Testing Program in 1982, students in French Immersion programs have had the option of being exempted from the tests. However, more and more superintendents want full participation for their French Immersion students. As a result, the number who take achievement tests has risen dramatically, and now over 80 per cent of French Immersion students in grades 6 and 9 write the tests.

Many critical issues and methodological problems surround the assessment of French Immersion students, however. Because of the importance of this matter to students, parents, educators, and the public at large, we believe that these issues and problems should be discussed by the broadest possible audience before changes are made to the existing testing program.

The purpose of this paper is to help focus the discussion, using achievement data collected recently. We believe that the new information in this report and the discussion that arises from it will help us find the best course of action.

Section 2 of this paper provides an overview of these issues and methodological problems. These must be resolved to ensure that French Immersion students can be assessed, and the results reported and interpreted, validly and reliably. Section 3 describes the results of a special study that was done by the Student Evaluation Branch to find out if language of testing is a variable that affects the way French Immersion students respond to test questions. Section 4 presents other related information that has been collected through the Achievement Testing Program already. The final section summarizes the information that is presented and identifies possible future directions.

SECTION 2

OVERVIEW OF ISSUES RELATED TO ASSESSING STUDENTS INSTRUCTED IN FRENCH IMMERSION PROGRAMS

BASIC ISSUE

The main issue that needs to be resolved when assessing students in French Immersion programs is:

Should students instructed in French Immersion programs be assessed and the results reported on a provincial basis?

Many pedagogical, technical, and political factors underlie this issue. These often conflicting or competing factors need to be understood and addressed for this issue to be resolved successfully. For example, while we want meaningful and useful information about student achievement, we want to ensure that French Immersion programs are not undermined by any psychometric practice or process. Yet we know that meaningful and useful achievement information can only be provided if assessment practices are valid and reliable. Unfortunately, what is most valid and reliable statistically may not be seen to be appropriate politically or pedagogically.

As with any issue as complex as this one, a number of related sub-issues must be addressed before we can resolve the main issue.

1. Should participation in the Achievement Testing Program be mandatory for students in French Immersion programs?
2. In what language should French Immersion students be tested?
3. Should all French Immersion students be tested in the same language?
4. Should the achievement of French Immersion students be assessed using the same tests that are used in the regular Achievement Testing Program?
5. Against what standards should the achievement of French Immersion students be compared?
6. Against what reference group should the achievement of French Immersion students be compared?

RELATED ISSUES

Each of these issues is discussed below.

1. *Should participation in the Achievement Testing Program be mandatory for students in French Immersion programs?*

Achievement testing for French Immersion students is currently optional, yet most school administrators choose to have their students participate. This high level of participation suggests that these educators want to know about the levels of achievement in their French Immersion programs. Presumably, this information is helpful in making decisions about the program. But to make such educational decisions, one needs to know not only the levels of achievement but also if those levels are "good enough". This type of judgment cannot be made without comparing test scores to some point of reference. In the case of French Immersion, that point of reference could be provincial standards and/or average scores for French Immersion students.

To be meaningful and useful as a point of reference, provincial test scores must be representative of the population under investigation. The results are representative only if everyone in the population or a representative sample of that population is tested. These conditions cannot be met as long as participation in the Achievement Testing Program remains optional for French Immersion students.

If participation in the Achievement Testing Program were mandatory for all students in French Immersion programs, then the only students who could be excused from participating would be those for whom a particular test was inappropriate. This would include students who were enrolled in a special needs program or those whose instruction in the course being tested was in another semester or year.

Educators opposed to mandatory participation argue that because of the nature of French Immersion programs (their recency and variability in the amount of daily instruction in French, for example), there is a high potential for misinterpretation of achievement test results. They feel that misinterpretation, if negative, could jeopardize support for these programs. Thus, they believe they should have the opportunity to opt out of a testing program.

The counter-argument is that it is precisely because of the recency and the variable nature of French Immersion programs that achievement in these classes should be monitored closely. Without such information, it is difficult to know the relative strengths and weaknesses of the program. In the absence of this awareness, one's ability to make program adjustments, where desirable, is limited. In short, rather than fearing negative repercussions from results that are unexpected or unfavorable, supporters of mandatory testing argue that one should use that information to make adjustments or improvements to the program.

2. *In what language should French Immersion students be tested?*

Participation in the Achievement Testing Program is optional for those students 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 notion that in order for participation to be appropriate for any given student, his or her language of instruction and language of fluency must match the language of testing. A match between the language of instruction and language of testing is important because conclusions about student achievement may only be valid if the test questions assess what students have learned in that language. Testing students in their language of fluency is also important to validity because what students know and can do can only be fairly assessed if examinees can easily read and understand the test questions.

In French Immersion classes, the language of instruction and the language in which most students are fluent are not necessarily the same. This poses a dilemma in choosing a language of testing because, regardless of which language is chosen, there is a potential risk. Choosing to test in the language of instruction when it is not the language of fluency could cause scores to be artificially low. These artificially low scores could cause negative political and pedagogical implications if they are interpreted to mean that lower achievement is both unexpected and unacceptable. On the other hand, a negative message could be seen to be sent to educators and the public about our confidence in our students' capabilities if French Immersion students were to be assessed in English (i.e. the language of fluency) rather than French, the language of instruction.

3. *Should all French Immersion students be tested in the same language?*

An issue that is related to the question "*In what language should French Immersion students be tested?*" is whether or not all French Immersion students should be tested in the same language. If the answer is that the language of testing should be variable, then a secondary issue concerns the level (individual, class, school, jurisdiction) at which the language of testing option should be allowed to occur.

The primary argument for mandating a single language of testing in French Immersion programs is that this decision would hold constant a variable that, it has been argued, affects how students respond to test questions. English tests and their French translation may not be equally difficult, because the translated questions might be easier or more difficult to answer. This means that students could have more or less trouble selecting the correct answer depending on which form of the test they write, the English original or its French translation. In other words, their scores could vary depending on the language of testing. Similarly, students may have more or less trouble reading and understanding the questions when they are presented in one language rather than the other because of unequal levels of first and second language reading ability. Since the selection of a correct answer to a test question depends in part on the ability to understand the question, then once again, student scores could vary depending on the language of testing.

Mandating one language of testing province-wide will not eliminate whatever effect choosing that particular language has on the students' ability to respond, but what it will do is hold that effect constant. For example, if scores are lower than expected because the form of the test chosen is more difficult, at least all groups of French Immersion students who are tested will suffer the same score depression, making it easier to interpret test scores. Essentially, the argument against mandating one language of testing is that such a policy may be seen as insensitive to local needs and conditions.

4. *Should the achievement of French Immersion students be assessed using the same tests that are used in the regular Achievement Testing Program?*

A strength of the Achievement Testing Program is that the tests are based on Alberta programs of study. Because of this match between what is expected and what is tested, the results of testing can be used in conjunction with other sources of information to make meaningful and useful decisions about curriculum, resources, instruction, and so on. Obviously, the goal of French achievement testing is to have results that are equally meaningful and useful, and therefore tests must be curricularly valid. In general, the learning objectives in the French Immersion programs are the same as those in the regular (English) program. This implies that, at least in principle, the English program achievement tests are also a fair representation of French Immersion program curricula. What may differ are the expected standards of achievement within a French Immersion program or subtleties in the way a curriculum is interpreted, for example. The question, then, is whether or not these differences are great enough to warrant producing separate achievement tests for the French Immersion programs in science, mathematics and social studies; language arts tests are currently produced separately for the English and French programs.

Assuming that the curricula match across programs but that the standards of performance vary, it is possible to address this difference without having to use separately developed tests. In the case of multiple-choice tests, all that would be required is to select different scores as the cut-off points representing the expected standards of achievement in each program. For written-response items, the scoring standard could be adjusted without having to alter the nature of the assessment tasks.

Three factors support the use of English program tests with French Immersion students. **First**, by using the same tests, it is possible to compare achievement between the two programs. It is precisely this point that causes some to argue for separate tests, since they see this form of comparison as undesirable. **Second**, it is more cost efficient to translate tests than it is to produce separate French and English versions. Thus, if the content of English program tests is representative of French Immersion curricula, then it is more desirable to use resources for other initiatives rather than to produce essentially parallel tests in two languages. **Finally**, since the technical merit of a test is determined in part by the size of the population in which it can be field tested, then English program tests are likely to be more valid and reliable indicators of

achievement than test forms developed separately for French Immersion students would be. Therefore, whatever validity that might be lost by using translations of English program tests would be gained by using instruments that are technically more sound.

The essential argument for not using English program tests in their translated form is that the level of language used on those translations is too difficult for many French Immersion students. The level of language used on the English forms, and therefore the French translations, is that which is readable to a native language speaker of the age and grade level being tested. It follows that because French is a second and not a native language, French Immersion students may be unable to fully comprehend the test material.

The argument follows that if separate tests were developed for French Immersion students, then levels of language appropriate for second language speakers could be used. The limitation of this argument is that if the text were simplified to the extent that second language readers could comprehend it, it may also be simplified to the point where the difficulty of the material being tested on English and French tests is no longer parallel. This parallelism could be lost because the abstractness of the concepts being measured on a test is closely related to the language used to express those concepts. In other words, the level of subject-specific content being tested could be reduced by simplifying the language used on the test.

5. *Against what standards should the achievement of French Immersion students be compared?*

By themselves, achievement test scores have no meaning. To have meaning, they must be given a context. One way of providing a context is to compare those scores with expected levels of performance. This process requires two distinct judgments. The first is to determine what percentage of students tested can be expected to achieve at least an acceptable level of skill and knowledge. The other is to establish the test score that represents that level.

For French Immersion scores to be reported meaningfully, this comparison to standards seems essential. Thus, the two judgments referred to above must be made. An issue that arises out of this concern is whether or not it is reasonable to expect French Immersion and regular (English language) program students to achieve the same standard of performance on learning objectives that are common. One line of reasoning is that French Immersion students divide their attention among five academic subjects (English language arts, French language arts, social studies, mathematics, and science) rather than four, and thus expectations should be lower for these students. Another line of reasoning is that since participation in French Immersion programs is decided by parents, these students may, in general, come from more supportive home environments than students in the regular program, and thus expectations for this group should be higher. Whatever decision about standards is made, a technical question that arises is how best to set standards so that appropriate expectations exist on French and English forms of a test.

6. *Against what reference group should the achievement of French Immersion students be compared?*

Another way to provide a context for test scores is to compare the performance of the group in question against that of a reference group. In the case of the current Achievement Testing Program, there is no appropriate reference group at the provincial level, since provincial averages are based on the achievement of the entire population of students who were tested. However, provincial averages can be used as the reference group scores against which individual school and jurisdiction levels of performance can be compared.

In the case of French Immersion tests scores, the central issue concerns what is the most appropriate reference group. From the point of view of individual schools and jurisdictions, an appropriate reference group would be all French Immersion students who wrote the test. This assumes, however, that participation in the testing program is mandatory, since reference group norms must be based on the scores obtained by all members of the population or a representative sample of that population. Other possibilities exist, however. If the learning objectives and the standards of performance for French Immersion and English language programs are essentially the same, then it may be appropriate, statistically, to compare French Immersion group averages with English program averages. One of the variables that may need to be considered in choosing an appropriate reference group is the language of testing and the comparability of forms if separate tests are used for French Immersion and English program achievement testing, since the difficulty of the test affects examinee performance and, therefore, the validity of comparisons.

SUMMARY COMMENTS

Clarification of the issues described in this section is essential to ensure that the best possible assessment and reporting plan for students in French Immersion programs is implemented. The views of educators delivering these programs will be helpful in ensuring that the methods used to collect data and report on student achievement reflect the needs of students in these programs.

In Section 3 of this report, the Language of Testing Study conducted in 1989 is presented. This study addresses, in part, the effect that language of testing has on student achievement for students instructed in French Immersion programs.

Finally, the process of developing reliable and valid achievement tests is carefully outlined in a brochure titled *Developing Achievement Tests, Grades 3, 6, 9*. To obtain copies of this brochure, please call the Assistant Director, Achievement Test and Diagnostic Evaluation Program, Student Evaluation Branch, Alberta Education, at 427-2948.

SECTION 3
LANGUAGE OF TESTING STUDY
INTRODUCTION

In a report prepared for Alberta Education, Carey (1980) noted that when French Immersion students respond to test questions their responses are shaped not only by their levels of knowledge and skill proficiency (the attributes under study) but also by the nature of the test and by their ability to read that test. He argued that because their responses were shaped by these factors and these factors could vary depending on the language of testing, French Immersion students could achieve different scores depending on whether they wrote a test in French or in English.

Because of the importance of the issue "*In what language should French Immersion students be tested*", Student Evaluation staff undertook a study to determine what effect, if any, language of testing has on French Immersion students' responses to achievement test questions. This language of testing study examined how French Immersion students in grades 3 and 6 responded to social studies test questions presented in French and in English.

DESIGN OF THE STUDY

The language of testing study had two parts. The design of each part is presented separately below.

Grade 3 Social Studies

In June 1989, 654 grade 3 French Immersion students from three urban jurisdictions wrote the original English version or a French translation of the 1988 *Grade 3 Social Studies Achievement Test*. All students taking part in the study had received their grade 3 social studies instruction in French. On average, they had received approximately 75 per cent of their daily instruction in French in grade 3.

Staff from the Student Evaluation Branch used class lists from the participating schools to assign randomly French or English forms of the test to equal numbers of students in each classroom.

The returned test booklets were scored under the direction of Student Evaluation Branch staff. The results were then analyzed to determine if there were significant differences in the scores achieved by the two groups of students.

Grade 6 Social Studies

In June 1989, 416 grade 6 French Immersion students from two urban jurisdictions took part in the language of testing study. These students, who had all received their social studies instruction in French, wrote French or English forms of the regularly scheduled 1989 Grade 6 Social Studies Achievement Test -- Part A: Multiple Choice and Part B: Written Response. On average, they had received 65 per cent of their daily instruction in French in grade 6.

Staff from Student Evaluation used class lists from the participating schools to assign randomly French or English forms of the test to equal numbers of students in each classroom.

The returned answer sheets and written-response test booklets were scored under the direction of Student Evaluation staff. Special steps were taken to ensure that the same standards were applied to the marking of the French and English written-response questions. The results from the multiple-choice and written-response portions of the test were then analyzed to determine if there were significant differences in the scores achieved by the two groups of students.

RESULTS OF THE STUDY

Grade 3 Social Studies

Table 3-1 presents a comparison, by reporting category, of the scores obtained by the students who wrote the English and French forms of the 1988 Grade 3 Social Studies Achievement Test. The results indicate that the group of students who wrote the test in French (F group) achieved significantly lower scores than did those who wrote the test in English (E group) on all reporting categories except one. These results support the hypothesis that, in grade 3, the responses of French Immersion students are sensitive to the language of testing.

A comparison of E group scores to the 1988 provincial averages indicates that there are significant differences for all reporting categories. This suggests that the French Immersion students tested in this study had levels of social studies achievement that were lower than those achieved by students in regular English language programs in 1988. These results are surprising for two reasons. First, since the grade 3 students in this study came from the same or similar schools and jurisdictions as the students in the grade 6 portion of the study, and the levels of performance of the grade 6 students were higher than those achieved provincially, it seems logical that the grade 3 students should have achieved average or above average levels of performance. Second, these results do not compare to those from 1988, in which a group of grade 3 French Immersion students from a jurisdiction similar to those participating in this study achieved levels of performance that were better than those achieved provincially. In short, based on other evidence, one would have predicted that the grade 3 French Immersion students tested in this study would have had levels of achievement that were at or above the 1988 provincial levels.

Because these results were unexpected, an explanation of their underlying cause is warranted. Two possibilities come to mind. The first is that the levels of achievement of the students in this study are equal to the 1988 provincial levels but the E group scores do not accurately reflect this.

*Table 3-1
Grade 3 Language of Testing Study 1989
Results by Reporting Category
for Grade 3 Social Studies*

Reporting Category	Average ¹		Standard Deviation		F ²	Effect Size Relative to Prov. S.D. ³
	E Group	F Group	E Group	F Group		
Total Test	29.3 (33.8)	21.8	8.4	8.0	137.3***	.83
Topic A	9.2 (11.7)	8.3	3.0	3.0	17.2***	.27
Topic B	11.0 (11.4)	7.2	3.2	3.5	206.9***	1.18
Topic C	9.1 (10.7)	6.3	3.7	3.0	111.9***	.74
Knowledge and Comprehension						
All Topics	12.7 (14.8)	9.6	4.2	3.9	96.2***	.70
Topic A	3.7 (4.8)	3.7	1.7	1.7	0.9	0
Topic B	4.8 (5.3)	3.3	1.8	1.8	112.4***	.83
Topic C	4.1 (4.7)	2.5	1.9	1.6	131.2***	.83
Value Concepts and Valuing Skills (All Topics)						
All Topics	3.3 (4.0)	2.3	1.4	1.4	90.1***	.68
Inquiry Skills I (All Topics)	7.6 (8.7)	6.2	2.3	2.3	52.7***	.56
Inquiry Skills II (All Topics)	5.8 (6.3)	3.7	2.2	2.0	160.8***	1.0

***p. < .001

¹The bracketed figures below the E group averages are the 1988 grade 3 provincial averages.

²F refers to the variance ratio, with the appropriate degrees of freedom.

³This is derived from the ratio:
$$\frac{\text{Average score for E Group} - \text{average score for F Group}}{\text{Standard deviation of E Group Average}}$$

This systematic underestimation of grade 3 students' social studies knowledge and skill could have resulted from their inability to understand completely what the English versions of the questions were asking; if they could not understand the questions, they could not demonstrate what they knew and could do in social studies. This hypothesis is based on the assumption that because English language arts was not introduced to these students until their second or third year in school, their reading skills in English may have been insufficiently developed to permit them to comprehend the test questions. What limits this hypothesis is that the results of the 1989 Grade 3 English Language Arts Achievement Test administration do not support the assumption that the students in this study had limited English language arts skills. Those results, presented in the next chapter, show that the French Immersion students who wrote the *English Language Arts Achievement Test* in June 1989 achieved scores that were approximately equal to provincial averages. This suggests that, in general, grade 3 French Immersion students have English language arts skills that are adequate or better.

An alternative hypothesis to explain the unexpectedly low E group scores is that because 1989 was not a testing year for grade 3 social studies, the emphasis given to social studies instruction in particular, in the classrooms tested in this study, was less than in 1988. As a result, the levels of social studies achievement in these French Immersion classrooms in 1989 were lower than the provincial levels of achievement in 1988. This alternative hypothesis, while disturbing, could be valid since it suggests a pattern that has been argued to exist in English language programs. Ironically, its truth would be good news from a measurement point of view because then it could be assumed that the scores of the E group of students were accurate reflections of their actual levels of achievement. This in turn would indicate that it is possible to assess accurately the levels of achievement in grade 3 French Immersion classrooms using the achievement tests that are designed for regular English language students.

Grade 6 Social Studies

Table 3-2 presents a comparison of the total test, multiple-choice, and written-response scores achieved by the two groups of students who wrote the 1989 Grade 6 Social Studies Achievement Test. The results indicate that the group of students who wrote the French form of the test (F group) achieved significantly lower scores than did those who wrote the test in English (E group). This suggests that, when their level of social studies achievement is measured, the responses of grade 6 French Immersion students are sensitive to the language of testing.

A comparison of E group scores to provincial averages indicates that the French Immersion students tested in this study had scores that were significantly higher than those achieved by students in the regular English language program. When the F group scores are compared to those achieved by all grade 6 French Immersion students who wrote the provincial achievement test in French, no significant differences are found. These results suggest that the level of achievement of the students selected for this study is representative of French Immersion students generally and that this level is somewhat higher than that of students in the regular English language program.

*Table 3-2
Grade 6 Language of Testing Study, 1989
Results by Major Reporting Category
for Grade 6 Social Studies*

Reporting Category	Average ¹		Standard Deviation		F ²	Effect Size Relative to Prov. S.D. ³
	E Group	F Group	E Group	F Group		
Total Test	69.9 (62.5)	53.8 (55.3)	14.8	14.5	126.1***	.99
Part A: Multiple Choice	36.1 (32.2)	27.1 (27.8)	8.2	7.8	134.4***	1.00
Part B: Written Response	19.4 (17.5)	15.8 (16.4)	4.8	5.3	49.3***	.70

***p. < .001

¹The bracketed figures below the E group averages are the 1989 grade 6 provincial averages. The average scores of the French Immersion students who were not part of the language of testing study but who wrote the French translation of the 1989 Social Studies Achievement Test are presented in brackets below the F group scores.

²F refers to the variance ratio, with the appropriate degrees of freedom.

³This is derived from the ratio:

$$\frac{\text{Average score for E Group} - \text{average score for F Group}}{\text{Standard deviation of E Group Average}}$$

Table 3-3 presents a comparison of the scores achieved by the E and F groups for each of the reporting categories specified by the blueprint for Part A: Multiple Choice. The results indicate that F group students achieved significantly lower scores on all reporting categories than did the E group students, although the size of that difference was variable. This suggests that while the effect on scores of varying the language of testing is systematic across all reporting categories, the size of that effect is not constant.

A comparison of E group and provincial scores suggests that the French Immersion students tested in this study achieved levels of performance in social studies that were as good as or better than those achieved by most students in regular English language programs. At the same time, their levels of performance on items related to Topic C were significantly lower than those of students in French Immersion programs generally. This suggests that when it comes to their level of achievement in this area, the students in this study are somewhat unrepresentative of all French Immersion students in grade 6 social studies.

Table 3-3
Grade 6 Language of Testing Study 1989
Results for Part A: Multiple Choice
for Grade 6 Social Studies

Reporting Category	Average ¹		Standard Deviation		F ²	Effect Size Relative to Prov. S.D. ³
	E Group	F Group	E Group	F Group		
Topic A	13.2 (12.3)	10.8 (10.9)	2.7	2.8	77.1***	.81
Topic B	11.4 (9.7)	8.4 (8.3)	3.1	2.8	102.4***	.87
Topic C	11.6 (10.1)	7.8 (8.6)	3.5	3.8	110.6***	1.01
Knowledge and Comprehension						
All Topics	15.2 (14.0)	12.0 (12.2)	4.0	3.6	72.6***	.76
Topic A	5.7 (5.5)	5.0 (4.8)	1.3	1.2	33.8***	.51
Topic B	4.6 (4.0)	3.6 (3.6)	1.6	1.5	46.5***	.59
Topic C	4.8 (4.0)	3.4 (3.8)	2.1	2.2	44.0***	.65
Value Concepts and Valuing Skills (All Topics)						
All Topics	4.8 (4.1)	3.8 (4.0)	1.2	1.6	47.0***	.68
Inquiry Skills I (All Topics)	6.0 (5.2)	5.3 (5.3)	1.6	1.8	17.6***	.40
Inquiry Skills II (All Topics)	5.8 (5.3)	3.0 (3.1)	1.8	1.5	302.1***	1.45
Inquiry Skills III (All Topics)	4.4 (3.9)	3.0 (3.2)	1.3	1.5	107.3***	.92

***p. < .001

¹The bracketed figures below the E group averages are the 1989 grade 6 provincial averages. The average scores of the French Immersion students who were not part of the language of testing study but who wrote the French translation of the 1989 Social Studies Achievement Test are presented in brackets below the F group scores.

²F refers to the variance ratio, with the appropriate degrees of freedom.

³This is derived from the ratio: $\frac{\text{Average score for E Group} - \text{average score for F Group}}{\text{Standard deviation of E Group Average}}$

Table 3-4 presents a comparison of the averages achieved by the E and F groups for the two reporting categories specified by the blueprint for Part B: Written Response. The results are parallel to those for the multiple-choice portion of the test: while the language of testing effect favored those who wrote the test in English, the size of that effect varied across the two reporting categories.

*Table 3-4
Grade 6 Language of Testing Study 1989
Results for Part B: Written Response
for Grade 6 Social Studies*

Reporting Category	Average ¹		Standard Deviation		F ²	Effect Size Relative to Prov. S.D. ³
	E Group	F Group	E Group	F Group		
Short Answer	10.2 (9.3)	7.7 (8.1)	3.2	3.6	57.0***	.76
Composition	9.1 (8.2)	8.2 (8.2)	2.5	2.8	14.2***	.32

***p. < .001

¹The bracketed figures below the E group averages are the 1989 grade 6 provincial averages. The average scores of the French Immersion students who were not part of the language of testing study but who wrote the French translation of the 1989 Social Studies Achievement Test are presented in brackets below the F group scores.

²F refers to the variance ratio, with the appropriate degrees of freedom.

³This is derived from the ratio:

$$\frac{\text{Average score for E Group} - \text{average score for F Group}}{\text{Standard deviation of E Group Average}}$$

DISCUSSION

The purpose of the language of testing study was to determine if grades 3 and 6 French Immersion students achieved different scores depending on whether they wrote the social studies achievement tests in French or in English. The results indicate that the scores for students who wrote the French forms of the tests are consistently and significantly lower than are the scores achieved by students who wrote the test in English.

The E and F groups in both the grade 3 and the grade 6 portions of the study were composed of students randomly selected from the same French Immersion classrooms. Because of this random assignment of students to groups, it can be assumed that all conditions (e.g., levels of ability; quality of instruction) except the language of testing were the same across the two groups within each grade. It is therefore possible to attribute the significant differences in scores across groups to the only variable that changed -- the language in which the tests were presented and written.

What makes the language of testing effect statistically as well as educationally important is its magnitude. A variation in scores as large as that which occurred in this study has implications for data interpretation. This is most clearly seen when the 1989 grade 6 provincial averages are compared with the E and F group data. Because the E group scores are significantly higher than the provincial averages, a reasonable interpretation of the data is that the students in this study had levels of achievement that were higher than provincial levels.

From this it is possible to conclude that through French Immersion instruction, these students were able to acquire some proficiency in French while at the same time achieving levels of academic performance that were equal to or better than those in regular English language programs. This same interpretation and conclusion would not seem reasonable were it based on a comparison of the F group and provincial averages, because the F group scores are significantly lower. In short, because of the magnitude of the language of testing effect, two very different conclusions could be made about the levels of achievement of the same group of students.

Aside from the systematic depression of F group scores relative to E group scores, two other trends are notable in the data. First, the size of the differences in E and F group scores varies across reporting categories within each grade. Since items are grouped into reporting categories according to their content and cognitive levels, this trend suggests that the size of the language of testing effect varies depending on the nature of the questions being asked (i.e., on their topic and type). The second notable pattern is that the size of the differences in scores varies between similar reporting categories when comparing grade 3 results to grade 6 results. This pattern suggests that grade level may be a factor that contributes to how students respond to French and English forms of tests.

These results suggest that the findings from this study may not be generalized to other grade levels and subject areas except in general terms. In other words, while it can be assumed that the scores of French Immersion students may also be sensitive to the language of testing when their levels of achievement in grades 3, 6, and 9 mathematics and science are assessed, the magnitude or pattern of those differences cannot reasonably be predicted on the basis of these results. Further study therefore is needed before decisions are made about the language in which French Immersion students should write achievement tests.

SECTION 4

APPROPRIATE STANDARDS AND REFERENCE GROUPS FOR STUDENTS INSTRUCTED IN FRENCH IMMERSION PROGRAMS

As discussed earlier in this report, test results must be compared with expectations before they can be interpreted meaningfully. This section will present some information that relates to identifying appropriate expectations for students instructed in French Immersion programs.

An important question is whether students in French Immersion programs should be expected to achieve at the same level as students in the regular program. One line of reasoning might be that these students are dividing their academic attentions among five subjects (English language arts, French language arts, social studies, mathematics, and science) rather than four, and thus expectations should be less. Another line of reasoning might be that, since student participation in the program is at the parents' option, French Immersion students may come from a more supportive home environment on the average, and thus expectations for the group should be higher.

The actual performance of French Immersion students on achievement tests may shed some light on which expectations are reasonable. Unfortunately, before the 1989 achievement tests, students were not required to identify themselves by program, so analyses of achievement test data for earlier years are generally unavailable.

Some information is accessible. Related data for grade 3 social studies (1988) and for grade 6 (1988) and grade 9 (1986) language arts have been analyzed.

Grade 3 Social Studies 1988

In 1988, many grade 3 students in the French Immersion program in one large urban jurisdiction wrote the *Grade 3 Social Studies Achievement Test*, which was available only in English. In co-operation with that jurisdiction, the students were identified and their results were compared to other students in the jurisdiction. Total scores for students receiving instruction in French in French Immersion programs were significantly higher than for students in other programs in the jurisdiction, and also higher than for students in the English language program in the same schools. The results are shown in table 4-1. Interpreting these results must take into consideration that the students were instructed in French and tested in English. For this case, at least, there seems to be no reason to have lower expectations for the French Immersion program students.

Table 4-1
Grade 3 Social Studies 1988
Results for Students Receiving
Instruction in Either French or English¹

GROUP	Test Scores Average (out of 50)	Standard Deviation	Number of Students
Students Instructed in French in French Immersion Programs	34.8	7.7	264
Students Instructed in English in Schools offering French Immersion	33.1	8.6	332
Students Instructed in English in Schools NOT offering French Immersion	31.1***	9.0	3996

***Differences between the last group's average score (31.1) and the first two groups' average score (33.1, 34.8) are statistically significant ($p < .001$).
The difference between the first two groups is not statistically significant.

¹All schools were in the same jurisdiction.

Grade 6 French and English Language Arts 1988

Tests designed to measure achievement in Grade 6 English Language Arts and Grade 6 French Language Arts were administered in June 1988. Student participation in the French language arts test was at the option of the superintendent. A total of 1550 students completed both tests.

Because writing the French language arts test was optional, and because the nature of the language education and background of students can vary greatly, caution should be exercised in interpreting the results. It is not known which students were Francophone and which were French Immersion, but the majority of the students were in French Immersion programs.

Each test consisted of two parts: a 50-question multiple-choice reading section and a written-response section scored on five categorical scales. Both sections were scored under the supervision of the Student Evaluation Branch. For the purpose of calculating a total score for the test, each section was weighted 50 per cent, giving total scores out of 100.

It is of some interest to compare the results of these two tests and to compare the English achievement scores of students who took the French test with the average for all those who took the English test.

The correlation of total scores for the 1550 students who completed both tests was 0.623. This correlation indicates a fairly high degree of relationship between the skills and knowledge measured by the French test and those measured by the English test. The average score for students who wrote both tests was 68.9 per cent on the English test, compared to a provincial average of 62.5 per cent. These results show that students who took language arts in both languages achieved higher scores in English language arts than students who received instruction in English language arts only. It is not known, of course, whether the French Immersion students involved would have achieved higher or lower English language arts test scores if they had been in a regular program. Students whose parents chose to place them in French Immersion programs may on average have higher levels of ability than students in the regular program. There is, at any rate, no evidence that the English language arts skills of students taking both courses are on the average less than for students taking only one language arts course.

The higher scores on the English language arts achievement test for students who wrote both tests were consistent in both the reading (multiple choice) and writing sections of the test. Students who wrote both averaged 35.9 out of 50 on the reading section and 16.5 out of 25 on the writing section, compared to the provincial results of 32.4 on the reading section and 15.1 on the writing section.

Students who wrote both tests averaged 65.7 per cent on the French test. However, there is no way to make valid comparisons between the results of the tests in the two languages. Although the design of the tests was much the same, the content was completely different.

The total multiple-choice scores on the two tests had a correlation coefficient of 0.704, and the written-response scores had a correlation of 0.378, indicating that reading skills in the two languages, as measured by these tests, are more closely related than writing skills.

Grade 9 French and English Language Arts 1986

In 1986, French and English language arts tests were given to grade 9 students. Although a smaller number of students wrote both the French and English tests that year, the results are consistent with the grade 6 results for 1988. The correlation of total scores for the 615 students who completed both tests was 0.480. The average score was 73.8 per cent on the English test, compared to a provincial average of 64.0 per cent. Students who wrote both tests averaged 62.8 per cent on the French test.

The total multiple-choice scores on the two tests had a correlation coefficient of 0.593, and the written-response scores had a correlation of 0.260, supporting again the interpretation that reading skills in the two languages are more closely related than writing skills.

French Immersion students performed very well when tested in English Language Arts. This may simply be because they represent a select group of students. Clearly, receiving instruction in French has not seriously jeopardized their skills and understandings in language arts. What we cannot tell from this information is how well these students would have performed had they received instruction only in English.

ACHIEVEMENT TEST RESULTS 1989

On the 1989 achievement tests, students were required to identify their program as English, Francophone, French Immersion, or other. Thus, it was possible to perform analyses based on language of testing and program.

Grade 3 English Language Arts 1989

Superintendents decided whether grade 3 students in French Immersion programs would write the *English Language Arts Achievement Test*. Table 4-2 shows that about 70 per cent of French Immersion program students in grade 3 wrote the test.

Table 4-2
Grade 3 English Language Arts 1989
Number of French Immersion Students

Program	Number of Students	Percentage of Students
French Immersion Total Population	2805	100.0
Tested (in English)	1960	69.9
Not Tested	845	30.1

Grade 6 Social Studies 1989

Writing the *Grade 6 Social Studies Achievement Test* was also optional for French Immersion program classes. The test was available in English and in French translation. Table 4-3 shows the participation percentages.

*Table 4-3
Grade 6 Social Studies 1989
Number of French Immersion Students*

Program	Number of Students	Percentage of Students
French Immersion Total Population	1872	100.0
Tested in French	931	49.8
Tested in English	77	4.1
Participated in Special Study	437	23.3
Not Tested	427	22.8

The results for French Immersion students are complicated by the fact that 23.3 per cent participated in the special language of testing study. Of the remaining students, the majority wrote in French. About 4 per cent of the total wrote in English, and nearly 23 per cent did not write.

Grade 9 Science 1989

Table 4-4 gives the participation rates for the Grade 9 Science Achievement Test. Most French Immersion students wrote the French translation.

*Table 4-4
Grade 9 Science 1989
Number of French Immersion Students*

Program	Number of Students	Percentage of Students
French Immersion Total Population	1097	100.0
Tested in French	946	86.2
Tested in English	54	4.9
Not Tested	97	8.9

Tables 4-2, 4-3, and 4-4 reflect the difficulty in identifying any reference groups for appropriate norms. Participation rates are consistently lower than for the regular program. In no case can students be considered a representative sample.

Grades 3, 6, and 9 Achievement Test Results 1989

Table 4-5 presents the achievement test results for students in French Immersion and English language programs. As well, results are shown for the different test forms (English or French) that were written.

*Table 4-5
Number of Students and Average Scores 1989*

Grade/Subject	French Immersion Program		English Language Program Tested in English
	Tested in French	Tested in English	
<hr/>			
<i>Grade 3 English Language Arts</i>			
Number of Students	none	1 960	31 998
<i>Total Test Score</i> (Maximum Possible = 100)	-	69.9	68.9
<i>Part A: Writing Score</i> (Maximum Possible = 25)	-	16.5	16.2
<i>Part B: Reading Score</i> (Maximum Possible = 40)	-	29.5	29.1
<hr/>			
<i>Grade 6 Social Studies</i>			
Number of Students	931	77	29 918
<i>Total Test Score</i> (Maximum Possible = 100)	55.3	69.2	62.5
<i>Part A: Multiple-Choice Score</i> (Maximum Possible = 50)	27.9	35.6	32.2
<i>Part B: Written-Response Score</i> (Maximum Possible = 30)	16.3	19.5	17.5
<hr/>			
<i>Grade 9 Science</i>			
Number of Students	946	54	27 137
<i>Total Test (Grade 9 Science Achievement Test consists of multiple-choice questions only with a maximum possible raw score of 75.)</i>	50.1	52.0	50.1
<hr/>			

As the language of testing study and the preceding discussion indicate, comparisons of the average scores among groups are of low validity. Results for schools and jurisdictions can in some cases be usefully compared. The regular program norms are probably the most appropriate for schools and jurisdictions evaluating Grade 3 English Language Arts results. For students in the other grades, the results for those students who wrote in English can be compared to regular program students who wrote in English. A jurisdiction's French Immersion students who wrote in French can best be compared to all French Immersion students who wrote in French. Caution is necessary for all these comparisons, however.

This situation, with limited possibilities for comparison of test scores to norms, will persist as long as clearly defined reference groups are not available. The absence of an adequate data base remains problematic. In 1989, standards were not set for students instructed in French. This is due in part to the optional participation of students in the French assessment component of the program. The representativeness of this group cannot be assured. Evaluation of provincial levels of achievement must await the establishment of provincial standards for these tests and programs.

SECTION 5

CONCLUSIONS AND FUTURE DIRECTIONS

The purpose of this report is to focus attention on issues related to the assessment of French Immersion program student achievement. Issues have been described and trends in French Immersion achievement testing data have been discussed in the hope that such an analysis will provide insight into how the different issues can best be resolved.

For French Immersion students, the results of the 1989 language of testing study suggest that when it comes to the assessment of grades 3 and 6 social studies achievement, it matters in which language student achievement is measured: scores for French Immersion students who wrote in French are depressed relative to those achieved by students who wrote in English. What is not clear from the data is whether this pattern occurs across all grade levels and subject areas. When the scores of French Immersion students who wrote the 1989 Grade 9 Science Achievement Test in French are compared to the scores achieved by students who wrote the test in English, the size of the difference in average scores of the two groups is less than that which is present in the grade 6 language of testing study. This could mean that the language of testing effect is smaller in science than in social studies or that the effect is smaller at the grade 9 level. On the other hand, since the groups who wrote grade 9 science in French and English were self-selecting, it could simply be that the same language of testing effect exists in grade 9 science but that this effect is nullified by possible group differences in ability.

What is clear from the data is that the language of testing issue is important and deserves further study. It is also clear that the decision to test French Immersion program students in one language rather than another must take into account possible language of testing effects and how these effects can be controlled or accounted for when interpreting test data. In other words, for any data about French Immersion students to be useful and meaningful, decisions about the language of testing must be based on more than just political considerations. They must also consider what is statistically valid and reliable.

When considering appropriate standards and norms against which to compare student achievement in French Immersion programs, the language of testing results are also of interest. What can be concluded from that study is that where there is a depression of scores as a result of the language of testing, it is inappropriate to compare the results of the group writing in that language against a test score that has been chosen to represent the standard on the test of the other language. It appears to be equally inappropriate to compare the scores of groups writing in French with those achieved by students writing in English where there is a language of testing effect, unless that effect has been accounted for in some way.

The 1989 grade 3 English language arts results for French Immersion students and the 1989 grade 6 social studies and 1989 grade 9 science results for French Immersion students who wrote the tests in English are all equal to or above the English language program provincial averages. These results suggest that the French Immersion students who were tested had levels of performance that were as good as or better than those achieved provincially. From this it seems reasonable to infer that it is appropriate to expect at least the same levels of academic performance from students in French Immersion programs as are expected of students in the English language programs.

From the language of testing study it was not possible to determine why French Immersion students achieved lower scores when they wrote in French rather than in English. It has been hypothesized that two variables could account for such a difference. These two variables include the nature of the English and French forms of the test and the first and second language reading abilities of the examinees. Further information about the effect of these variables on student responses to test questions must be acquired before a decision can be made about the appropriateness of using French translations of English program achievement tests to assess French Immersion student achievement. Such a decision will also depend on whether or not it seems appropriate to apply the same expectations to French Immersion program students as are applied to English program students.

Whether or not testing should be mandatory for French Immersion program students will depend in part on how successfully some of these other issues can be resolved. The same caveat applies to the issue of whether or not the achievement of students instructed in French Immersion programs should be assessed and reported on a provincial basis. For these reasons, it seems important that further study and discussion of the issues and problems identified in this report should occur.

Student Evaluation staff are addressing the issues raised in this report in two ways in 1990. First, a plan for ensuring a full discussion of the issues presented is being prepared. A cross section of educators involved in the delivery of French Immersion programs will be involved in these discussions. Second, a follow-up language of testing study at the grades 3 and 6 levels in 1990 is being planned. As well, grade 9 students will write the French and English Language Arts achievement tests. Information collected from these activities will assist in further resolving the issues presented. Other activities may be planned following in-depth discussions with educators in French Immersion programs.