DOCUMENT RESUME

ED 280 270 FL 016 507

TITLE French Immersion Studies, Year 3 (1985-86). Tests of (English) Reading Skills.
INSTITUTION York Region Board of Education, Aurora (Ontario).
PUB DATE Jan 86
NOTE 27p.; For a related document, see FL 016 508.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Comparative Analysis; Elementary Education; *English; Foreign Countries; French; Grade 3; Grade 4; Immersion Programs; Language Tests; *Program Effectiveness; Reading Comprehension; *Reading Skills; Reading Tests; Second Language Programs; Test Format; *Test Results; Test Theory

ABSTRACT

To determine whether students enrolled in one Ontario region's early French immersion (FI) programs developed English reading skills comparable to their non-FI peers, a monitoring process was begun in the first FI program year (grade 3) in which formal English instruction is given. The FI cohort and a control group matched for mental abilities and communities were administered reading tests in third and fourth grade. As predicted, the FI students performed below the control group on the first test but attained scores that were at least equal in the fourth grade. A test of inference and generalization administered in fifth grade to the two groups showed consistently but marginally superior scores in all skill areas for the FI group. Item analysis and examination of subgroup performance indicated some areas for improvement or future investigation, including use of the more advanced test at the fourth and seventh grade levels and expansion of the language testing to include vocabulary, punctuation, grammar, and capitalization. A supplement that reports and compares the English reading comprehension scores of FI students and non-FI students is appended. (Author/MSE)

FRENCH IMMERSIONS STUDIES, YEAR 3 (1985-86)

TESTS OF (ENGLISH) READING SKILLS

Research Department

Division of Planning
and Development

THE YORK REGION BOARD OF EDUCATION

JANUARY 1986

FRENCH IMMERSION STUDIES, YEAR 3 (1985-86)
TESTS OF (ENGLISH) READING SKILLS

A B S T R A C T

To determine whether students enrolled in the Early French Immersion (FI) programs develop English reading skills comparable to their non-FI peers, a monitoring progress was begun in the first year (grade 3) that the FI students had formal English language instruction. A control group, matched on mental abilities and communities, was also tested at that time, April 1984, using a modified, multiple-choice CLOZE methodology. In April 1985, the two cohorts were again tested on their literal comprehension skills. As postulated, the FI students, on average, scored significantly below their peers on the first test then, in the second year (grade 4), attained scores that were at least equal.

In October 1985, the study group (183 FI, 196 non-FI students) was again tested on their English reading skills. But this time the test (CTBS) went beyond factual comprehension skills to include inference and generalization abilities. The FI cohort, on average, attained consistently, if marginally, superior scores on all three major skill areas, but the differences were not statistically significant. Both cohorts scored above the national norm and at a level consistent with their mental ability as measured in 1983.

By item analysis and examination of the performance of sub-groups within the cohorts, certain areas for improvement or for future investigation were identified. Among the recommendations for staff study with respect to further monitoring: (1) analysis of the 1985 CTBS reading scores attained by FI students at the grade 4 level; (2) similar analysis of CTBS reading scores at the grade 7 level beginning in 1987; (3) expansion of the language testing program to include vocabulary, punctuation, language usage (grammar), and capitalization.

FRENCH IMMERSION STUDIES, YEAR 3 (1985-86)
TESTS OF (ENGLISH) READING SKILLS

C O N T E N T S

TABLES AND FIGURE

APPENDICES

Research Office
Division of Planning
and Development

January 1986

FRENCH IMMERSION STUDIES, YEAR 3 (1985-86)
TESTS OF (ENGLISH) READING SKILLS
(By the Canadian Tests of Basic Skills, Reading Comprehension test)

Purpose of the study and recap of Year 1 and 2 findings

This is the third year of a monitoring study of the progress in English reading (comprehension) skills of the board's initial cohort of "early French immersion" (FI) students. A "matched" cohort of non-FI students chosen as a local comparison or "control" group has been followed through the study period with the same tests. The study began in 1983-84 when the students were in grade 3.

A fuller explanation of the purpose and methods of this monitoring can be found in Part C of Early French Immersion, Three Evaluation Studies (August 1984) received by the Standing Committee on Program Policy and Program Management on August 20, 1984, and in the subsequent French Immersion Studies, Year 2 (1984-85) received by the PP & PM Committee on July 2, 1985. In summary, there are two questions of concern to trustees, staff and the parents of FI students:

(1) How do the English reading comprehension skills of FI and non-FI students compare after the FI students begin formal instruction in English at school? (Note: for the FI cohort under study, English-language instruction began in grade 3. Currently, such instruction begins in grade 4.)

(2) Apart from how the test results of FI and non-FI York Region students compare with each other, how well do they both compare with external standards? (In Years 1 and 2 this question involved levels of "mastery", as defined in the previous reports.)

These expectations were held at the beginning of the monitoring:

(1) Initially (grade 3) the non-FI students would out-perform the FI students, even though the groups were matched on mental aptitude and came from the same communities;

(2) Subsequent to their first year of instruction in English, the FI students' scores would "catch up" to the non FI's, then the FI cohort would gradually surpass the control group's reading performance;

(3) As both of the study cohorts initially scored somewhat higher than the York Regional mean average mental abilities test score, their reading comprehension scores would be at least equal to the regional and national norm.

In Years 1 and 2, reading comprehension ability was measured by a locally-developed version of a "multiple choice modified" cloze test battery, formatted as a game. The Cloze test mainly measures literal (factual) comprehension. The reading selections had previously been field-tested and normed in Ontario by the Ontario Institute for Studies in Education. Full item analysis data were available and the grade level difficulty indices established in the field trials were reconfirmed by a panel of YRBE primary teachers who rated the selections under consideration. Further details can be found in the Year 1 report and the test scores can be obtained for review from the Research Department.

On the grade 3 Cloze tests (April 1984), the non-FI students on average scored significantly higher and a greater percentage of non-FI children scored at or above the "mastery" (M) level. However, results at the highest scoring FI school were comparable to the lowest of the non-FI schools and average FI scores at that school were just short of the M level (82% rather than 84%).

On the grade 4 Cloze tests (April 1985), the FI students scored higher on average (51.5 items correct out of 60) than the non-FI cohort (50.8). Although this difference was neither statistically significant nor of practical significance, it indicated that any gap in literal comprehension abilities had been closed. In addition, in this second year a higher percentage of FI children than non-FI attained the M level (72% vs. 63%). This also should be taken as a sign that the gap had closed, rather than as a conclusive indication that the FI cohort had significantly surpassed their control group.

In both years, both FI and non-FI groups attained average scores higher than the norming population on the same stories. This was in keeping with expectations based on their higher than average (grade 3) mental abilities test results.

In Year 2 (1984-85), when the study groups were in grades 4, the students were tested with the (English) Reading comprehension battery of the Canadian Tests of Basic Skills (CTBS). This test, administered in October, assesses abilities to make inferences and generalizations as well as literal (factual) comprehension. It was anticipated that the grade 4 administrations would lead to the equation of Cloze and CTBS scores and permit tracing thereafter from grade 4 through the other grades in which CTBS is routinely used (grades 7, 10, 12).

Due to the problems with the test scoring facility, the 1984 grade 4 CTBS testing program did not produce usable reading scores. Although the opportunity to equate CTBS and Cloze results in the same school year was thus lost, it was decided to test the study groups with the CTBS reading comprehension battery in October of the grade 5 year and to use those tests. The grade 4 Cloze testing was conducted in April and thus there were still six months between tests. This time the CTBS followed rather than preceded the Cloze testing. The present report will deal with the "equating" exercise and its findings and conclusions.

Changes in the Year 3 study group

Long-term school-based "field research" differs from "laboratory research" in the manner in which it deals with the inevitable study population changes that occur over the years. Families move out of and into the schools in the study; some children transfer to programs for the gifted; some children change from the immersion program; some few are not advanced to grade 5. "Laboratory research" tends to follow only the "survivors" and the original study population shrinks. The present study accepted the volatile nature of York Region's growth and opted to accommodate changes in class compositions and even where the children were located (e.g., Jefferson P.S. FI students relocated in Beverley Acres P.S. in 1984-85; many Woodland P.S. FI students relocated in Dickson Hill P.S. in 1985-86).

Either method of dealing with change presents threats to the validity of longitudinal studies (especially those factors ominously known as "selection" and "mortality"). While there is no practical way of totally avoiding such threats, they can be minimized. In this study, the situation has been treated by the addition of students in the attempt to keep the groups matched on IQ and community. (See the Year 2 report for some details).

The original study group totalled 262 students (including 179 FI children), the 1984-85 population totalled 352 (FI=189) and the 1985-86 study group grew to 379 (FI=183). The major reasons for increases in Years 2 and 3 were the addition of "gifted" and other non-FI students to retain the IQ equivalencies (the non-FI group was of slightly lower IQ in Year 1, slightly higher in Year 2) and to include children of the same grade in the "new" locations of the FI programs).

As of Year 3, we cannot assert that the IQ scores of the two cohorts are "matched," only that they were as of Year 1 when the testing was done and that efforts have been made to maintain equivalencies. These IQ scores are not invariant over time. Thus we can only say that if they are no longer matched, then the programs themselves (among other factors) may account for subsequent changes. The original study design did not call for another administration of a mental abilities test. This might well be considered for, say, October 1987, when this study group is due to sit for the CTBS batteries. The purpose of mental abilities testing at that time would not be to reshuffle the cohorts in order to "match" on IQ, but to explore possible relationships between French immersion programming and mental abilities over time.

Findings, Year 3

The objectives of the October 1985 administration of the CTBS Reading Comprehension test in this study population were:

(1) To determine whether the observed scores of the FI and non-FI cohorts bore the same relationship to each other as the grade 4 Cloze literal comprehension test scores (Note: the CTBS reading test covers inference and generalization as well as factual -- literal -- comprehension skills);

(2) To determine how well the study population's score compared with the national norm (Note: YRBE regional normative data does not exist for grade 5 as testing is not normally done at this level);

(3) To determine whether there were differences in the performances of the cohorts on factual, inferential, or generalization skills.

Inasmuch as there was little "school" time between the April 1985 Cloze testing and the October 1985 CTBS testing, it was postulated that there would be no significant change in the relative position of the cohorts, i.e., the FI students would have slightly higher scores than the non-FI cohort. However, since there had not previously been substantial testing of inference or generalization skills, it was thought possible that significant differences might be found in these reading abilities.

FI and non-FI performance levels

How well did the FI and non-FI students compare with each other and with the national norm? Table 1 arrays the average performance of these cohorts in the eight schools that house the study population (Note: three schools, D, E and G, have no FI program; school H has no non-FI grade 5 students; and one school, E, houses a program for 19 gifted grade 5 students who come from the neighbourhood and nearby communities).

For the 379 students tested (96 per cent of those in grade 5 in the study schools, the remainder being absent during the test), the mean average score on the 54 items on the CTBS reading comprehension battery of eight "stories" (representing grade 3 through to early grade 7 prose selections) was 31.4. The national norm performance (autumn administration) is 29.0 items correct. On average, the study group is reading at the 5.4 grade equivalent, that is to say, about two months beyond the national norm (5.2) for grade 5 students on an autumn administration.

Table 1: By-school, by-program, average performance (mean average raw score, "building grade equivalent" and "building" percentile rank) on the CTBS Reading test Form F, Level 11 (grade 5), Autumn 1985

	All* Grade 5s (N=379)			Grade 5 FIs (N=183)			Grade 5 Non-FIs (N=196)		
School	Mean	BGE**	%ile	Mean	BGE	%ile	Mean	BGE	%ile
"A"	30.1	5.3	60	30.6	5.4	68	28.2	5.1	45
"B"	31.8	5.5	75	32.8	5.6	82	27.7	5.1	45
"C"	31.3	5.4	68	33.1	5.6	82	28.0	5.1	45
"D"	28.8	5.2	52	-	-	-	28.8	5.2	52
"E"***	37.0	6.1	99	-	-	-	37.0	6.1	99
"F"	29.9	5.3	60	31.0	5.4	68	28.6	5.2	52
"G"	30.9	5.4	68	-	-	-	30.9	5.4	68
"H"	31.2	5.4	68	31.2	5.4	68	-	-	-
TOTALS	31.4			31.9			31.1		
TOTALS EXCLUDING "E"	30.6			31.9			29.1		

* That is, of the eight schools in this study. Note that where a school has both a French Immersion and Non-FI cohort, each has been treated as a separate "building". Composite building data are also supplied (in the "All" column).

** The GE permits comparison of a student's score with the normative attainment of grade peers; the "building grade equivalent" (BGE), which is reported as a percentile, permits comparison of a school grade average relative to the averages attained in other schools. The BGE relates to norms for schools rather than for individuals. School and individual norms may differ markedly, most noticeably when school performance is much above or below average. For example, an individual score of x may translate to a GE equivalent to the 65th percentile but a school average score of x may translate to the 95th percentile among schools. If further clarification is needed, consult the CTBS Manual for Administrators or the Research or Testing Offices.

*** Includes 19 students in the program for the gifted (38% of the grade 5 cohort in this school). These gifted students averaged 41.3 correct responses (average GE = 6.6).

For an individual, two months roughly corresponds to the standard error of measurement and might be considered a chance variation from "true score." But for a large group, this 0.2 GE or a 2.4 mean raw score difference can be said to represent a performance substantially above the mean. We believe the study group to be of high average or above-average mental ability. The observed mean score is consistent with expectations of results at least equal to the national norm since they are significantly higher.

To see how well the students in each school collectively compare to national "building" norms, we can look at the "BGE" and "%ile"* sub-columns in the "All" column. Here we note the lowest performance: one school is at the mean (actually 28.8) and at the median (actually 52%ile). At the other extreme is one school, E, where performance reflects the presence not only of gifted students (average score = 41.3), but of other very proficient readers (average score = 34.5 items correct). The other schools are in the high average (e.g., 60%ile) to above-average (the 68%ile) to the high range (75%ile).

Since school E is so atypical, separate calculations were made excluding these students. As might be expected, the weighted average of school E (37.0 correct responses, on average) makes a difference (0.8 items), but the overall performance (30.6) of the remaining study population is still well above the national mean (29.0).

By comparison, the FI cohort scored on average 0.8 more items correct than all the non-FI students, including the gifted (31.9 vs. 31.1). This October result is consistent with the Cloze testing findings of the previous April, namely, a small but scarcely significant difference in favour of the FI students.

It should be noted that in the four schools where both FI and non-FI grade 5 students are housed, the FI performance was markedly superior. However, it must be remembered that the FI students come from larger catchment areas than their non-FI schoolmates. Moreover, the FI students initially showed higher IQ scores than the non-FI students in the same schools even though, overall, the average mental aptitude scores of the cohorts (as measured by the Otis-Lennon Mental Abilities Test in mid-grade 3) were virtually equivalent.

---

* "%ile," percentile rank, indicates the relative standing of a student (or group of students) relative to other students (as individuals or groups). The percentile rank tells the per cent of students in the reference group who obtain lower scores. Hence if a student earns a percentile rank of 50, half of the other students earned a lower score. A percentile rank of 70 indicates that 70 per cent of the reference group earned lower scores. Percentile ranks range in magnitude from 0 to 99.

Notwithstanding this observation, it is important for teachers of FI and non-FI classes in the schools where both programs are housed together to recognize differences in performance and ability levels and to respond appropriately in their instructional programs.

Analysis of performance differences

In addition to the previously mentioned small difference in average number of correct items (0.8, favouring the FI students), there were other differences between the two cohorts of this study group. Before describing these, it should be noted that the differences within each cohort are greater than the differences between the FI and non-FI sub-populations. The between-cohort differences appear to be systematic, if not substantial, and fairly consistent in favouring the FI students, though not invariably, as we shall see.

These differences were found by what is called "item analyses" and involved an examination of the performance of all 379 students in the study group on each of the 54 items that constitute the Reading Comprehension battery's eight reading passages. The percentage of "correct" responses for each item was calculated for both the FI cohort and the non-FI students. Their combined statistics produced an average "success rate" for the total YRBE study group. Comparisons were then possible among the two sub-groups, the total study group and the national norm.

With a view to determining whether there were significant performance differences between the FI and non-FI cohorts and between the total YRBE group (thought to be somewhat above average) and the national norm group, comparisons yielded the following information.

(1) Differences on grade 5 and grade 6 reading passages.

The FI students were more successful than their non-FI comparison group on reading passages 6 and 7, items 60 through 75 inclusive, which appear to be at the late grade 5 to early or mid-grade 6 level, but only by about seven per cent more correct responses on average.

The total YRBE group performed more successfully on these stories than the national norm population. The differences on an item-by-item basis were not great, typically about eight per cent, but mainly favoured the YRBE students.

These findings are consistent with the overall results and expectations that the YRBE would at least match the national norm group.

(2) Differences in the grade 3 to mid-grade 5 reading passages

The first five reading passages (items no. 30 through 59 inclusive) include stories at the late-grade 3 or early grade 4 level through to about mid-grade 5. Average grade 5 readers would not have much difficulty with most items in these passages, but they include some very difficult items which test abilities to see relationships between facts, to infer cause and effect, and to apply information through generalization, etc.

An analysis of performance on items in the first five passages, items that the great majority of students would attempt, seemed a potentially useful way of determining differences in the reading skills in FI and non-FI students. When, by inspection, it was found that about 85 per cent of all test-takers actually attempted up to item 64 (middle of the sixth reading passage, about late grade 5 difficulty level), it was decided to extend the analysis to include the first 35 test items. At the same time it was decided to exclude the "easiest" 17 items therein. Since these 17 were answered correctly by at least two of every three test-takers, the differences between the cohorts were essentially meaningless at such high success rates (e.g., when one cohort achieves 94 per cent correct on an item and the other achieves 96 per cent, random error is as likely as any explanation of the apparent difference.)

So the analysis narrowed to the 18 "hardest" of the first 35 test items. The skills that each of the 18 items purported to test were identified. (A description of the skills tested in the CTBS Reading Comprehension battery is appended as "Skills Objectives.") Table 2 shows how the FI and non-FI cohorts performed and also how the total group performed. The national norm for each item is displayed. The items are grouped by skill.

Inspection, item by item, shows that the total YRBE grade 5 group outperformed the national norm on all items except 35* and 39*. And the FI sub-group scored, on average, more correct responses than the non-FI students on all items except number 49. The

* Item 35 has two logical and perhaps equally plausible responses. Perhaps this is why 43 per cent of the national norm group and 42 per cent of YRBE gifted students selected the "wrong" answer. Does this item selectively discriminate against students capable of high-order logical operations? This anomaly has been brought to the attention of the test publishers and they are examining this situation. Item 39 is keyed as skill F2 ("...understand factual details relating to classification"). The item might also be considered as inference skill I2 ("...draw conclusions from information and relationships") since it requires understanding of the anatomical relationship of ankles to legs and the lexical relationship of "swelled" to "puffy". The skills categorizations, some contend, are artificial, (not uncommonly) arbitrary and ought not to be used.

Table 2: Performance on 18 difficult items of the CTBS Reading Comprehension Test, Form 6, Level 11 (Grade 5) Fall Administration, 1985

		Correct Response Rate (%)			
			** York Region Performance		
*Skill	Item#	National Norm	Combined (N=379)	FI(N=183)	Non-FI(N=196)
F1	45	45%	48%	56%	41%
F1	60	34	47	53	42
F1	64	41	54	56	52
F2	39	70	65	67	62
F3	48	40	48	50	47
F3	49	48	51	50	52
F3	55	54	66	69	63
F3	63	43	50	52	48
I1	35	57	51	53	49
I1	53	60	63	68	58
I3	31	57	65	69	60
I3	51	39	54	60	49
I4	36	54	57	61	52
G1	50	38	40	43	38
G3	58	50	59	60	59
G3	59	45	54	61	46
G6	61	43	47	51	42
G6	62	44	45	48	41

* See the "Skills Objectives" sheet for descriptions of these skills

** For 379 students in eight schools

differences are not always substantial, but they are consistent in confirming the relative performances predicted by the Cloze tests in April 1985 and the mental abilities testing in 1983. That is, they give a slight edge in literal (factual) comprehension to the FI cohort and show the total YRBE group to be clearly above the norm group in performance on almost all items in this "F" skill area.

(3) Differences on the grade 7 reading passage

The eighth reading passage (items no. 76 through 83 inclusive) appears to be at the early grade 7 level, more or less, and its eight items include all three of the major skill areas. The grade 5 national norm population average correct response rate for the eight items (autumn administration) is only 31 per cent. Not only is the reading selection demanding, but many students do not reach these items until the test time (40 minutes) is almost expired. The test publisher's agent states that students typically do reach the end of the test, but this is not our experience.

As we can see from Table 3 (a), neither the FI nor non-FI cohort fared as well as the national norm on the majority of these questions, including the very last three items. The YRBE group, on average, scored below the norm group on all three "generalization" skill questions, but were more-or-less comparable on the four "factual" and one "inferential" items. (NOTE: Administrations of the CTBS at other grade levels have shown below-average "generalizations" reading skills among YRBE students).

It would appear, but is difficult to prove, that the cause of this lower performance by the YRBE group is the failure of about two in five students to complete the test. Some 82 of the 183 FI students and 76 of 196 non-FI students (42 per cent overall) did not provide an answer to the last item. In fact, some 65 FI and 60 non-FI students (33 % overall) did not answer any of the last eight items. This probably accounts for the below-norm performance and for the relatively superior showing of the non-FI cohort (29.7 % correct on average for the last eight items vs. 28.0% for the FI). Tables 3(b) and 3 (c) provide details.

The failure of so many to respond to the last passage can be traced, at least in part, to one error that apparently occurred in four classes. Some or all of the students in these classes filled in solidly the "bubbles" on the answer sheet when signifying their answers. The instructions for this test administration call for just a line to be drawn through the

Table 3(a): Average correct response rate (%)
by item for the last reading passage of
The CTBS READING COMPREHENSION TEST
Level 11, Grade 5, Autumn 1985

			** York Region Performance		
*Skill	Item#	National Norm	Combined (N=379)	FI(N=183)	Non-FI(N=196)
I2	83	19	17.7	16.4	18.9
G2	82	28	19.5	18.6	20.4
G3	81	25	20.3	19.7	20.9
F1	80	19	25.1	23.4	26.5
F2	79	30	33.2	33.3	33.2
G2	78	48	45.1	43.7	46.4
F1	77	45	38.3	41.0	35.7
F3	76	32	31.7	27.9	35.2
Avg. correct response rate		31	29	28	30

Items 76-83 are based on the last reading selection in the test

* See the "Skills Objectives" sheet for descriptions of these skills

** For 379 students in eight schools

Table 3(b): Average correct response rate (%)
by item for the last reading passage of
THE CTBS READING COMPREHENSION TEST
Grade 5, French Immersion Students (N=183)

	York Region* FI average performance					
School Code =	A	B	C	F	H	By-item %
No. of students) =	(26)	(41)	(27)	(44)	(45)	
Item #						
83	15.4	12.2	14.8	13.6	24.4	16.4
82	23.1	22.0	22.2	11.4	17.8	18.6
81	15.4	17.1	29.6	20.5	17.8	19.7
80	23.1	22.0	11.1	22.7	33.3	23.4
79	26.9	43.9	18.5	25.0	44.4	33.3
78	50.0	56.1	29.6	31.8	48.9	43.7
77	38.5	46.3	33.3	34.1	48.9	41.0
76	30.8	29.3	22.2	20.5	35.6	27.9

Avg. correct response rate for this story = 28.0%

*Total grade 5 FI cohort except for students absent during testing

Table 3(c): Average correct response rate (%)
by item for the last reading passage of
THE CTBS READING COMPREHENSION TEST
Grade 5, Non-French Immersion Students (N=196)

	York Region* non-FI average performance							
School Code =	A	B	C	D	E	F	G	By-item %
No. of students) =	(6)	(19)	(14)	(30)	(50)	(38)	(39)	(196)
Item #								
83	0.0	15.8	21.4	20.0	32.0	10.5	12.8	18.9
82	0.0	10.5	14.3	36.7	18.0	18.4	23.1	20.4
81	0.0	15.8	28.6	16.7	28.0	21.1	17.9	20.9
80	0.0	10.5	21.4	30.0	40.0	23.7	23.1	26.5
79	16.7	10.5	14.3	36.7	54.0	28.9	28.2	33.2
78	33.3	31.6	42.9	40.0	74.0	28.9	43.6	46.4
77	16.7	31.6	57.1	30.0	44.0	28.9	33.3	35.7
76	16.7	15.8	35.7	33.3	40.0	26.3	25.6	35.2

Avg. correct response rate for this story = 29.7%

I.e., for these seven schools only

appropriate answer "bubble" when making a choice among alternative answers. It is hard to say just how much time the extra work took. From trials in the Research and Testing Offices it is estimated that at least one and a quarter minutes -- and probably two minutes or more -- were lost by this error. Quite possibly this would translate to two or more items that might have been attempted.

Tables 3(b) and 3(c) were compiled in an attempt to see among-school differences within the FI and non-FI cohorts. Clearly, the within cohort differences are greater than the between cohort differences. This probably traces to two main factors, (1) the administration error mentioned above and (2) the presence (in school (E) of a large number (19) of gifted grade 5 students. These inconclusive tabulations may be of most value to staff in the schools, as they will be in the best position to know (from the answer sheets) the probable impact of the administration error on their students' performance.

Summary of performance differences

(1) Both the FI and non-FI cohorts performed on average above the national norm population on each of the three major skill areas of the CTBS Reading Comprehension battery. This was anticipated on the basis of previous attainment on a mental abilities test and on tests of literal comprehension.

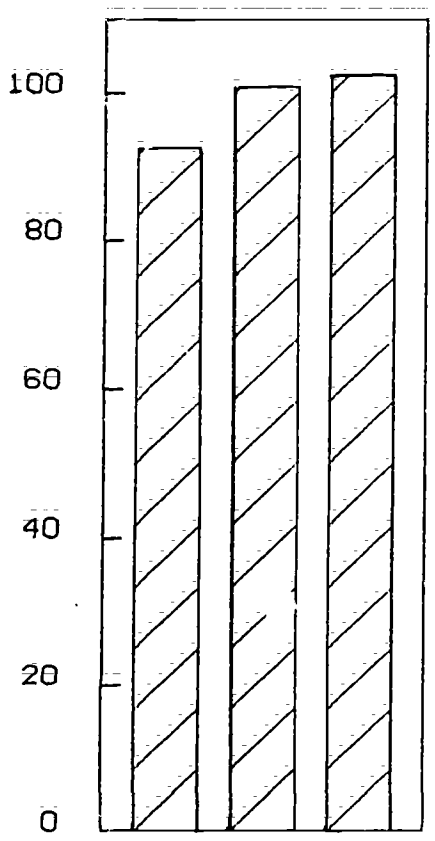
(2) The FI cohort's average performance was slightly but consistently above that of the non-FI students on the "factual" (literal comprehension) skill test items, as was postulated from the grade 3 and grade 4 Cloze test results.

(3) Many YRBE students in each cohort did not "complete the test" (provide answers to latter questions). This may in part trace to errors in administration, but the failure of about 15 per cent of students to provide answers beyond the 35th item (of 54 in the battery) almost certainly traces to other difficulties.

(4) On the "inference" and "generalization" skill items the FI cohort on average equalled or did better than their non-FI peers for the first seven of eight stories.

The relative performances on the two groups over the duration of the study and the three testing periods is illustrated in Figure 1 (following page). The FI students' average performance is shown as a percentage of the non-FI students' average score. The latter has been given a value of 100 at each testing period. (This avoids scaling anomalies arising from the varying number of questions on the tests.)

FIGURE 1: FI students' average reading score as a percentage of non-FI students' score, 1984-1985



Date:	Spr. 84	Spr. 85	Aut. 85
Grade:	3	4	5
Test:	C L O Z E		CTBS
# items:	50	60	54
Skills	F	F	F,I,G

Conclusions and recommendations

(1) Despite some threats to the validity of this exercise, it would appear that the grade 5 CTBS Reading Comprehension battery produces literal ("factual") comprehension results comparable to results attained through use of Cloze tests in grade 3 and grade 4 with essentially the same population. The grade 5 and subsequent CTBS reading results (i.e., in grades 7, 10, 12) may therefore provide a continuing yardstick for estimating the relative performance of these study cohorts.

R1. It is recommended that the grade 7 CTBS Reading Comprehension scores, autumn 1987, for these FI and non-FI students be examined to determine their reading achievement levels both with respect to each other and to the national norm population.

R2. It is also recommended that the grade 4 CTBS Reading Comprehension and Mathematics test scores, autumn 1985, for the FI students be examined and compared with those of YRBE's non-FI grade 4 students and also with the national norm population. Depending on findings, it may be desirable to compare FI students' scores with those attained by students in programs for the gifted and those attained by "non-gifted" and non-FI students in schools housing gifted plus mainstream programs.

(2) While, on average, the performance levels of both the non-FI and FI students are what might be expected relative to the norm group, it appears that many YRBE students are not proceeding successfully as far into the Reading Comprehension battery as the norm population. The "mechanical error" (the filling in of the answer bubble referred to in this report) has been drawn to the attention of participating schools and even greater efforts will be made to avoid this sort of problem in the future. Beyond this, about 15 per cent of students stop answering before proceeding two-thirds of the way through the test. There are several possible reasons for this, including being tested beyond their ability level.

R3. It is recommended that the Research and Testing Officers study this situation together with school staff, the test publishers, et al as appropriate, and prepare a report (with recommendations, if indicated). It is intended that this study include the current CTBS results and, if necessary, the autumn 1986 CTBS response patterns, too.

(3) While the FI cohort's average performance is presently only marginally, though rather consistently, better than the non-FI students' achievement, it is anticipated, on the basis of studies elsewhere, that the difference will increase over time. However, future comparisons may not be fruitful for two reasons: (1) the within group differences are already much greater than between group differences and (2) it is the nature of multilevel standardized tests to appear to create even greater differences (in grade equivalent scores, for example) over time. The former situation leads to spurious "no significant difference" findings; the latter leads to apparently substantial differences that are (in part) merely artifacts of test scaling.

R4. It is recommended that the Research and Testing Officers explore these situations with respect to CTBS test scores for the current longitudinal study group and subsequently advise on means of obtaining reliable and relevant data respecting achievement differences.

(4) Research with the YRBE's initial French immersion cohort should lay to rest the earlier concerns about their ability to comprehend written English at a level commensurate with their ability. This is not to say that the FI students have yet demonstrated appropriate mastery of other aspects of language (e.g., breadth of vocabulary, punctuation, spelling). If recommendations 1 and 2 (above) are acted upon, some further indications of the reading and vocabulary skill levels of the FI students will emerge. However, the present region-wide standardized testing program does not include the spelling, punctuation, language usage (grammar) or capitalization tests. Concurrent with the present study, the grade 5 FI students in one school took the CTBS spelling test. Their performance almost exactly matched the national norm. This "spot check" is comforting to parents or others who expressed concerns at the time that FI was initiated. The results for this one school are not, however, consistent with the mental abilities and reading comprehension scores (which were somewhat above the norm). Of course, nothing is "proved" by a one-shot test.

R5. It is recommended that the Research and Testing Officers, together with FI school staff, consider what other language testing activities be undertaken (possibly as pilot studies) on the basis of perceived needs of professional staff and concerns that parents may still have.

(5) The original study design also called for monitoring of the French language skills of the FI students during the year that instruction in English began (i.e., grade 3, 1983-84, for this cohort). The results of French reading comprehension testing showed an acceptable performance level (close to the norm population, even though the norm group had one more year of French immersion than the YRBE group). During the recent English language testing cycle, two FI principals requested that consideration be given to replicating the previous French reading comprehension and to expanding the testing in French (not necessarily at the expense of existing English language testing program).

R6. It is recommended that the Research Officer consult with FI staff, with supervisory officers, and with other interested parties on the need to collect additional data on the achievement levels (including diagnostic data) of FI students; determine the feasibility of collecting such data; and report to the Superintendent of Planning and Development on the findings of these enquiries.

Appendix A: Test R: Reading Skills Objectives

Test R: Reading
SKILLS OBJECTIVES

F **Facts:** To Recognize and Understand Stated Factual Details and Relationships (Literal Meaning)

- F1 Description: To understand factual details relating to description of people, places, objects, and events
- F2 Categorization: To understand factual details relating to classification
- F3 Relationships: To understand functional relationships, time, and sequence
- F4 Contextual Meaning: To deduce the meanings of words or phrases from context

I **Inferences:** To Infer Underlying Relationships (Interpretative Meaning)

- I1 Cause and Effect: To understand cause, effect, and interaction
- I2 Draw Conclusions: To draw conclusions from information and relationships
- I3 Traits and Feelings: To infer traits, feelings, and emotions of characters
- I4 Motives: To infer the motives and reasons for the actions of characters

G **Generalizations:** To Develop Generalizations from a Selection (Evaluative Meaning)

- G1 Main Idea: To recognize the main idea or topic of a paragraph or selection
- G2 Organization: To understand the organization of a paragraph or selection
- G3 Application: To apply information through generalization or prediction
- G4 Purpose: To recognize the author's purpose, motive, or intention
- G5 Viewpoint: To recognize the author's viewpoint, attitude, or bias
- G6 Figurative Language: To interpret figurative language
- G7 Mood: To recognize the mood or tone of a selection
- G8 Style: To recognize qualities of style or structure

For further detail on the three major skill categories (facts, inferences, generalizations) and on the number of test items for each of the 16 skill objectives see the CTBS Teacher's Guide, pp 35-37 (available on loan: contact the Research Office).

The Teacher's Guide also provides information on how to conduct individual and group analysis of performance (pp 31-34) and also offers suggestions for developing skills in each of the three major skill categories. (pp. 38-39).

A Further Look at English Reading Scores of French Immersion Students (1985-86)

Research Office

Division of Planning
and Development

June 1986

A Further Look at the English Reading Scores of French Immersion Students (1985-86)
A supplement to "French Immersion Studies, Year 3 (1985-86)

Background

From the program's inception, there was a concern whether students in the Board's Early French Immersion (FI) program would develop English reading skills comparable to their non-FI peers. Therefore, a monitoring plan, to begin with the first FI cohort when it reached grade 3 (the start of formal instruction in English language), was developed. Annual English reading comprehension testing began in April 1984. Reports were presented to the board on the progress of the initial FI cohort and a control group (matched on IQ) of their non-FI grade peers. The latest report, French Immersion Studies, Year 3 (1985-86); Tests of (English) Reading Skills, January 1986, recaps findings over the three years.

Findings from this longitudinal study generally confirm predictions that the FI students would not do quite as well as their control group in their grade 3 year but would match, then surpass, the non-FI students in subsequent years. By grade 5 (October 1985), the FI students were performing, on average, "slightly but consistently above ... the non-FI students on literal comprehension" and were equal to or marginally better than the control group on inference and generalization skills. In addition to these observations, the study raised questions about the relative performances of the various "streams" that one can find amongst a grade cohort.

The Year 3 report made six recommendations for further inquiry, one specifically related to the relative performances of FI, non-FI, "gifted" and "non-FI, non-gifted" student streams in schools offering "gifted" programs. The recommendation proposed using the second FI cohort and their grade 4 peers. These students had sat for the Canadian Tests of Basic Skills (CTBS) in October 1985. Their reading comprehension scores achieved at this time were to be examined and comparisons among the groups (and with the CTBS national autumn norm scores) were to be made. More exactly, the scores of the grade 4 FI cohort (eight classes in five schools) were to be compared with:

(a) all grade 4 students in the "congregated" programs for the gifted (five classes at five schools);

(b) all the other grade 4 students in the five schools housing these "congregated" programs for the gifted (six classes);

(c) non-FI students in selected "comparison" schools (schools that share an attendance boundary with an FI school), some 32 schools with a total of 50 classes;

(d) all grade 4 students in the YRBE except those whose scores are flagged as possibly unreliable (e.g., students for whom English is a second language) and including FI, "gifted", and non-FI students;

(e) the national norm on the CTBS reading comprehension test.

Findings

In comparing the reading comprehension test achievements of these four groups, each in turn against the FI students' scores, the null hypothesis (No Significant Difference) was tested. Differences that could occur by change more than once in twenty (.05) times were rejected as Not Significant (NS)

Table 1: Comparisons with FI students' reading scores (grade 4, 1985-86)

Group	No. Students	Mean Score	Standard Deviation	Significance Level
French immersion	220	23.4	9.5	-
In "gifted" program	97	33.9	6.7	.0001
In school with "gifted" program but not in that program	132	20.6	7.8	.05
Non-FI comparison cohort	1,124	22.9	8.9	NS
All YRBE grade 4	2,925	23.2	8.7	NS
Autumn national norm	2,939	23.4	9.3	NS

Discussion

The findings were such that further analysis (e.g., item analysis as were conducted for the grade 5 data reported in the Year 3 study) would not likely produce anything relevant to the principal question: Do French Immersion students develop English reading comprehension skills comparable to their non-FI peers?

The FI mean score is, in fact, equal to the national sample (autumn administration) of the CTBS norming population and is marginally higher than

- the YRBE non-FI comparison group of grade 4s,
- the mean YRBE grade 4 score,

but these differences (half a raw score point at most) are not statistically significant. Score differences in this range could happen by chance more than once in twenty times (i.e., if alternative forms of the test were administered, there is a reasonable though small chance that a difference would not be found.

The size of the differences is comparable to that found for the original FI cohort in grade 4 in the spring of 1984 and again in the autumn of 1985 when that population had moved on to grade 5. This may be of special interest to those who wondered whether the initial FI intake was an especially parent-screened group whose achievements would not be equalled by subsequent FI cohorts.

The difference (a mean of 10.5 raw score points) between the FI students and the gifted cohort is such that it would be expected by chance less than one time in 10,000 administrations of alternative forms of the CTBS reading test. This addresses (but not conclusively) the suggestion that the FI is an "elitist" group of superior achievement, more like the "gifted" rather than the mainstream student. Previously (1983-84), the initial FI cohort was found to have a higher than average IQ (Otis-Lennon Mental Abilities Test). But English reading achievement (as only one of many possible criteria) has not shown the FI population as a whole to be outstanding achievers. That the FI students match the non-FI students after only two years of formal instruction in English argues the case for the FI students as a cut above the pack. It should be noted that very high scores (95 percentile or higher) were achieved by individuals in each of the sub-groups in this study.

Previous observations had shown that "non-gifted" students in one school with a "congregated" gifted population had achieved very high reading scores. These were high enough to suggest that, in a school where programs for the gifted are run, there is a "spin-off" benefit for the "non-gifted," as reflected in reading achievement. There was speculation that the presence of "gifted" programs led teachers of other students to elevate their expectations or to use the methods or materials employed in the gifted programs; the result would be a palpable response, higher achievement, from the "non'gifted". This time the scores of all "non-gifted" students in all five of the schools with programs for the gifted were analysed. And this time the results suggest something quite different: the 20.6 average turns out to be significantly lower than the FI and the total YRBE grade 4 averages. This finding was not further explored, but the spin-off theory appears to be discredited, unless the spin is in the opposite direction first indicated. This may be worth pursuing with the autumn 1986 CTBS results; (my informed) guess is that one more set of results would not produce conclusive evidence. At best, it might help us to ask better questions for further inquiry.

Summary

Concerns for the English reading comprehension skills of French Immersion students do not seem justified by the grade 4 CTBS results for the autumn of 1985. As with the earlier FI cohort, these students match or exceed the reading skills of all other comparison groups save those of students selected for the board's programs for the gifted. Findings also discredit a theory that "non-gifted" students in schools with "congregated" programs for the gifted derive a spin-off benefit that is reflected in elevated reading comprehension skills.