Miles Turnbull
University of Prince Edward Island

Doug Hart
and

Sharon Lapkin
Ontario Institute for Studies in Education of the University of Toronto

Grade 6 French Immersion Students' Performance on Large-Scale Reading, Writing, and Mathematics Tests: Building Explanations

We analyzed data from Ontario's provincial testing program to ascertain if the reading, writing, and mathematics skills of grade 6 immersion students were comparable to those of regular English program students. Various immersion program designs were taken into account. The analysis confirms the results of earlier program evaluations that any lags in immersion students' achievement in reading, writing, and math disappear by grade 6. We offer two explanations to account for this result. The lag explanation holds that taking reading, writing, and math in French until the end of grade 3 creates a lag in achievement until English is introduced into the curriculum, after which immersion students catch up to regular students' performance. The selection explanation suggests that immersion test performance improves by grade 6 relative to regular English program counterparts because the composition of the grade 6 cohort is more select than that of earlier cohorts.

Nous avons analysé des données du programme provincial d'évaluation de l'Ontario dans le but d'évaluer les compétences en lecture, en écriture et en mathématiques des élèves de la sixième année en immersion par rapport à celles d'élèves dans le programme anglais régulier. L'analyse a tenu compte de divers programmes d'immersion. Les résultats confirment ceux d'évaluations de programmes antérieures selon lesquelles les retards que peuvent manifester les élèves en immersion en lecture, en écriture et en mathématiques, disparaissent avant la sixième année. Deux propositions sont présentées pour expliquer ces résultats. Selon l'explication basée sur le retard, le fait de lire, d'écrire et de faire des mathématiques en français jusqu'à la fin de la troisième année provoque un retard dans le rendement des élèves jusqu'à ce que l'anglais soit intégré au programme d'études; c'est à partir de ce moment que les élèves en immersion rattrapent ceux dans les programmes réguliers. L'explication reposant sur

Miles Turnbull is an associate professor in the Faculty of Education. He works in the preservice program in French second-language teaching and also in the graduate program in leadership and learning.

Doug Hart is an institutional researcher and senior research officer and has been involved in a large number of program evaluation projects, primarily in the areas of second-language education and professional education in education, health, and law. He is co-principal investigator for the biennial OISE/UT Survey of Educational Issues, which has tracked public opinion on educational issues since 1978.

Sharon Lapkin is a professor in the Modern Language Centre and Second Language Education Program. Her research projects center on French second-language education in Canada and range from program evaluations of core French and immersion to qualitative studies of language learning in progress through detailed analysis of transcribed learner dialogues. She is Co-Editor of the *Canadian Modern Language Review*.

l'acheminement veut que l'amélioration dans la performance des élèves en immersion que l'on constate à partir de la sixième soit due au fait que ces élèves représentent une cohorte plus sélective que les précédentes.

Introduction

As large-scale curriculum-based testing has become more prevalent in school jurisdictions in Canada and the United States over the past 10 years, the "Harris Tories" joined the bandwagon, and in 1996 the Minister of Education introduced a reading, writing, and mathematics testing program in Ontario, managed and administered by the Education Quality and Accountability Office (EQAO). Although these tests were introduced principally as part of Harris' Common Sense Revolution (Gidney, 1999) aimed at increasing the government's accountability to Ontario taxpayers (see e.g., Earl & Nagy, 1998), we provide an example in this article of one way the EQAO test results can be used to understand program processes and dynamics.

We recently conducted a study (Turnbull, Hart, & Lapkin, 2000) that focused on French immersion students' grade 3 and grade 6 results from the provincial testing program in Ontario.[1] In this article we focus on the grade 6 results from that study while providing a brief summary of the grade 3 results (see Turnbull, Hart & Lapkin, 2001, for more details).

This new testing program in Ontario provides a unique opportunity to investigate French immersion students' performance on reading, writing, and mathematics tests close to three decades after many of the large-scale evaluations of French were first conducted. Current programs are mature and developed and attract a more diverse student population compared with those evaluated in the 1970s and 1980s. Moreover, the new EQAO tests differ in nature from standardized tests of English and mathematics used in earlier evaluations in that they are curriculum-based and include a greater variety of testing approaches. It is particularly important to revisit student achievement in French immersion programs at this time, especially because early immersion programs are currently under scrutiny, in times of budgetary restraint, by school boards and parent groups who continue to question whether French immersion students develop reading, writing, and mathematics skills and knowledge that are comparable to those of students in the "regular" English program (Canadian Parents for French, 2000).

The large-scale program evaluations conducted in the 1970s and 1980s relied on standardized tests such as the Canadian Tests of Basic Skills to provide information on the achievement of French immersion students that could be compared with those of students in the regular English program. Results showed that initially immersion students' achievement in English lags behind their academic potential. However, the results also suggested that within one year of the introduction of English language arts into the curriculum, immersion students catch up with comparable students in the regular English program (for summaries of relevant evaluation data,[2] see Genesee, 1987; Swain & Lapkin, 1982; Lapkin & Swain with Shapson, 1990, which include long lists of evaluation reports conducted across Canada in the 1970s and 1980s).

A recent evaluation study in British Columbia (Bournot-Trites & Reeder, 2001) has examined a related but different issue: academic achievement in

French immersion relating to variation in results according to (a) the language of instruction for the content area subject; (b) the language of testing; and (c) the overall amount of instructional time in the "weaker" or immersion language. Bournot-Trites and Reeder (2001) conducted a small-scale but well-controlled study in a single immersion school, which increased instructional time in French from 50% in grades 4 through 7 (following an early total immersion experience in grades 1 through 3) to 80%. Two consecutive cohorts (the last 50% cohort and the first 80% cohort) were tested in mathematics toward the end of their grade 6 year. The 50% group studied math in English, whereas the 80% group took math in French; both groups were tested in English using the Stanford Diagnostic Mathematics Test. The results showed that the 80% group outperformed the 50% group in mathematics: "the students who had acquired their mathematical knowledge in French were able to retrieve it successfully in French" (p. 39). The increased instructional time in the target language produced uneven results in English reading and writing skills: the 80% group did not perform as well as the 50% group on a standardized test of English reading comprehension, but did better than the regular program group with which it was compared in English writing for both narrative and descriptive modes (Reeder, Buntain, & Takakuwa, 1999).

Our study has allowed us to also explore the effect of various approaches to English literacy instruction in French immersion. Currently across Ontario (and across Canada for that matter), boards introduce formal instruction in English at different grades in their early immersion programs. Thus the EQAO testing created a natural laboratory for exploring the effect of program organization on student achievement and possible lags in achievement.

We also use the EQAO test results to explore two possible explanations as to why at grade 3 French immersion students perform as well or only somewhat better than nonimmersion students on reading and writing tests, yet appear to catch up and substantially outperform their peers in the regular program at grade 6 (Turnbull et al., 2000, 2001). The first explanation is an extension of the view that French immersion students' performance in English lags behind their true abilities due to lack of formal instruction in English. Once formal instruction begins, however, students show rapid gains in performance. This phenomenon can be explained by the theoretical premise that in situations of additive bilingualism, linguistic, cognitive, and academic advantages accrue because L1 skills are already well established when exposure to the L2 begins. Because some aspects of language proficiency are interdependent across languages, skills in French literacy acquired in the first years of an early immersion program transfer to the L1 when instruction in English is introduced (Cummins & Swain, 1986).

However, some critics of immersion question whether French immersion students actually catch up to their peers in the regular English program or whether test results are different from earlier to later grades because the grade 3 and grade 6 groups are in fact not comparable. Do French immersion test results at grade 6 appear inflated because the weaker immersion students drop out of the program?[3] The second explanation, then, holds that immersion students' performance does improve, not because instruction in English allows their true potential to show through, but rather because selective attrition with

progress through the grades means the remaining group has a higher average ability than the initial cohort as whole.

Research Questions and Organization

In this article we examine the following research questions:

1. How did French immersion students perform on the provincial grade 6 tests of reading, writing, and mathematics compared with regular English program students?

2. How can we account for the difference in patterns between grade 3 and grade 6? Why do immersion students perform as well or only somewhat better than nonimmersion students at grade 3, yet substantially outperform their peers in the regular program at grade 6?

3. Did French immersion students' performances on the provincial tests differ depending on (a) the grade at which English literacy instruction was introduced; and (b) the proportion of instructional time in French?

The remainder of this article is divided into three sections. The first section, methodology and sample, describes the design and the sample for our study. The second section presents the findings related to the students' performance on the provincial tests. This section also includes our exploration of two possible explanations for differences between test results at grade 3 and grade 6. The final section of the article provides a discussion of the implications of this study.

Methodology and Sample

Secondary Analysis[4] of EQAO Test Data

Our study is based on an analysis of data provided to us by EQAO to compare the overall achievement results of all immersion and regular English program students who completed the grade 6 reading, writing, and mathematics tests in the 1998-1999 school year. We also conducted a short survey of all school districts in Ontario that currently offer early immersion programs to obtain information on: (a) immersion programs (if any) currently or recently offered; (b) program format, particularly starting grade for English literacy instruction and amount of time allocated to this English instruction. Using information from the questionnaire component of the study, then, along with board identifiers in the test data sets (provided by EQAO), we compared students' performance according to (a) the grade at which English reading and writing instruction was introduced, and (b) the relative intensity of French and English instruction in the early grades.[5]

Sample

In 1998-1999 EQAO testing protocols required immersion students in all districts to participate in the grade 6 language and mathematics testing. However, our analyses are based on the sample of districts that responded to our questionnaire on program characteristics. Districts responding to our survey represent 70-75% (depending on skill area) of immersion students participating in the EQAO testing. Moreover, data collected by EQAO indicate that the home language background of immersion and nonimmersion students was similar in the districts that responded to our survey.[6]

Table 1
Program Categories: Early Immersion at Grade 6

Accumulated hours of instruction in English, Grades 1-6	Early Immersion				
	<950	1,050-1,200	1,250-1,400	1,480-1,900	2,800
	% of instructional time in English (ranges)				
Grade 1	0	0-13	0-20	5-23	50
Grade 2	0	0-20	0-25	10-23	50
Grade 3	0-20	0-20	20-30	20-50	50
Grade 4	13-20	13-30	25-30	30-50	50
Grade 5	13-30	20-50	25-50	20-30	50
Grade 6	13-30	30-50	30-50	45-50	50
Number of boards	4	7	7	8	5

A Typology of Immersion Programs at Grade 6

Immersion students approached the grade 6 EQAO tests in 1998-1999 with an array of program backgrounds; the principal differences were related to starting grade for instruction in French and in English language arts. Early immersion programs generally became more similar after grade 3; the proportion of teaching in English increased in programs that introduced English late or held class time in English to a minimum in the early grades. By grade 6 in most programs, students were receiving 30-50% of instruction in English (see Table 1). Despite this convergence, early immersion students reached the end of grade 6 with differing amounts of exposure to instruction in English. This range was further expanded by the inclusion of middle immersion students, instructed wholly in English (with the exception of core French) before starting immersion in grades 4 or 5. Middle immersion programs themselves can be categorized as intensive or partial (see Table 2) regarding the proportion of French.

In summary, the accumulated hours of instruction in English ranged from under 1,000 to 2,800[7] among early immersion programs. The more intensive middle immersion programs had totals approximately equivalent to the partial early immersion programs. However, students in partial middle programs had much higher totals exceeding 4,000 hours.

Table 2
Program Categories: Middle Immersion at Grade 6

Accumulated hours of instruction in English, Grades 1-6	Intensive	Partial
	2,900-2,970	4,100-4,500
	% of instructional time in English (ranges)	
Grade 4	0-27	87
Grade 5	0-27	30-50
Grade 6	13-27	30-50
Number of boards	5	6

Table 3
Numbers of Boards and Students Tested by Program Category

Immersion program	Number of boards	Number of tested students*
Early Only		
<950	3	332
1,050-1,200	5	491
1,250-1,400	6	541
1,480-1,900	7	677
2,800	4	1,067

*Based on 1998-1999 mathematics test results. Students in boards for which we have program information from the questionnaire survey represented 75-76% of tested students depending on skill area.

Our typology of programs at grade 6 was based on accumulated hours of instruction in English where all immersion students in a district were in early immersion programs. Districts with both early and middle immersion programs formed a separate category as did districts with middle immersion alone. Table 3 presents a detailed picture of immersion program types as well as the number of boards and students tested in each category. This typology is used throughout the article when test results are reported.

The Tests
The EQAO tests measure students' knowledge and skills in reading, writing, and mathematics and are based on the Ontario Curriculum grades 1-8 (Ontario Ministry of Education, 1998). At grade 6 all tests are completed in English. Tests are completed independently over a five-day period (2.5 hours per day). Before testing sessions, students complete sample introductory pre-activities to help prepare for the actual tests that assess the knowledge and skills that students are expected to have achieved by the end of grade 6. All tests are developed and assessed by teams of trained classroom teachers and school principals. Tests are scored by more than one assessor using the following criterion-referenced assessment scales: Level 3 identifies a high level of achievement and is the provincial standard. Parents whose children are achieving at this level can be confident that their children will be well prepared for work in the next grade. Level 1 identifies achievement that falls much below the provincial standard. Level 2 identifies achievement that is approaching this standard. Level 4 identifies achievement in the grade 6 curriculum that exceeds the provincial standard. The following overview of the reading, writing, and mathematics tests is based on information available on the EQAO Web site in 2000 (www.eqao.com):

The Reading Tests
Students are required to read several different types of texts (factual information, stories, and poetry) and then answer a variety of questions that assess students' (a) reasoning (general understanding, inferencing); (b) use of ideas from the text; (c) ability to identify the characteristics and organizational ele-

ments in texts; and (d) ability to identify correctly and explain the use of the required language conventions identified in the curriculum.

The Writing Tests

Students are required to write two texts that demonstrate (a) reasoning skills (use of ideas related to the task); (b) ability to write clearly with a sense of voice and audience, using a range of writing techniques; (c) ability to organize ideas and information in a logical and focused way, using appropriate and logical connectors; and (d) general accuracy in spelling, grammar, and punctuation according to the grade-level curriculum-based language conventions. Students also select and submit for evaluation a third writing sample from their work completed during the school year.

The Mathematics Tests

Students work independently to solve problems, write responses to these problem-solving items, and answer multiple-choice questions that assess knowledge, understanding, and use of concepts and procedures required by the mathematics curriculum at each grade level. Students are assessed on their ability to choose, apply, explain, and justify appropriate operations, procedures, and problem-solving strategies to complete tasks related to (a) number sense and numeration; (b) geometry and spatial sense; and (c) patterning and algebra.

Results[8]

Summary of Grade 3 Test Results

Overall, grade 3 results from districts where immersion students wrote the EQAO tests indicate that in the case of reading and writing, immersion students were somewhat more likely to be rated at levels 3 or 4. Distributions for immersion and regular program students were virtually identical in mathematics.

Reading and writing test results for total immersion students at grade 3 (i.e., a small minority of students who had no instruction in English before grade 4) were below the range for those in the regular program. However, in all other cases, immersion students equaled or outperformed regular program students in the same district grouping (Turnbull et al., 2000, 2001).

Although districts had the option of not administering the EQAO reading and writing tests to immersion students at grade 3, the mathematics test was compulsory (either in English or in French translation).[9] In contrast to reading and writing test results, students in total immersion at grade 3 did not appear at a disadvantage. Test results for mathematics did not show a systematic relationship to starting grade or accumulated hours of instruction in English. The grade 3 test data did allow us to examine the link between language of testing and performance (Turnbull et al., 2000). However, because we are not tracking the same students from grade 3 to grade 6 in this article, and because testing at grade 6 was completed in English only, we do not discuss the connections between language of testing and performance in this article.

Overall Grade 6 test results (Research Question 1)

At grade 6 immersion students outperformed those in the regular program in all skill areas. Differences were notable even when we compared these with the differences in performance between regular program students as a whole and

those in enrichment. Immersion students were compared with nonimmersion students in districts that offer immersion programs. The English program enrichment group was also restricted to these districts.

Differences in test performance were most noticeable for reading. Just over half of nonimmersion students achieved level 3 or 4, compared with slightly more than 70% of immersion students (see Table 4). The performance of immersion students was in fact closer to that of English program students in enrichment, although a higher proportion of the latter group reached level 4. On the writing tests (see Table 5), 51% of regular program students obtained level 3 or 4; two thirds of immersion students reached this level. However, the writing performance of immersion students was less similar to that of English program students in enrichment than was the case for reading. Results for mathematics (see Table 6) paralleled those for writing. Just under half of regular program students reached level 3 or 4 compared with 62% of immersion students and almost 75% of English program students in enrichment.

Examining the Results Across Grades 3 and 6 (Research Questions 2 and 3)
How can we account for the difference in patterns between grade 3 and grade 6?[10] Why do immersion students generally perform as well or only somewhat better than nonimmersion students at grade 3, yet substantially outperform their peers in the regular program at grade 6?

It is important to remember that we are dealing here with cross-sectional data;[11] that is, both our grade 3 and grade 6 data were for 1998-1999. We did not have longitudinal data for the same student cohort. Even if we had 1996-1997 data for grade 3, there would have been no way to link students across test years. Hence we could not be sure what proportion of our immersion students consisted of the same students who had been tested in grade 3. In the intervening three years, there would have been an influx of middle immersion students and attrition from early immersion between grade 3 and 6.

However, although our data documented the experiences of students at two grade levels at one time, our explanations are framed in terms of what might happen to the same cohort of students between grades 3 and 6.

Two Explanations of Immersion Students' Performance
In this section we explore two competing explanations of the divergence in pattern between grade 3 and grade 6 test results. Each explanation, with some additional assumptions, can be made to yield specific predictions about patterns of results we should expect in the EQAO test data. The bulk of this section is devoted to setting out and testing these predictions. We should stress that this is an exploratory exercise in which both the predictions and the assumptions on which they are based are open to challenge. Moreover, although the explanations compete in that each invokes a different process, in practice both processes (or neither) may be at work. We think, however, that this exercise is worthwhile as a first attempt to formalize thinking about immersion students' performance in English in ways that might be testable.

The "Extended Lag" Explanation
The first explanation is an extension of a key finding from the first wave of research on French immersion students. As Swain and Lapkin wrote in a 1982 book that reviewed results in three immersion programs,

Table 4
Distribution of Overall Reading Ratings for Immersion and Nonimmersion Students: Grade 6*

Immersion Student	Reading Level				Count
	1 %	2 %	3 %	4 %	
Yes	2.8	26.5	55.7	15.0	8,833
No	10.0	39.4	42.5	8.1	104,318
Students in enrichment	5.2	20.9	48.4	25.5	1,819

*Comparisons are restricted to boards where immersion students participated in testing.

Table 5
Distribution of Overall Writing Ratings for Immersion and Nonimmersion Students: Grade 6*

Immersion Student	Writing Level				Count
	1 %	2 %	3 %	4 %	
Yes	3.5	29.8	48.4	18.3	8,779
No	9.7	39.3	40.1	11.0	102,769
Students in enrichment	5.0	23.6	44.6	26.8	1,802

*Comparisons are restricted to boards where immersion students participated in testing.

Table 6
Distribution of Overall Mathematics Ratings for Immersion and Nonimmersion Students: Grade 6*

Immersion Student	Mathematics Level				Count
	1 %	2 %	3 %	4 %	
Yes	5.3	32.4	45.2	17.0	8,865
No	13.4	38.2	36.5	11.9	105,199
Students in enrichment	6.9	18.5	41.0	33.7	1,821

*Comparisons are restricted to boards where immersion students participated in testing.

It is important to point out that although there were temporary lags in some groups (especially in early total immersion) on tests involving literacy skills, the long-term trend is for immersion students to perform as well or better than, comparison groups on both standardized and non-standardized tests. (p. 41)

The temporary lag typically vanished by grades 3 or 4 in early immersion programs. This was a consistent result of immersion program evaluations in the 1970s and 1980s. A review of immersion programs by the Carleton Board of Education (1989) also concluded:

> Most studies since ... including those carried out in the Ottawa area ... have found a consistent pattern of results. Typically some English language lags are found in grades 1 to 3. For example, first graders who have not yet been taught to read in English obtain lower scores on an English reading test than those of children in an English language grade one program. The same reasoning applies to the finding of relatively lower EFI [early French immersion] scores on certain other language skill measures (e.g., spelling) before grade 4. In all cases, and independent of the proportion of English instruction time introduced into the EFI program from grade 3 on, these observed lags disappear by grade 4 or, at the latest 5. (pp. 3-17)

At the time the most important feature of the lag was that it was temporary, providing reassurance to immersion educators and parents that students' English language skills would not be sacrificed by an intensive French program. However, there was little attempt to provide a theoretically grounded account of the lag or to investigate the mechanisms of its disappearance. Thus early research yielded an empirical description of the lag effect, but did not elaborate an explanation.

A rudimentary conceptualization of the lag effect might be as follows. Total French immersion students' performance in English lags behind the students' academic potential due to lack of formal instruction in English. Once formal instruction begins, however, students show rapid gains in performance. The threshold amount of English-language instruction needed to set in motion gains in English language performance is quite low and easily met by virtually all early immersion program formats. Our analysis of the grade 3 EQAO test results for early immersion and regular program students yielded results consistent with this scenario. Only one group of early immersion students performed less well than regular program students in the same school districts. These were students in total immersion programs through to the end of grade 3; in other words, students who approached the EQAO tests having had no formal instruction in English.

It is possible to extend this rudimentary scenario to account for the fact that although immersion students (except those in total immersion to the end of grade 3) do as well as their regular program peers on tests at grade 3, they outperform them at grade 6. We propose that there is a second, higher threshold of English-language instruction that once crossed results in added gains in immersion students' performance in English.[12] We leave the potential mechanisms at work here as an open question. Two possibilities, however, suggest themselves. The first is that early immersion students may on average have greater academic potential than those in the regular program. This greater aptitude does not translate into better performance, however, until supported by the amount of formal instruction in English represented by the second threshold. A second possible mechanism might involve metalinguistic effects associated with bilingual education, but only after a certain level of formal instruction is reached. For our purposes here, however, the key premise is

simply the existence of a second, higher threshold of formal instruction in English (beyond what is needed for immersion students to reach parity with their regular program peers). This second threshold results in an extended lag before immersion students' performance overtakes that of students in the regular program.

The Selection Explanation

We can construct an alternative explanation of the different patterns in immersion student performance at grades 3 and 6 by focusing on possible changes in the composition of cohorts at each grade level. Attrition from immersion programs has long been a concern for both educators and parents on the grounds of equity for students and effect on both immersion and regular programs. As a 1989 Carleton Board of Education (1989) review of immersion states, "The students who transfer out of French immersion programs are viewed not only as being under stress themselves but also as a stress which is created for the regular English program" (pp. 2-10). It should be noted that the report is citing perceptions, not documented facts. There is, however, probably a consensus among educators that students who transfer out of immersion in the early grades are generally those whose performance is below average. Where educators divide is on the issue of whether students' difficulties are specific to being in an immersion program or more general, and hence likely to reappear when they enter the regular program. There is also debate over how frequently weaker students are effectively "counseled out" of immersion rather than leaving on their (or their parents') own initiative.

For our purposes here, however, the key point is that attrition in early immersion programs is usually related to academic performance. Thus attrition "winnows" each immersion cohort as it progresses through the grades. At grade 6 any given cohort would include a higher proportion of academically strong students than it did at grade 3. Similarly, any contrast of academic ability between immersion and regular program students would be more marked at grade 6 than it was at grade 3. (We do not consider here the effect of any other cohort changes such as movement between school districts.) Thus the stronger performance of immersion students relative to their regular program peers at grade 6 could be attributed to the transfer of weaker students from immersion to the regular program between grades 3 and 6.

Both explanations are supported by some anecdotal evidence. The selection explanation corresponds to educators' concerns about elitism in French immersion and its effect on regular English programs. The extended lag explanation is congruent with persistent, lingering parental doubts about the effects of French immersion on English language skills and subject learning. However, there are few hard data to test these claims. Unfortunately, the EQAO data do not allow us to test these explanations directly. We have neither developmental data on immersion students' aptitudes, academic preparedness, or abilities, nor data on characteristics of students moving out of and into immersion programs between grades 3 and 6. We do, however, have information on test outcomes by program design. Given certain assumptions (outlined below), we can translate these explanations (extended lag, selection) into predictions about patterns of program differences. The outcomes regarding these predictions can then be used to test the explanations.

Differing Implications of the Extended Lag and Selection Explanations for Relative Outcomes in Early Immersion

The extended lag explanation applies only to students in early immersion programs. We have found that at grade 3, test performance bore little relationship to the extent of instruction in English (except for students in total immersion). Following the extended lag explanation, we might reason on this basis that all programs fell below some threshold of instruction in English sufficient to alleviate the lag. That is, no program had provided up to grade 3 a sufficient number of hours of instruction in English to allow immersion students' strong academic ability to translate into better test performance than that of their peers in the regular program. Variations in accumulated hours of instruction in English are not associated with systematic differences in test performance as all lie below the threshold.

At grade 6 there is considerable variation in accumulated class time in English. It is plausible to expect that some but not all programs reached whatever (unknown) threshold is needed to dissipate the extended lag effect. On this assumption, following the extended lag explanation, we would predict that greater accumulated hours in English at grade 6 will be associated with better student performance on the EQAO tests. We can be somewhat more specific about the pattern of differences. We would expect the effect of greater accumulated hours to be clustered about the (unknown) threshold point. Below the threshold, differences in hours are relatively inconsequential for test performance, as they were at grade 3. Above the threshold, we would expect rapidly diminishing returns to further hours of instruction in English. This is because beyond the threshold, we might expect differences in hours of instruction to become much less important in comparison with other factors, in particular individual aptitudes.

The selection explanation might lead to opposite predictions. If we assume that attrition is higher the more French-intensive the program, then the effect of selection should be highest in programs with the least number of accumulated hours of instruction in English and lowest in programs where immersion students have had the greatest exposure to English. This is speculation that cannot be confirmed by empirical evidence that we are aware of and is a topic for future research.

On this basis we would expect stronger test performance in programs with fewer accumulated hours in English. Moreover, we would expect higher test performance to be related to fewer hours of instruction in English over the full range of program differences rather than being localized around some threshold level.

The extended lag and selection explanations can support competing predictions about patterns in EQAO test performance according to program typology. It should be noted that the predicted effects bear an asymmetric relationship to the effect of instruction in English itself. The effects predicted by the extended lag explanation incorporate the actual effect of instruction in English but assume that this effect is much greater at some points (near the threshold) than at others. The selection explanation, in contrast, assumes that any positive effects of greater instructional time in English are overwhelmed by differential rates of selection across programs. That is, the concentration of

higher-aptitude students in more French intensive programs is more conse-
quential than the effect of differences in instructional time in English. For this
reason the predicted effects of the selection explanation may seem less
plausible than those of the extended lag explanation. A more modest version of
the selection explanation might be that factors of differential attrition and
differences in instructional time tend to work against one another, tending to
minimize program differences. We will call the prediction of an inverse rela-
tionship between test scores and accumulated hours of instruction in English, a
prediction of the "strong version" of the selection explanation.

Test Results by Type of Program at Grade 6 (Research Questions 2 and 3)
The reading test results offer no support for the extended lag explanation (see
Table 7). Superficially, there is some support for the selection explanation in
that the districts offering immersion programs with the smallest number of
accumulated hours in English (<950 hours) register the strongest test perfor-
mances. However, reading test performance in these districts was exceptional-
ly high for both immersion and nonimmersion students in comparison with
students in other program categories. This suggests that there were factors
(e.g., student motivation, parental pressure) that affected student outcomes in
both programs that had little to do with the extended lag or selection explana-
tions.

Once we had excluded the districts that offered French immersion pro-
grams with fewer than 950 hours of English instruction, the pattern of test
results revealed little evidence of a threshold point within our range of pro-
gram categories. There was some increase in the proportion of immersion
students who scored at level 4 as we moved to programs with higher total
accumulated hours of instruction in English. However, this tracked a parallel
increase for regular program students, suggesting that factors extraneous to
differences in immersion programs were at work.

The pattern of test results similarly provides little support for the strong
version of the selection explanation. There is clearly no inverse relationship
between reading test scores and accumulated hours of instruction in English

Table 7
Distribution of Overall Reading Ratings by Program Type: Grade 6

| Immersion Program Type | Immersion Student | Overall Reading Level | | | | |
		1 %	2 %	3 %	4 %	Count
EI <950	Yes	1.5	24.0	59.6	14.9	329
	No	8.0	34.9	46.4	10.7	6,977
EI 1,050-1,200	Yes	2.7	29.0	57.5	10.8	489
	No	10.9	41.2	41.3	6.6	9,607
EI 1,250-1,400	Yes	2.4	27.7	57.3	12.6	541
	No	9.5	40.1	43.3	7.1	8,493
EI 1,480-1,900	Yes	2.2	29.0	53.9	14.8	675
	No	9.7	37.8	43.7	8.8	9,200
EI 2,800	Yes	4.9	25.9	54.7	14.4	1,060
	No	8.0	38.7	44.8	8.4	7,183

Table 8
Distribution of Overall Writing Ratings by Program Type: Grade 6

Immersion Program Type	Immersion Student	Overall Writing Level				
		1 %	2 %	3 %	4 %	Count
EI <950	Yes	3.4	24.8	53.1	18.6	322
	No	7.5	36.2	42.1	14.2	6,641
EI 1,050-1,200	Yes	3.4	34.1	46.9	15.6	493
	No	9.3	41.4	39.3	10.0	9,492
EI 1,250-1,400	Yes	4.1	32.9	46.8	16.3	541
	No	10.0	40.3	40.2	9.4	8,383
EI 1,480-1,900	Yes	2.6	25.0	51.4	21.0	663
	No	9.5	38.1	40.3	12.1	9,081
EI 2,800	Yes	3.6	33.6	47.8	15.0	1,073
	No	8.1	38.0	42.2	11.7	7,062

once we exclude the anomalous results for districts with immersion programs that offer 950 or fewer hours of instruction in English.

The writing test results (see Table 8) were broadly similar to those for reading in that the most French intensive grouping (<950 hours of instruction in English) again turned in exceptionally high performances whether they were immersion or nonimmersion. Excluding this group, we find some support for the notion of a threshold effect in the sharp increase in test performance between immersion programs with 1,250-1,400 accumulated hours of instruction in English and those with 1,480-1,900 hours. The proportion of immersion students at level 4 increases from 16.3% to 21.0% and the proportion at levels 3 and 4 combined increases from 63.1 to 72.4%. Moreover, these differences are far in excess of those for regular program students where there is only a small increase in the proportion at level 4. Thus in contrast to reading, the difference in the performance of immersion students on the EQAO writing tests can plausibly be attributed to program differences rather than differences in the districts housing the programs. The gap between the performance of immersion and regular program students on the writing test is substantially wider in districts with programs in the 1,480-1,900 range than in districts with programs accumulating fewer hours in English by the end of grade 6.

This evidence for the existence of a threshold effect and thus for the extended lag explanation is, however, immediately undercut by the pattern of student performance in early immersion programs with the highest accumulated hours of instruction in English (2,800 hours or more). Here immersion students' performance is lower than in the 1,480-1,900 category. The achievement of regular program students is slightly higher than in districts with less English-intensive immersion programs. The result is that differences in the performance of early immersion and regular program students are narrowest in the category of districts with immersion programs that offer the highest levels of instruction in English.

Although the writing test results fail to support the extended lag explanation, there is also little evidence here for the strong version of the selection

Table 9
Distribution of Overall Mathematics Ratings by Program Type: Grade 6

| Immersion Program Type | Immersion Student | Overall Mathematics Level | | | | |
		1 %	2 %	3 %	4 %	Count
EI <950	Yes	1.8	26.8	47.3	24.1	332
	No	9.2	33.1	40.0	17.7	6,737
EI 1,050-1,200	Yes	5.1	44.0	42.8	8.1	491
	No	15.4	41.5	34.8	8.3	9,679
EI 1,250-1,400	Yes	2.4	27.0	49.2	21.4	541
	No	12.6	38.4	37.6	11.4	8,583
EI 1,480-1,900	Yes	4.7	34.9	43.1	17.3	677
	No	13.0	37.3	37.5	12.1	9,227
EI 2,800	Yes	8.9	35.1	43.2	12.8	1,067
	No	12.0	40.7	36.4	10.9	7,267

explanation. The pattern of program differences does not reveal a consistent inverse relationship between writing test scores and accumulated hours of instruction in English.

The mathematics test results (Table 9) are even more ambiguous in terms of support for the selection explanation. Immersion and nonimmersion students in the <950 hours program group again turned in exceptionally high performances. Among other categories, we again find a sharp increase at a particular point both in immersion students' performance and in the gap in performance between immersion and regular program students. In this case, however, the dividing line is between programs with 1,050-1,200 accumulated hours of instruction in English and those with 1,250-1,400 hours. However, immersion students' performance is less strong, and the margin between immersion and regular students is narrow in categories above 1,400 hours and particularly above 2,800 hours. It would be difficult in any circumstance to explain why a threshold in mathematics would be reached before one became measurable in English (because higher achievement on the mathematics tests is deemed to be a consequence of better English-language mastery). But an unsustained pattern of gains is clearly inconsistent with the threshold effect. Again, however, the pattern of results is too irregular to support the strong version of the selection explanation.

Summary and Discussion

At grade 6 in contrast to grade 3, students in immersion clearly outperformed those in the regular program on EQAO tests in reading, writing, and mathematics. Differences were notable even in comparison with the performance of the highly selected English program enrichment group (Turnbull et al, 2000). What had changed since grade 3? One explanation is that test results at grade 3 did not really reflect the greater academic strengths of immersion students. There was an extended lag in development that was remedied some time after grade 3 with further accumulated hours of instruction in English. We have termed this the extended lag explanation. A second explanation is that attrition leads immersion students as a group to be academically stronger, as weaker

students leave early immersion programs. We have called this the selection explanation.

We derived predictions from these two explanations regarding differences in performance among program groupings. In testing these, we found little support for either the extended lag or the selection explanations. If the extended lag explanation is in fact correct, then the threshold level of accumulated hours of instruction in English must lie below the total accumulated attained in all current early immersion programs in Ontario. Alternatively, if selection is the key factor, this is undetectable with our current data either because there is no differentiation of attrition rates (by ability) among our program categories or because selection effects are masked by the effect of increased exposure to English in programs where immersion students are less self-selected (i.e., where attrition linked to academic potential) is lower. Although these results are hardly definitive, they do suggest directions for further investigation.

Conclusion

The results from this study corroborate those of large-scale evaluation studies of French immersion programs conducted some 20 and more years ago. Our findings should reassure parents who are considering French immersion programs for their children: this study establishes that French immersion does not have a negative effect on students' reading, writing, and mathematics skills in English. French immersion students are not disadvantaged in English in the medium to long term. Although research from the 1970s and 1980s had already clearly communicated this message, the findings from our study, conducted in a different sociopolitical climate in education, with tests that are more curriculum-based, add further support to this research and may be perceived as a more accurate reflection of the current situation than studies from the 1970s and 1980s. Although these tests were conducted in Ontario, the results are relevant to immersion educators across Canada and internationally.

Notes

1. We argue that the analyses conducted here are relevant for immersion programs everywhere even though the testing we describe occurred in Ontario.
2. More qualitative analyses of immersion students' writing in English indicate that they are as creative and articulate as regular English comparison students (Lapkin, 1982; Swain, 1975).
3. There is also the issue that students who self-select into middle and late immersion programs are stronger students; this question is not addressed in this article.
4. The initial analysis of test results by EQAO focused on providing profiles of student achievement at different levels: the province as a whole, school districts, and individual schools. Our secondary analysis or reanalysis of the data addresses a different set of issues concerning differences in achievement between regular and immersion program students.
5. Note that statistical testing is not appropriate to this study as the results are subject to nonsampling rather than sampling error. EQAO testing in 1998-1999 yielded population data in that all grade 6 classes in Ontario were expected to participate. Any exclusions would be nonrandom. Test results presented for districts that responded to our survey of programs represent a subset of this population. However, our sample reflects self-selection rather than random selection, so statistical testing cannot be used to estimate error.
6. Immersion students were somewhat less likely than regular program students to be officially designated ESL. However, the proportions of designated ESL students were very small in each program: 1.7% in immersion and 2.9% in the regular program. Immersion students were also somewhat less likely to be identified in remedial programs: 5.3% versus 8.2% for the regular program.

7. 2,800 hours reflects a 50-50 program in French and English.

8. Unfortunately, no data were available from EQAO to allow us to control for socioeconomic and sex differences between the immersion and nonimmersion students. Comparisons made in the same school districts allow us to ensure some control over these factors.

9. Boards with full immersion through to the end of grade 3 tended not to administer literacy testing. However, most others participated in the literacy testing. In the case of mathematics testing, boards with relatively little instruction in English showed a tendency before the 1998-1999 school year to participate in the testing program and an overwhelming preference for translated tests when they did.

10. Although not the focus of our study, an important finding of the EQAO testing program has been of systematic differences in the performance of boys and girls, favoring the latter. Given that girls are frequently overrepresented in immersion, it is worth considering possible the effects of sex. In fact the overrepresentation of girls among immersion students is virtually identical at grades 3 and 6. Girls account for 57% of grade 3 immersion students, compared with 48% of regular program students. At grade 6 the proportions are 56% and 49% respectively. Thus any changes in relative test performance cannot be ascribed to an increasing "feminization" of immersion between grades 3 and 6.

    In the case of reading and writing, the effects of sex and program appear to be additive. Girls do better than boys in each program, and immersion students do better than regular program students across both sexes. In the case of mathematics, sex effects virtually disappear. Immersion students do better than regular program students, but boys and girls perform similarly in each program.

11. The EQAO testing program has yielded a series of cross-sectional datasets with no means of linking individual student results across test years. It is virtually impossible to conduct large-scale longitudinal studies in Ontario due to high levels of student mobility across district boundaries and the absence of a province-wide student number system.

12. We anticipate a second threshold point rather than a simple cumulative effect of accumulated hours of instruction in English. Our analysis of the grade 3 EQAO results showed that immersion students who approached the EQAO grade 3 tests with no background of instruction in English did less well than regular program students. However, among those who had received prior instruction in English, test performance was not systematically related to the accumulated hours of instruction in English mandated in their programs up to the end of grade 3. In other words, past the threshold of some instruction there seemed to be no systematic cumulative effect of additional hours. This is what we would expect if additional effect on test performance depended on reaching a second threshold.

References

Bournot-Trites, M., & Reeder, K. (2001). Interdependence revisited: Mathematics achievement in an intensified French immersion program. *Canadian Modern Language Review, 58*(1), 27-43.

Canadian Parents for French. (2000). *The state of French second language education in Canada.* Ottawa: Author.

Carleton Board of Education. (1989). *Immersion/regular program study.* Ottawa: Author.

Cummins, J., & Swain, M. (1986). *Bilingualism in education.* London and New York: Longman.

Earl, L., & Nagy, P. (1998). Assessment and accountability. *Orbit, 29*(1), 23-28.

EQAO. (1998-1999). *Ontario provincial report on achievement. English language schools.* Toronto, ON: Author. [On-line]. Available: www.eqao.com

Genesee, F. (1987). *Learning through two languages.* Cambridge, MA: Newbury House.

Gidney, R.D. (1999). *From hope to Harris: The reshaping of Ontario's schools.* Toronto, ON: University of Toronto Press.

Lapkin, S. (1982). The English writing skills of French immersion pupils at grade 5. *Canadian Modern Language Review, 39*, 24-33.

Lapkin, S., & Swain, M., with Shapson, S. (1990 and 1992). French immersion research agenda for the 90s. *Canadian Modern Language Review, 46*(4), 638-674. Reprinted in S. Rehorick & V. Edwards (Eds.), *French immersion: Process, product and perspectives*. Welland, ON: Canadian Modern Language Review.

Ontario Ministry of Education. (1998). *Ontario curriculum, grades 1-8*. Toronto, ON: Queen's Printer. [On-line]. Available: www.edu.gov.on.ca

Reeder, K., Buntain, J., & Takakuwa, M. (1999). Intensity of L2 instruction and biliterate proficiency in intermediate years of a French immersion program. *Canadian Modern Language Review, 56*(1), 49-72.

Swain, M. (1975). The writing skills of grade three French immersion pupils. *Working Papers on Bilingualism, 7*, 1-38.

Swain, M., & Lapkin, S. (1982). *Evaluating bilingual education: A Canadian case study*. Clevedon, UK: Multilingual Matters.

Turnbull, M., Hart, D., & Lapkin, S. (2000). *French immersion students' performance on grade 3 and grade 6 provincial tests: Potential impacts on program design*. Toronto, ON: Final report submitted to Education Quality and Accountability Office (EQAO).

Turnbull, M., Hart, D., & Lapkin, S. (2001). Grade 3 immersion students' performance in reading, writing and mathematics: Province-wide results from Ontario (1998/99). *Canadian Modern Language Review, 58*, 9-26.