Evaluation of the Effects of Medium of Instruction on the Science Learning of Hong Kong Secondary Students: Performance on the Science Achievement Test

Din Yan Yip, Wing Kwong Tsang, and Sin Pui Cheung
The Chinese University of Hong Kong

Abstract

This paper is the first of a series of articles reporting the findings of a longitudinal study on the impact of a new language policy about the medium of instruction on science learning of secondary students in Hong Kong. This paper compares the science achievement of Chinese students learning science through a second language, English, with that of students receiving instruction in their mother tongue, Chinese. Based on the scores on a science achievement test made up of multiple-choice and free-response questions, the English-medium students, despite their higher initial ability, were found to perform much more poorly than their Chinese-medium peers. They were particularly weak in problems that assess understanding of abstract concepts, the ability to discriminate between scientific terms, and the ability to apply scientific knowledge in novel or realistic situations. This result implies that the English-medium students were handicapped in science learning by their low levels of English proficiency, and learning English as a subject through the primary years is not sufficient to prepare them for a full English immersion program in secondary school.

Introduction

The medium of instruction in secondary schools has long been a controversial issue in Hong Kong. Prior to 1998, individual schools were free to decide their own instruction medium. Because there is a great demand for graduates with high English proficiency, most schools opted for English as the medium of instruction without considering whether or not their students were capable of learning effectively through English. However, in 1998, the government imposed a language policy for schools, which stipulates that most schools must adopt Chinese as the medium of instruction (MOI), and only about a quarter of the schools that take in the more able students are allowed to teach in English. As this policy will have a far-reaching impact on secondary education in Hong Kong, the present study investigated the effects of this policy on students' learning of science in their first 3 years of secondary schooling.

Background and Rationale of the Study

The majority of the population in Hong Kong is Chinese and speaks Cantonese, a dialect of Chinese, as the language for daily communication. For the purposes of this paper, when the authors use the term "Chinese language," we mean "Cantonese." However, Hong Kong was a British colony from 1842 to 1997; hence, the government uses English as the official language. English is a language of power because proficiency in English confers advantages in securing well-paid posts in the government and in the commercial sector. The high status of the English language is reflected in the education system, which aims to produce graduates proficient in both Chinese and English.

Primary schools in Hong Kong use Chinese as the MOI and teach English as a subject. After Primary 6 (Grade 6), students are streamed according to MOI into different types of secondary schools: those with English as the medium of instruction (EMI) and those with Chinese as the medium of instruction (CMI). Prior to 1998, secondary schools were free to choose their MOI, and most schools opted for EMI to meet the demand for graduates with high levels of English proficiency. One problem with this mode of late immersion is that most students entering English-medium secondary schools were not equipped with an adequate level of English proficiency for them to learn content (nonlanguage) subjects effectively in English. Furthermore, some subject teachers could not teach proficiently in English.

An undesirable outcome of this situation is that many teachers used a mixture of Chinese and English for instruction (Johnson, 1983). Shek, Johnson, and Law (1991) showed this mixed mode of instruction has become more prevalent since the implementation of compulsory education from Primary 1–6 (Grades 1–6) to Secondary 1–3 (Grades 7–9) in 1978. In many of the EMI schools that took in less able students, the teachers used Chinese most of the time for teaching, classroom control, and interpersonal interaction. Interaction in English was mainly restricted to asking simple recall questions that required one-word or single-phrase answers from the students, highlighting vocabulary in English textbooks with explanations in Chinese, and going through worksheets or notes written in English (Johnson, 1997; Pennington, 1999; Evans, 2000).

The weakest students could not understand even very simple text written in English. Their learning style basically consisted of translating content words in a text by looking in the dictionary and writing the Chinese characters alongside the English vocabulary in their notes or textbooks. In order to prepare for tests and examinations, they had to commit to memory terms and isolated chunks of texts in English that they did not quite understand. Under these conditions, it was very unlikely that these students could develop an intrinsic interest in and motivation for learning. Given students' poor English abilities and the pressure to cover syllabi heavily loaded with factual content, many teachers considered a mixed mode of instruction (Chinese and English) as inevitable and even desirable. This situation was also common in some of the schools that admitted students of higher abilities.

Based on the belief that students can learn better through their mother tongue, the Education Department of Hong Kong recently implemented a new policy on the language of instruction in junior secondary schools (Education Department, 1997). Beginning in 1998, most secondary schools (about 300) were required to use students' first language, Chinese, as the MOI. However, a small number of schools (114) that take in the top 25% of students in Secondary 1 (S1) use the second language, English, as the MOI. This figure is based on the findings of local studies (e.g., Brimer et al., 1985; Education Commission, 1990) that suggest that about 30% of students may be able to learn effectively through English. Under this policy, instruction in mixed mode is strongly discouraged. The first cohort of secondary students affected by this policy were those who entered S1 in September 1998.

In order to assess the impact of this language policy, a large-scale project was launched to study the effects of this policy on students' achievement in a number of school subjects, including Chinese, English, mathematics, and science. The project seeks to understand the impact of language immersion on student learning, particularly in relation to late-immersion programs that begin in secondary schools. This paper reports and discusses some of the findings on the science achievement of students studying in the two different streams of schools, the EMI and CMI schools.

Literature Review: Effects of Medium of Instruction on Achievement in Content Subjects

The present paper addresses the impact of MOI on science learning. It seeks to contribute to understanding how science achievement is affected by using Chinese, the native language, or English, a second language, as a medium of instruction in Hong Kong. Although a considerable amount of research has been done on immersion programs, most of these studies have focused on the development of students' language abilities, in both their first and second languages, and only a small number are concerned with the learning of content subjects. Among the studies that explore the effects of MOI on content subjects, the findings are inconclusive and sometimes contradictory. A review of these studies can provide a useful background for understanding the possible impact of Hong Kong's new language policy on the science learning of EMI students.

In Ontario and elsewhere in Canada, a key concern of parents and educators about immersion education was whether native English-speaking students, if taught in their second language (French), would be able to keep up with their English-medium peers in content subjects. To address this concern, a number of standardized tests of mathematics achievement were conducted on immersion and nonimmersion students. Analyses of covariance, with IQ as the covariate, were carried out in order to compare the achievement of students of these two groups in mathematics. In a few cases, the English-taught, nonimmersion students scored significantly higher than the immersion students, but in some other cases, the immersion students demonstrated better performance on the mathematics test. Despite these variations, the results of most studies (e.g., Edwards, Colletta, Fu, & McCarrey 1979; Genesee, 1987; Swain & Lapkin, 1982) show that there are no statistically significant differences between the mathematics achievement of the two student groups, indicating that the immersion students were not at a disadvantage by receiving instruction in French.

In most of the standardized science achievement tests administered to Grades 5 to 8 in Canada, the average scores of the immersion students and their English-taught peers are equivalent (Cummins & Swain, 1986; Genesee, 1987). These results indicate that the science achievement of students is neither positively nor negatively affected by receiving instruction in the second language. Genesee (1976) reported that 1-year-late immersion students scored as well as their English control peers on science tests, after controlling the student samples for IQ using analysis of covariance procedures. Swain (1978) reported that early total-immersion students in Ontario scored as well as the nonimmersion students on standardized science tests. However, late-immersion students were found to score significantly lower than the English-program students at the end of the first year of the program, but they caught up by the end of the second year. These students started the immersion program in Grade 8 and studied the French language as a subject in Grade 7, just a year before immersion. A possible interpretation of the lag in the science achievement of the immersion students at the end of Year 8 is that their second-language skills interfered with the learning of subject matter. This explanation is supported by the subsequent observation that there was no difference in achievement in science, mathematics, geography, and history between the French-immersion and English-program students in Grades 9 and 10.

These findings on the effects of immersion on academic achievement, however, must be treated with caution for a number of reasons. Some of the results (e.g., Andrew, Lapkin, & Swain, 1979; Swain & Barik, 1977) are based on a small number of classes, and so the findings could vary according to class-specific factors. The content of standardized tests is not specific to the curriculum of the schools studied. Moreover, these results are often based on the performance of earlier grades, and the long-term effects of immersion have not been adequately assessed. In order to obtain results that could be generalized to a wider context, it would be preferable to conduct achievement tests constructed according to the curriculum of the schools, on a larger number of immersion classes, for a longer period of time.

In Cameroon, most schools use the official language, either English or French (it varies by region), for instruction starting from the primary level. This is a foreign language for most students. Gfeller and Robinson (1998) report the results of a language-teaching project initiated in the 1980s to compare the effects of teaching young children through their mother tongues or through a foreign language. For the first 3 years of the primary school curriculum, students in the experimental schools learned basic knowledge—including reading, writing, and arithmetic—through their native language. At the same time, they learned the official language as a school subject. From the fourth year onwards, the official language became the language of instruction, although the native language was still used for the teaching of some subjects, such as local culture, history, and geography. When the experimental classes were compared with the control classes at the end of the first year, the experimental students' performance in French, the official language, was the same as for the control classes. However, the students of the experimental group were found to perform slightly better in arithmetic than students of the control group. In addition, they had learned their own native language. These students also showed preference for learning through their own language rather than through the official language because they could express themselves with more facility. Similar findings are reported in studies in other African countries that use similar language policies to revive indigenous languages through early immersion programs (Bunyi, 1999; Collison, 1975; Ehindero, 1980). In contrast to the Canadian findings, these results suggest that early immersion may have negative effects on the learning of content subjects, probably because the children are still not proficient in the second language.

The language situation in the Philippines is multilingual in nature, with the vernaculars used at home and in the neighborhood. Filipino is the national language, a symbol of unity and linguistic identity, whereas English is the language of academic discourse and business. The schools use both Filipino and English. Filipino is used for most subjects, which are mainly related to social studies, while English, the second language, is used for instruction in science and mathematics. For subjects taught in English, the performance ranges from average to below average, with a score of 50% in mathematics and 40% in science. While the poorer performance in science has been attributed to inadequate proficiency on the part of the teachers (González & Sibayan, 1998), the results also suggest that science learning is hindered by the low English proficiency of the students and teachers. English belongs to a totally different language family than Filipino, and competence in English is closely related to socioeconomic status. Children of the poorer classes, due to poor teaching and living conditions, are seldom functionally literate in English for carrying out higher cognitive activities that are required for the effective learning of science (González, 1998).

Foreign-language immersion programs in elementary schools in the United States have grown significantly since the early 1970s. Most immersion programs involve Spanish, as there are a large number of Spanish-speaking immigrants from Central and South America, but there are also programs in French, German, Japanese, and Chinese. These programs are mainly early total immersion or early partial immersion, in which only half of the school day is spent in the immersion language. According to Met and Lorenz (1997), teachers involved in partial-immersion programs reported that their students could handle concrete objectives in the immersion language in the primary grades, but students were frustrated in the learning of abstract concepts in higher grades, probably because their cognitive development was at a level higher than their language proficiency. To facilitate learning in higher grades, some teachers used English when dealing with abstract concepts or allowed their students to communicate in English. These observations point to the need for matching the cognitive demand of the subject curriculum with the second-language ability of the students in order to achieve desirable effects of immersion. The same problem is also observed in Hong Kong. When dealing with more abstract or complex concepts in content subjects at the senior secondary levels, many EMI teachers tend to use a mixed code of Chinese and English to enhance student understanding because they believe that the English proficiency of many students is not sufficiently developed to enable them to think abstractly in their second language.

Meta-analysis is a statistical procedure that can assess the effect of variations in programs. Using this technique, Willig (1985) compared the bilingual programs conducted in American schools with traditional programs, in which non-native English-speaking students were taught exclusively in English. After controlling for prior student differences, minority-language students who were taught in their native language performed significantly better than their English-taught peers throughout the curriculum, such as in reading, language, mathematics, and overall achievement. The bilingual students also showed more positive attitudes towards self and the school. However, such programs are significantly different from the immersion models implemented in Hong Kong and other places, as they aim to protect and to develop children's native languages while developing the majority language. Another important aim of these programs is to build ethnic identity in the minority-language children in a majority-language society (Fishman, 1989).

Although not one single study in this meta-analysis included science achievement, these findings suggest that children can learn more effectively in their mother tongue than in a second language in certain contexts. A criticism of Willig's meta-analysis is that it only included 23 studies, which showed great variation in the nature of students, the social and cultural ethos of the program, and the variety of language intake within the programs. For example, some programs start with children at a similar level of language skills. In other programs, the children start at different language abilities, making classroom teaching more difficult. Such variations make simple generalizations difficult and unreliable.

In a large-scale study involving 9,095 secondary school students in Hong Kong, Siu et al. (1979) investigated the effects of language of instruction on student learning and cognitive development. Four lessons each of mathematics, science, and world history were designed and taught to classes in either English or Chinese. The results indicated that the use of CMI facilitated intellectual development, and the CMI students generally learned subject matter more effectively than the EMI students. However, MOI did not seem to cause any significant difference in the acquisition of subject knowledge for the high-ability students or among the low-ability students. The validity of these conclusions is questionable because the fundamental assumptions underlying the study were flawed. The students in this study were not in fact learning consistently through either language. In the EMI classrooms in particular, the students normally studied through mixed code. Furthermore, the conclusions drawn from an experimental situation of learning in Chinese or English for only four lessons are invalid for implying the long-term effects of MOI. Drawing on similar studies, research on the Canadian immersion programs, for example, has shown that immersion students require 5 to 7 years to catch up with the nonimmersion students in achievement in the first language and in content subjects (Cummins & Swain, 1986; Genesee, 1987).

In Lo's study (1991), standardized tests were used to assess achievement in Chinese, English, and mathematics. The mean achievement scores of 2,638 Secondary 3 (S3) students in each subject were compared according to the MOI used in lessons. Students taught in Chinese showed higher achievement in mathematics than those receiving either English-only instruction or mixed-code instruction in Chinese and English. However, there was no significant difference in mathematics achievement between the students instructed in English and those instructed in the mixed code. One interpretation of these findings is that CMI facilitates mathematics learning because students and teachers can communicate and interact more effectively in Chinese than in English during lessons. It is thus inferred that CMI would have even greater positive effects on the learning of other nonlanguage subjects that are more verbal and less symbolic and figural in nature. An unexpected result from this study is that students receiving instruction in mixed code demonstrated greater achievement in English than those receiving instruction in English. A number of reasons have been proposed to account for the positive effect of mixed-code instruction relative to instruction in English. For instance, the S3 students might not have reached the required level of English proficiency for them to benefit from instruction in English, or the teachers might have lacked the English competency required for effective instruction in English. Based on his findings, Lo proposed that it is not advisable to abolish mixed-code instruction in content subjects, and that well-planned language teaching is more effective than immersion in promoting the development of competence in English for secondary students, including those with high ability. The conclusions from this study should, however, be viewed with caution because the sample sizes of the different groups are not comparable, and the comparisons were made without controlling for students' initial differences in academic ability. Furthermore, the validity and reliability of the achievement tests for Chinese, English, and mathematics, which were constructed and administered by the Education Department, have not been verified.

In a recent study involving more than 12,000 secondary students in Hong Kong schools, Marsh, Hau, and Kong (2000) traced the achievement of native Chinese-speaking students in EMI and CMI schools in language subjects and content subjects for 3 years starting from S1. While the CMI schools basically used Cantonese for teaching, the medium of instruction in the EMI schools varied greatly according to the abilities of students. For many EMI schools with less able students, the teachers might use mainly Cantonese or a mixed code of Cantonese and English for teaching content subjects, though the textbooks and the examinations were in English. Prior student achievement was based on a placement score that represents an aggregate of achievement of a student in all academic subjects at the end of primary schooling, moderated by external examinations. In each of the 3 years following entry into secondary school, the Education Department administered standardized achievement tests in English, Chinese, mathematics, science, geography, and history. The achievement tests were administered to all students near the end of the school year (May to June) in the language of instruction in which the student studied the particular subject.

Marsh et al. (2000) report the findings as follows: After controlling for students' prior ability and other factors, comparison of students' achievement indicates that EMI had positive effects on English proficiency and, to a lesser extent, Chinese proficiency. However, the effects of EMI were negative on all other subjects, being relatively slight for mathematics and greater for history, geography, and science. The positive effects of EMI on English and Chinese achievement were expected. These results support the parental belief that immersion in English promotes the development of both English and Chinese. However, a possible reason for the strong negative effects of EMI on history, geography, and science is that these three subjects are new content areas for secondary students. Learning these subjects in a second language is particularly demanding for students because they have to master the basic terminology as well as develop conceptual understanding of the subject matter and comprehend the textbooks in English. The results also suggest that the English-language skills of the EMI students might be insufficiently developed to cope with the complex curriculum materials in these content subjects. These problems are less serious with mathematics, as mathematics learning involves the use of symbolic terminology that may not be so dependent on the language of instruction. For history, geography, and science, the negative effects associated with instruction in English were the same irrespective of students' initial academic ability. However, students who were initially more proficient in English were less disadvantaged by instruction in English (Marsh et al., 2000).

A problem with the design of the Marsh et al. study, which might affect the validity of data interpretation, is that many of the so-called English-medium schools used Chinese or mixed code for instruction, so only a small number of the EMI schools were truly English-medium. Given that in the Marsh et al. study, most EMI schools were in fact mixed code, it is necessary to analyze student achievement in these two types of EMI schools, English-only and mixed code, separately in order to determine the effects of English-medium instruction on the learning of content subjects and to identify the optimum conditions for effective English immersion. Nevertheless, the findings of the study suggest useful criteria for identifying students who would benefit from English immersion, for example, postulating that the negative effects of instruction in English would be minimized if the selection of students into EMI schools is based on prior English ability.

While the findings of the above studies regarding the effects of MOI on student achievement are inconclusive and sometimes conflicting, some generalizations can be drawn to guide policy making on the medium of instruction for schools in Hong Kong. There is evidence that instruction in English or in mixed code has negative effects on learning for low-ability students. However, the negative effects may decrease as students' English proficiency improves. For high-ability students who have reached a threshold proficiency in both languages, using English as the MOI may enhance language acquisition, particularly in English. For these students, achievement in different content subjects may be affected to a lesser degree.

The Education Department's rationale for the 1997 language policy is consistent with these findings. The designation of most secondary schools into the CMI stream and a small number of schools into the EMI stream, and the strict observance of the language of instruction, are based on the observation that children learn best in their mother tongue. Learning through English, a second language to the Chinese students, can be effective only when the students have reached an adequate level of English proficiency. This policy can hopefully provide an optimal learning environment for students of different language proficiencies and capabilities.

Design of the Study

In order to investigate the effects of the MOI policy on the science learning of secondary students, the present study tracked the progress in science achievement of a cohort of students who entered S1 in September 1999 for 3 consecutive years. The science achievement of the students in the different MOI streams was assessed by a written test near the end of each academic year in S1, S2, and S3.

The 100 secondary schools involved in this study were sampled by the stratified random sampling method. Twenty-five schools were randomly selected from the 114 EMI schools, which in general take in students of the highest academic ability. Approximately 300 CMI schools were divided into high-, medium-, and low-ability strata (CHIG, CMID, and CLOW) according to the mean Academic Aptitude Index (AAI) scores of their S1 student intake. The AAI of a student is a score based on the student's academic performance in school-based examinations in Primary 5 and Primary 6 (Grades 5 and 6), the last 2 years of primary education, moderated by a public aptitude test. Twenty-five schools were randomly selected from each stratum, resulting in a total of 75 CMI schools. This sampling design ensured that the study would cover a representative sample of junior secondary school students in terms of academic ability.

The study used both quantitative and qualitative methods to assess the effects of the MOI on science learning. The quantitative study was based on the performance of the four school strata on the multiple-choice and free-response questions in the science achievement test administered for 3 consecutive years starting in the 1999–2000 academic school year. The Education Department provided the AAI scores of the students participating in this project, which served as a measure of students' pretest scores of academic performance.

The quantitative part of this study was complemented by a student questionnaire eliciting their self-concepts and feelings of competence in science. This gave us some indication of students' psychological development. The questionnaire also explored the classroom climate. In addition, we observed science lessons for a small number of schools selected from the EMI, CHIG, and CMID strata. The CLOW schools were excluded from the lesson observation exercise because the students at these schools tend to have a poor attitude toward learning, and there are more serious classroom management problems that may interfere with the comparison of MOI effects. [The qualitative part of the study, i.e., lesson analysis, will be presented in a separate paper.] Some lessons were videotaped for further analysis. The information collected from questionnaires and from classroom observations regarding classroom climate provided some insights into the effects of the MOI on the differential science achievement of the EMI and CMI students.

Compared with previous local studies on the effects of MOI on science learning, the present study is more systematically and vigorously designed. Its methodology has the following strengths:

1. The project involved a longitudinal study of a large representative sample.
2. The differences in prior achievement of individual students, in terms of the AAI, and other student variables were controlled.
3. The use of multilevel modeling technique allowed better differentiation of the effects of individual students and schools than the multiple regression analyses used in many of the previous studies.
4. As no mixed-code teaching is allowed in the new language policy, the effects of the MOI on student learning can be identified more accurately. Interviews with teachers and students indicate that since the implementation of the new policy, mixed code is generally not practiced in science lessons, though occasionally teachers may use Chinese to supplement their English explanations when teaching difficult or abstract concepts.
5. The new policy ensures that most students in EMI schools have reached a threshold level of English proficiency (based on conclusions drawn from previous research, e.g., Brimer et al., 1985; Education Commission, 1990). This allows for a more valid evaluation of the effects of late-immersion programs on the development of additive bilingualism and acquisition of subject contents.

Focus of the Present Paper

This paper will be the first of a series of publications on the results of the study of the effects of the MOI policy on science learning in Hong Kong. It will focus on the performance of the 1999 cohort of students on the S2 Science Achievement Test (SAT) conducted in April through June 2001. The data analysis includes: (a) multiple comparison of the mean performance on the S2 SAT among the four school strata (i.e., EMI, CHIG, CMID, and CLOW); (b) multilevel analysis of factors affecting students' science achievement in S2; and (c) item analysis of students' performance on individual multiple-choice and free-response questions of the EMI and CHIG strata.

Design of the Secondary 2 Science Achievement Test

The 1999 cohort of students entered S1 in September 1999. Between April 2001 and June 2001, these students took the S2 SAT, after the new MOI policy had been in effect for almost 2 years.

The S2 SAT consists of two parts. Part A contains 26 multiple-choice questions with a total score of 26 marks (the equivalent of "points"). Each question has four options. This part assesses students' knowledge and understanding of science. Part B contains three free-response questions with a total score of 9 marks. For the questions on this part, students are required to organize their ideas and present them accurately and concisely. This means that these questions assess students' higher cognitive and communicative skills as well as their knowledge and understanding of science. The questions in both parts are mainly based on the content of the Hong Kong S2 science curriculum (Curriculum Development Council, 1986). The questions on the S2 SAT show a satisfactory reliability with a Cronbach value of .63.

The CMI students took the Chinese version of the achievement test. However, two versions of the test had been prepared for the EMI students, in order to study the impact of the test language on their understanding and interpretation of the written questions. Most EMI students were given the bilingual test with questions presented in both English and Chinese. They were allowed to answer in either English or Chinese for the free-response questions. At the same time, about one third of the EMI students in each class were given the test in English only. The English-Chinese and English-only versions of the S2 SAT were distributed randomly in the class during the data collection process.

Some may argue that having three versions of the S2 SAT may have confounding effects on the study. Initially, we had considered giving both EMI and CMI students the test in Chinese, but we found that many EMI students had greater difficulties in understanding the questions written in Chinese than in English (for more information, see Appendix). One reason for this might be that the EMI students learned the science content, terminology, complex discourse, and grammatical structures in English; thus, they were unfamiliar with the content when it was written in Chinese. Therefore, we concluded that EMI students would be able to better understand the science questions written in English than in Chinese, and that EMI students would prefer the questions in English rather than in Chinese. However, we felt that bilingual questions would be helpful so the EMI students could refer to the Chinese version when they encountered difficult English expressions. Nevertheless, it should be acknowledged that using three language versions of the science test may confound the results by introducing variables other than students' scientific knowledge that may affect students' performance. For example, the EMI students might experience greater difficulty in comprehending certain questions, or in communicating their ideas and arguments in the free-response items, if they chose to respond in English.

Analysis of Student Performance on the Secondary 2 SAT

Students' performance on the S2 SAT was examined by two different statistical methods and item analysis.

Multiple Comparisons of Mean Scores of Academic Aptitude Index and Science Achievement

The 100 participating schools were divided into four strata according to their medium of instruction and students' initial academic abilities at S1 entry: the EMI, CHIG, CMID, and CLOW. The EMI stratum consists of schools that use English as the medium of instruction; the CHIG, CMID, and CLOW schools use CMI. Initially, the AAI means of the students in the four strata indicated that the S1 students of these four strata fell into four distinct levels of abilities, the highest being in the EMI stratum and the lowest in the CLOW stratum. The difference between the AAI mean of any two successive strata is significant at the .001 level for all cases (see Table 1).

Table 1

Mean Academic Aptitude Index of School Strata and Differences Between Mean Academic Aptitude Index of Successive Strata

		EMI	**CHIG**	**CMID**	**CLOW**
N		4716	4614	4347	3939
Mean		115.81	108.45	99.07	83.66
Standard error		.15	.15	.16	.16
Differences of the means	EMI	-	7.35*	16.73*	32.15*
	CHIG	-	-	9.38*	24.80*
	CMID	-	-	-	15.42*

*The difference of the means is significant at the .001 level.

If the students progress at a similar pace in science learning through S1 to S2, the science achievement scores of the students of the four strata near the end of S2 should correlate to students' initial academic abilities as indicated by their AAI scores. This means that the students of one stratum should perform better than those of the next lower stratum. Table 2 shows whether this is the case by comparing the mean scores of the S2 SAT of any two successive strata.

Table 2

Mean Science Achievement Scores of School Strata and Comparison of Mean Scores

		EMI	**CHIG**	**CMID**	**CLOW**
N		4624	3630	4552	3969
Mean (%)		44.91	48.98	43.11	35.64
Standard deviation		12.54	13.23	13.90	14.08
Differences of the means	EMI	-	-4.07*	1.80*	9.27*
	CHIG	-	-	5.87*	13.34*
	CMID	-	-	-	7.469*

*The difference of the means is significant at the .001 level.

Table 2 shows the mean scores of the four school strata on the S2 SAT and the analysis of variance of the mean scores among the four strata. Among the three CMI strata, the performance is closely related to the initial student abilities at S1 entry, as the mean score of each stratum is significantly higher than that of the next lower stratum, that is, CHIG > CMID > CLOW. However, the students in the EMI stratum performed less satisfactorily than the CHIG stratum, and the difference between their mean scores is significant at the .001 level. The EMI students not only failed to maintain a higher achievement in science than the CHIG students after 2 years of secondary education, but also showed a significantly lower performance on the S2 SAT.

A preliminary conclusion from this analysis is that the EMI students were hindered in science learning by receiving instruction in English, or the CMI students were facilitated in science learning by using Chinese as the MOI. This observation challenges the initial assumption held by the Education Department that the EMI students, who have a higher AAI, are capable of learning effectively in English.

Multilevel Analysis of Factors Affecting Science Achievement

To make a quantitative comparison of the effects of MOI on science achievement among the four school strata, it is necessary to use a statistical technique to control and identify the effects of students' prior attributes that may have significant effects on students' performance. This will measure the science achievements of the different school strata on the basis of a common index. The analyses involve two levels of variables, one at the individual student level, such as pre-entry ability and gender, and the other at the school level, such as the school mean AAI and the MOI. The statistical technique needed for such analyses is multilevel analysis (Bryk & Raudenbush, 1992; Goldstein, 1987). More specifically, the statistical program Hierarchical Linear and Nonlinear Modeling, Version 5 (HLM5) (Raudenbush, Bruk, Cheong, & Congdon, 2000), is used here. By examining the effects of different variables on the science achievement scores, we can construct a baseline model for genuine comparison among the sampled schools. From the baseline model, we can calculate the value-added measures of the sampled school strata in science achievement. By computing the value-addedness of the schools in the four sampling strata, we can compare the degree of improvement of students' science achievement in schools adopting either Chinese or English as MOI.

Effects of Academic Aptitude Index on science achievement

The first set of hierarchical linear models constructed is a two-level hierarchical model for students nested within schools. It is designed to examine the effects of prior ability on science achievement, and does not include any explanatory variables such as gender or socioeconomic status of the students. The independent variables are individual students' AAI and schools' mean AAI, and the dependent variable is the achievement scores on the S2 SAT. This baseline model shows the degree of segregation, or segregation index, in students' initial academic ability among the sampled schools (see Table 3). According to Willms and Raudenbush (1989), segregation index is the percentage of total variance in a variable that lies between schools, that is, the ratio of between-school sum of squares to the total sum of squares.

Table 3

Partition of Variance in Science Achievement Scores

	AAI	**Science achievement scores**
Student-level variance	.439	.767
School-level variance	.922	.237
Percent of within-school (student-level) variance	32.26	76.39
Percent of between-school (school-level) variance	67.74	23.61

The total variance s_T^2 of the test scores is made up of two components, the student-level variance s_P^2 and the school-level variance s_S^2, that is, $s_T^2 = s_P^2 + s_S^2$. The segregation index of AAI shows that Hong Kong students are highly segregated in academic ability when they enter secondary schools. The high percentage of between-school variance (67.74%) indicates that two thirds of the total variance of AAI is among schools, and only one third of the total variance (32.26%) is among students within schools. This means that students within each school are relatively homogeneous in terms of academic ability, while the schools show wide variation in the academic ability of their S1 intake.

For the science achievement of the student cohort of this study, the segregation index is 23.61% (see Table 3), which is much lower than the segregation index for AAI. This indicates that the initial wide gap in students' academic ability among the sampled schools narrowed over time with reference to science achievement. A possible interpretation of this finding is that the EMI students, though they began with higher initial ability, were handicapped in science learning when the MOI was a second language. Conversely, a possible explanation for the CMI students' gain in science achievement is that they were able to learn science more effectively in their mother tongue. Thus, the gap in science achievement between EMI and CMI schools narrowed after 2 years. This interpretation can be tested by item analysis of the performance of the EMI and CMI students on the S2 SAT.

Apart from the baseline model, it is also meaningful to examine students' background and the contextual effects of AAI on science achievement. In the present study, these refer to the effects of individual students' AAI and the school mean AAI respectively. The results of the HLM analysis are summarized in Table 4.

The results in Table 4 show that the effects of individual AAI and school mean AAI are statistically significant. The individual AAI has a positive effect on science achievement. As the AAI of individual students increased by one unit of standard deviation, the science achievement score increased by .561 of a standard deviation unit.

Table 4

Effects of Prior Ability on Science Achievement Scores

	Null Model			**Model 1**		
Fixed effects	**Estimate**	**Standard error**		**Estimate**		**Standard error**
Pupil level (Level 1)						
Intercept	.458	.049		.535*		.033
AAI				.561*		.019
School level (Level 2)						
AAI effects on mean scores				-.175*		.038
Random effects	**Esitmate**	**df**	**Chi-square**	**Estimate**	**df**	**Chi-square**
Level 1 variance	.767			.707		
Percent of Level 1 variance explained				7.86		
Level 2 variance	.237	99	4799.888	.104	98	1201.385
Percent of Level 2 variance explained				56.28		

*The difference of the means is significant at the .001 level.

However, the contextual effect of AAI on science achievement is much different. The coefficient (-0.175) for the effect of the school mean AAI on science achievement is statistically significant but negative in value. The negative value indicates that the school mean AAI has a negative effect on the intercept, implying that a higher school AAI mean is associated with a lower science achievement score. This apparently irrational relationship may be a consequence of the different MOI, that is, mother tongue versus instruction in a second language. It seems logical that students with a higher initial ability will demonstrate higher science achievement as supported by the results in Table 2, which show that among the CMI schools, the science achievement of students correlated with the academic levels of the school strata. However, though the EMI students started with a higher initial ability than the CHIG students, the EMI students did not perform as well as the CHIG students on the science test. One reason for this discrepancy may be that while the EMI schools initially had a higher mean AAI than the CMI schools, the students at EMI schools may have been handicapped in their science learning due to instruction in English because they are less proficient in English than in Chinese. This interpretation would be consistent with the observation that the science achievement of the student cohort has a much lower segregation index than that of AAI.

Effects of sampling strata on science achievement

In order to differentiate the school contextual effects on science achievement, we can replace the school mean AAI with three dummy variables. The dummy variables are the CHIG, CMID, and CLOW strata, using the EMI stratum as the reference point. When these dummy variables are injected into the baseline model to replace the school contextual measure, that is, school mean AAI, a new hierarchical linear model is constructed. This model will enable us to compare the EMI students with the students of the three CMI school strata, with respect to their performance on the SAT, and compute the magnitude of the effects of MOI on science learning. The equations for this multilevel regression model are as follows:

Level 1 model: $Y = b_0 + b_1(ZAAI) + b_2(ZSES) + b_3(Female) + r$

Level 2 model: $b_0 = g_{00} + g_{01}(ZAAI) + g_{02}(ZSES) + g_{03}(CHIG) + g_{04}(CMID) + g_{05}(CLOW) + m_0$

The effects of the sampling strata on science achievement as identified by this hierarchical linear model are summarized in Table 5. At the student level, student AAI has a positive effect on science achievement, whereas being a female has a small negative effect on science achievement. However, the socioeconomic status of a student’s family does not have any statistically significant effect on science achievement.

Table 5

Effects of Sampling Strata on Science Achievement Scores

	Null Model			Model 1		
Fixed effects	**Estimate**	**Standard error**		**Estimate**		**Standard error**
Pupil level (Level 1)						
Intercept	-.064	.048		-.483*		.113
AAI				.384*		.017
Female				-.080*		.016
SES				.003		.007
School level (Level 2)						
AAI effects on mean scores				.196		.106
SES effects on mean scores				.047		.074
CHIG effects on mean scores				.627*		.088
CMID effects on mean scores				.483*		.141
CLOW effects on mean scores				.777*		.235
Random effects	**Estimate**	**df**	**Chi-square**	**Estimate**	**df**	**Chi-square**
Level 1 variance	.675			.638		
Percent of Level 1 variance explained				5.45		
Level 2 variance	.214*	94	5269.437	.080*	69	382.665
Percent of Level 2 variance explained				62.69		

*The difference of the means is significant at the .001 level.

The coefficients of the three CMI school strata on science achievement, with reference to the EMI stratum, are all statistically significant and positive in value. The data indicate that the three strata of CMI schools substantially outperformed the EMI stratum in the science achievement test. More specifically, the CHIG, CMID, and CLOW students outperformed the EMI students by .627, .483, and .777 of a standard deviation in science achievement, respectively, after controlling for the school mean AAI.

This finding supports the previous implication that using EMI has a hindering effect on science learning. Although the EMI students began with a higher initial academic ability than the CMI students, their English proficiency may not have been high enough to enable them to learn science effectively. Science involves a lot of abstract concepts and complex relationships, the accurate use of scientific terminology, and the application of scientific knowledge and skills to solve problems. All these demand a high level of language proficiency (González, 1998; Rollnick, 2000). Thus, the CMI students, who are able to communicate effectively in their mother tongue, may have an advantage over the EMI students in learning science.

Item Analysis of the Multiple-Choice and Free-Response Questions

How can we account for the unsatisfactory performance of the EMI students in comparison with the CMI students? In what aspects are the EMI students particularly constrained in science learning? Is it related to the mastery of scientific terminology, the understanding of abstract concepts, or the development of higher cognitive skills? One way to seek the answers to these questions is to compare the performance of the EMI and CMI students on individual items of the SAT. The CHIG students, despite their lower pre-entry achievement, obtained a significantly higher science achievement score than the EMI students. By performing item analysis between these two strata of students, we can find some clues about the nature of the problems experienced by the EMI students when they learn science through a second language instead of in their mother tongue.

Multiple-choice items in which the EMI students outperformed the CHIG students

Of the total 26 multiple-choice questions on the S2 SAT, the students in the EMI stratum performed better than the students in the CHIG stratum on only five items (Items 1, 7, 8, 21, and 22). This was a significance of $p < .05$ or lower (see Table 6).

Most of these items require direct recall or simple application of science concepts. For example, Item 1 asks to identify the gas present in the greatest amount in the exhaled air; Item 7 requires the identification of substances that turn blue litmus paper red; Item 8 is concerned with a visual problem caused by a diet lacking vitamin A. The EMI students also outperformed the CHIG students in questions that assess simple chemistry concepts, for example, the

Table 6

Multiple-Choice Items on Which English as the Medium of Instruction Students Outperformed Chinese Medium of Instruction High-Ability Students

Item	Stratum	Mean	Standard deviation	Difference of the means	Significance level
1	EMI	.310	.463	-.033	.005
	CHIG	.277	.448		
7	EMI	.600	.490	-.039	.003
	CHIG	.561	.496		
8	EMI	.97	.489	-.030	.031
	CHIG	.368	.482		
21	EMI	.517	.500	-.030	.036
	CHIG	.486	.500		
22	EMI	.783	.412	-.072	.001
	CHIG	.711	.453		

reaction between water and metals (Item 21), and the test for hydrogen gas (Item 22).

Multiple-choice items on which the EMI and CHIG students showed similar performance

For most of the multiple-choice items, the performance of the CHIG students is as good as or better than that of the EMI students. There are 12 items for which the mean score of the CHIG stratum is comparable to that of the EMI stratum (see Table 7).

Some of these items involve the application of science concepts to explain everyday experiences or experimental setups, such as to state the importance of using a water bath for heating alcohol when performing the iodine test for a green leaf (Item 2), to account for the changes observed in a respirometer (Item 13), to identify the conditions for producing an alkaline solution (Item 23), to explain why microscopic algae are abundant at the water surface (Item

24), and to identify the forces acting on a golf ball flying through the air (Item 26) (see Figure 1). Some items require the students to apply science concepts to predict experimental results, such as identifying the paper card that will burn most quickly under converged sunlight (Item 5), comparing the resistance of wires of different length and diameter (Item 15), predicting the brightness of different light bulbs connected in series (Item 16) or in parallel (Item 19), and estimating the effect of short circuiting on the brightness of light bulbs (Item 17) (see Figure 1).

Table 7
Multiple-Choice Items on Which English Medium of Instruction and Chinese Medium of Instruction High-AbilityStudents Showed Comparable Performance

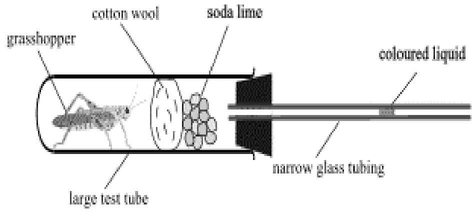
Item	**Stratum**	**Mean**	**Standard deviation**	**Difference of the means**
2	EMI	.268	.443	-.018
	CHIG	.286	.452	
5	EMI	.459	.498	-.072
	CHIG	.531	.499	
6	EMI	.300	.458	-.038
	CHIG	.338	1.383	
11	EMI	.044	.205	-.001
	CHIG	.042	.202	

Table 7 (con't)

Multiple-Choice Items on Which English Medium of Instruction and Chinese Medium of Instruction High-Ability Students Showed Comparable Performance

Item	Stratum	Mean	Standard deviation	Difference of the means
13	EMI	.319	.466	-.008
	CHIG	.312	.463	
15	EMI	.515	.500	-.010
	CHIG	.505	.500	
16	EMI	.652	.476	-.023
	CHIG	.629	.483	
17	EMI	.377	.485	-.008
	CHIG	.368	.482	
19	EMI	.147	.354	.000
	CHIG	.147	.355	
23	EMI	.273	.446	-.024
	CHIG	.297	.457	
24	EMI	.638	.480	-.027
	CHIG	.665	.472	
26	EMI	.068	.252	-.006
	CHIG	.062	.242	

13. The diagram below shows a set-up to measure the respiration rate of a small animal.



After some time, the coloured liquid in the narrow tubing moves towards the left hand side. This is because

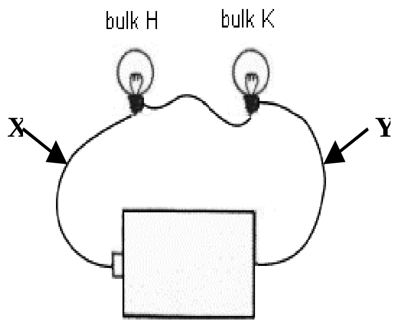
A. the temperature of the air in the large tube decreases.

B. the animal uses up the oxygen in the large tube.

C. more air is breathed in by the animal than is breathed out.

D. the soda lime absorbs the carbon dioxide given off by the animal.

17. The diagram below shows an electrical circuit with two identical light bulbs.



What will happen if a copper wire connects points X and Y?

A. Bulbs H and K will become less bright.

B. Bulbs H and K will become more bright.

C. Bulbs H and K will be burnt out.

D. Bulbs H and K will not light up.

26. An

Wh

(i

A. (

B. (

C. (

D. (

Figure 1. Examples of multiple-choice items on which the EMI and CHIG students showed similar performance.

Multiple-choice items on which the Chinese medium of instruction high-ability students outperformed the English medium of instruction students

For the remaining nine multiple-choice items, the CHIG students performed significantly better than the EMI students at a level of significance of $p < .05$ or lower (see Table 8).

Table 8

Multiple-Choice Items on Which Chinese Medium of Instruction High-Ability Students Outperformed English Medium of Instruction Students

Item	Stratum	Mean	Standard deviation	Difference of the means	Significance Level
3	EMI	.604	.489	-.056	.001
	CHIG	.660	.474		
4	EMI	.513	.500	-.277	.001
	CHIG	.790	.407		
9	EMI	.380	.485	-.037	.003
	CHIG	.417	.493		
10	EMI	.358	.479	-.084	.001
	CHIG	.442	.497		
12	EMI	.651	.477	-.076	.001
	CHIG	.727	.446		
14	EMI	.298	.458	-.028	.038
	CHIG	.326	.469		
18	EMI	.770	.421	-.049	.001
	CHIG	.819	.385		
20	EMI	.842	.364	-.035	.001
	CHIG	.877	.329		
25	EMI	.722	.448	-.043	.001
	CHIG	.765	.424		

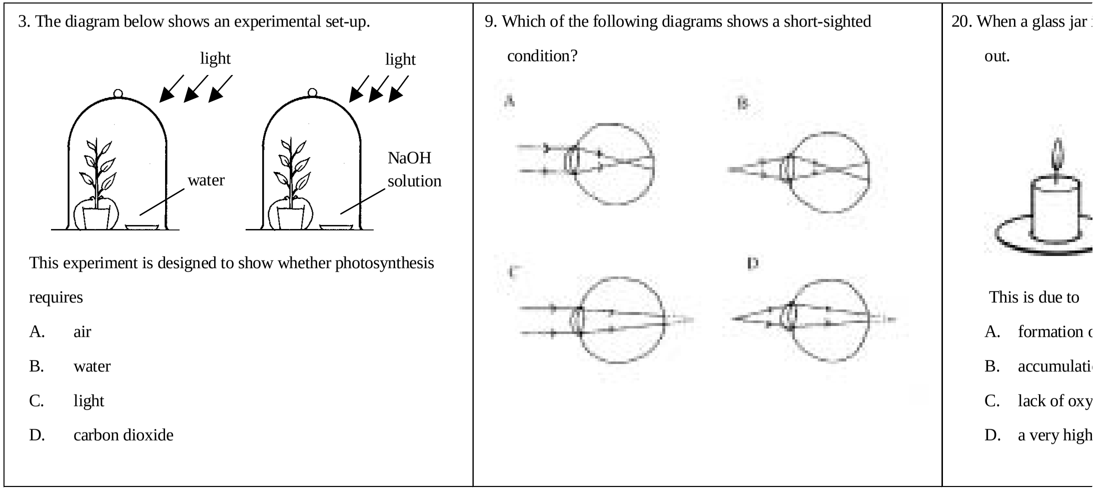


Figure 2. Examples of multiple-choice items on which the CHIG students outperformed the EMI students.

Some of these items are concerned with experimental designs. For example, Item 3 requires the students to identify the aim of an investigation of photosynthesis by comparing two experimental setups; Item 4 assesses the concept of destarching a plant before the experiment; Item 30 asks the students to suggest how to obtain the results from a respirometer more rapidly (see Figure 2).

For Item 3, students should know that sodium hydroxide removes carbon dioxide from the jar, and they have to compare the two setups to find out the effect of carbon dioxide on photosynthesis (Option D). A much greater number of the EMI students than the CHIG students erroneously thought that the experiment is designed to test whether water is necessary for photosynthesis (Option B). This indicates that these students had not mastered the concept of control in designing investigations, so they failed to appreciate the role of the sodium hydroxide solution in removing carbon dioxide from the control setup. Although destarching a plant is a basic precondition for experiments on photosynthesis, the poor performance of the EMI students in comparison with the CHIG students indicates the EMI students did not understand this concept, and many of them erroneously thought keeping the plants in the dark before the experiment was to allow time for the plants to absorb water (Option B).

Items 9, 20, and 25 are concerned with the application of science concepts in realistic or unfamiliar situations. These questions assess the understanding of and the ability to integrate various concepts rather than memory work (see Figure 2). For example, Item 9 involves the identification of the ray diagram that indicates a short-sighted condition. In order to answer this question correctly (Option A), students need to determine the nature of light rays from a distant object, and to identify the location of the image formed inside the eyeball. A significant proportion of the EMI students chose Options C (32%) and D (22%), indicating that many of them could not distinguish between the effects of short sight and long sight.

Item 20 asks students to explain why a lighted candle goes out when covered by a jar. While most students could point out correctly that it is due to a lack of oxygen (Option C), a much greater proportion of the EMI students than the CHIG students erroneously thought the accumulation of carbon dioxide inside the jar extinguishes the flame (Option B). This is a preconception commonly shown by students before they receive formal instruction on burning (Driver, 1993). The result indicates that the EMI students have greater difficulty than CMI students in understanding the concept of combustion. One reason for this may be that it is easier for children to replace their informal preconceptions with proper science concepts if they learn science in their native language instead of in a second language.

Item 25 assesses understanding of the cause of global warming, a complex phenomenon that involves a number of concepts such as photosynthesis and greenhouse gases. A much greater proportion of the EMI students than the CHIG students mixed up global warming with the depletion of the ozone layer (Option D). This misunderstanding is widespread among students, and it persists even after formal instruction (Potts, Stanisstreet, & Boyes, 1996). The confusion probably arises from the fact that the two phenomena have some common characteristics; for example, both are of global dimension and have a partially common cause, in that CFCs damage the ozone layer and also act as greenhouse gases (Stanisstreet & Boyes, 1996). The present result provides further support to the above assertion that students can dispel their informal preconceptions more readily if they receive instruction in their native language rather than in a second language. This is particularly true for alternative conceptions that are very resistant to change.

The EMI students performed more poorly than the CHIG students on some multiple-choice items that appear to be quite straightforward and of low cognitive demand (Items 10, 12, and 18). Closer examination of these questions reveals that they assess the understanding of the precise meaning of certain concept terms in science. For example, when answering Item 10, which is about the functions of the ear bones, the students had to distinguish between the meanings of "to magnify vibrations," "to transmit vibrations," and "to equalize the pressure on both sides of the tympanum." The scientific vocabulary very likely poses particular difficulty for children learning through a foreign language. The EMI students might experience a similar problem for Item 12; they had to understand the scientific meaning of "force" and associate it with the concepts of "gravity," "friction," "weight," and "mass." It is not easy for children to distinguish between the everyday meaning of "force" and the scientific meaning of "force," particularly when the teaching is conducted through a second language. Furthermore, the abstract nature and the subtle differences between the meanings of some of these concepts are always difficult for adolescent students. The same reason can also account for the poorer performance of the EMI students compared with the CHIG students on Item 18, which assesses understanding of the science terminology concerned with heat transfer, such as "radiation," "conduction," and "convection."

Performance on the free-response items

The free-response items require students to apply their scientific concepts and organize their reasoning to explain various physical phenomena. Such questions have a high demand for concept understanding, high-level thinking, and communication skills. The CHIG students performed better than the EMI students on all three items, and the differences in mean scores are significant at the .001 level (see Table 9). This implies that the CHIG students had a better grasp of the concepts involved, and they demonstrated a higher ability to integrate and articulate their scientific knowledge. This is reflected from a comparison of the answers given by the two strata of students.

Table 9

Performance of EMI and CHIG Students on Free-Response Items

Item	Stratum	Mean	Standard deviation	Difference of the means	Significance level
27	EMI	.649	.596	-.561	.001
	CHIG	1.211	.605		
28	EMI	1.064	.719	-.193	.001
	CHIG	1.257	.766		
29	EMI	.918	.871	-.186	.001
	CHIG	1.105	.941		
Total score	EMI	2.632	1.442	-.941	.001
	CHIG	3.573	1.551		

Item 27 asks students to explain why plants usually grow better in soils that contain dead bodies of plants and animals. When answering this question, many CHIG mentioned the presence of dead organic matter and the decomposition process, whereas only a small number of EMI students could state these points. For Item 28, which asks students to explain why a stone would drop more slowly on the moon than on the Earth, more CHIG students than EMI students were able to compare the size or mass of the Earth and the moon, and related this to the difference in gravity on the stone. For Item 29, which is concerned with the rate of melting of an ice block when it is covered by a blanket, more CHIG students predicted correctly that the ice covered with a blanket will melt more slowly and explained their answers in terms of the heat-insulating property of the blanket.

When answering the free-response questions, the students had to organize their thoughts and present them in a logical and concise way. The EMI students apparently had greater difficulty doing this than the CHIG students, as most of the answers of the EMI students were made up of sentences with ambiguous meaning and isolated ideas. Furthermore, few of the EMI students were capable of using appropriate concept terms for their explanations, such as minerals, humus, decomposition, gravity, mass, heat insulation, and heat gain. This shows that the EMI students had greater difficulty in mastering and using the terminology and language of science to verbalize their ideas than their CMI peers.

Conclusion

For the 1999 student cohort, analysis of variance reveals that the differences in mean scores in the S2 SAT among the four school strata are all statistically significant. Among the CMI schools, the mean scores occur in descending order according to the sampling strata, that is, CHIG > CMID > CLOW. Despite the higher mean AAI of the EMI stratum, the mean score of the EMI students was found to be statistically significantly lower than that of the CHIG students. In other words, the CHIG students demonstrated higher achievement in science than the EMI students based on their performance on the S2 SAT. Multilevel analysis shows that all CMI strata substantially outperformed the EMI stratum in the S2 SAT after controlling for the school mean AAI. These findings indicate that using EMI has a potentially negative effect on science learning and support the observations made in local studies (e.g., Johnson & Cheung, 1992; Johnson & Lee, 1987; Johnson & Yau, 1996) that many of the EMI students did not have adequate English proficiency to learn effectively in English.

A comparison between the performance of EMI and CMI students on individual test items provides some clues as to the underlying causes for the lower science achievement of the EMI students and, thus, helps us to identify some possible problems associated with learning science in a second language. Item analysis reveals that the EMI students outperformed the CHIG students only on a small number of multiple-choice items, which made relatively low cognitive demands. Despite their initial higher academic ability as reflected by the AAI scores, the performance of the EMI students on most multiple-choice items was similar to or less satisfactory than that of the CHIG students. These items mainly assess understanding of abstract concepts, terminology used in science, the nature of experimental design, and the ability to apply scientific knowledge in realistic or novel situations. All these require high-level thinking and the mastery of the language of science.

These observations thus provide evidence that the EMI students experienced particular difficulties in mastering scientific terminology and developing higher cognitive skills and conceptual understanding of science subject matter, when compared with their CMI peers who learned science through their native language. As MOI is the key difference between the school setting of the EMI and CMI schools, a likely cause of the lower science achievement of the EMI students, relative to the CHIG students, is that their language proficiency was not good enough for them to learn science effectively in English. This hurdle in science learning as incurred by learning through a second language is also suggested by the much poorer performance of the

EMI students on the free-response items, which demand deeper understanding and better communication skills than the multiple-choice items. These findings do not support the assumption made in the new language policy that students of EMI schools are capable of learning effectively through English.

To appreciate the difficulties encountered by the EMI students, we should bear in mind that these students experienced a sudden change in the MOI from their native language to a second language when they began their secondary education. Throughout their primary education, most EMI students had been instructed in their mother tongue, Chinese, whereas English was only taught as a single subject. While the EMI students were academically more able than the CMI students at the end of primary schooling, they might not have had sufficient English proficiency to learn a content subject such as science in English. This view is substantiated by the findings of some local studies. For example, in Johnson and Cheung's study (1992), designed to compare achievements in Chinese and English using tests from the International Association for the Evaluation of Educational Achievement, they found that S3 students had a lower standard of reading in English than in Chinese. In another study, Johnson and Lee (1987) showed that about a third of S3 students scored little above what could be achieved by chance on tests of listening and reading in English, and that for the same student cohort, the best scores on the English tests were equivalent to the lowest scores on the Chinese tests. Even after 7 years of immersion in English, S7 students demonstrated much better reading skills in Chinese than in English (Johnson & Cheung, 1995). As Hong Kong students in general achieve a much lower standard of language proficiency in English than in Chinese, it is expected that the EMI students will be negatively affected in their learning despite their higher prior ability, at least in the initial years of immersion in English in secondary schooling.

Thus, when students learn science in a second language that they have not yet mastered, they are placed at a distinct disadvantage relative to those who learn in their native language. This implication is supported by the findings on students' performance on the SATs in this study. It may be that, for instance, teachers in EMI schools are seriously constrained in what they can present to their students, and the EMI students may not be able to understand complex concepts in English. Furthermore, some teachers may not be able to present the lesson content fluently and coherently in English. While most EMI students can master concrete concepts, their ability to construct, apply, and present the more abstract and complex scientific ideas in English is frustrated by their limited language proficiency (Willig, 1985). These problems of learning science in a second language have been reported in a number of late-immersion programs (e.g., Marsh et al., 2000; Met & Lorenz, 1997; Swain & Johnson, 1997).

The above analysis and conclusions lead to further questions on the possible impact of the medium of instruction on science learning:

1. What are the effects of MOI on the instructional activities in science lessons? Can these effects account for the difference in science achievement between the EMI and CMI students?
2. Is the constraint in science learning experienced by the EMI students transitory or long lasting?
3. Will the gap in science achievement between the EMI and CHIG students be narrowed after a number of years as the EMI students become more proficient in English or more used to receiving instruction in the second language?
4. Will the gap between the EMI and CMI students widen as the cognitive demands of the science curriculum increase at higher secondary levels?

To seek answers to these questions, an important element of the present study is to explore and identify the causes that may have led to the differential science achievement of the EMI and CMI students. To do this, we need to study the characteristics of science lessons as conducted in the EMI and CMI schools. The classroom climate and nature of instructional activities will have a direct impact on the process of learning. When conducting the S2 SAT, the students were asked to respond to a questionnaire of 16 items, some of which are concerned with their perception of the classroom climate regarding science lessons. The responses on these items can provide useful information on the effects of MOI on the processes of learning and teaching, such as the pedagogical styles of the science teachers and the types of activities carried out during science lessons. Observation of science lessons can also help us understand the realism of classroom contexts in EMI and CMI schools and compare these classrooms in relation to the different MOI used. Other items on the questionnaire explore students' self-concept in science, which is another important outcome of the learning process. These items will contribute to identifying the effects of MOI on students' motivation and interest in learning science, and their self-esteem and self-confidence in the subject.

Thus, through an analysis of student questionnaires and classroom observations, we obtained some insights into the constraints and problems of adopting EMI for science in the early years of secondary schooling in Hong Kong. Such information will hopefully guide the development of appropriate strategies for implementing the MOI policy for the EMI schools, so that their students can have a smooth and efficient transition from CMI to EMI for the effective learning of science and other content subjects through a second language.

References

Andrew, C. M., Lapkin, S., & Swain, M. (1979). *Report on the 1978 evaluation of the Ottawa and Carleton French immersion programs, Grades 5–7*. Toronto, Canada: Ontario Institute for Studies in Education.

Brimer, A., Cheng, W., Ip, B., Johnson, K., Lam, R., Lee, P., Leung, J., et al. (1985). *The effects of the medium of instruction on the achievement of Form 2 students in Hong Kong secondary schools*. Hong Kong: Joint project of the Educational Research Establishment, Education Department and Faculty of Education, Hong Kong University.

Bryk, A. S., & Raudenbush, S. W. (1992). Toward a more appropriate conceptualization of research on school effects: A three-level hierarchical linear model. *American Journal of Education, 97*(1), 65–108.

Bunyi, G. (1999). Rethinking the place of African indigenous languages in African education. *International Journal of Educational Development, 19,* 337–350.

Collison, G. O. (1975). Concept formation in a second language: A study of Ghanaian school children. *Harvard Educational Review, 44*(3), 441–457.

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

Curriculum Development Council. (1986). *Science curriculum: S1–S3*. Hong Kong: Author.

Driver, R. (1993). Beyond appearances: The conservation of matter under physical and chemical transformations. In R. Driver, E. Guesne, & A. Tiberghien (Eds.), *Children's ideas in science* (pp. 145–169). Buckingham, England: Open University Press.

Education Commission. (1990). *Education Commission Report No. 4: The curriculum and behavioural problems in schools*. Hong Kong: Government Printer.

Education Department. (1997). *Medium of instruction: Guidance for secondary schools*. Hong Kong: Printing Department.

Edwards, H. P., Colletta, S., Fu, L., & McCarrey, H. A. (1979). *Evaluation of the federally and provincially funded extensions of the second language programs in the schools of the Ottawa Roman Catholic Separate School Board: Annual report 1978–79*. Toronto, Canada: Ministry of Education, Ontario.

Ehindero, O. (1980). The influence of two languages of instruction on students' level of cognitive development in science. *Journal of Research in Science Teaching, 17*(4), 283–288.

Evans, S. (2000). Classroom language use in Hong Kong's English-medium secondary schools. *Educational Research Journal*, *15*(1), 19–43.

Fishman, J. A. (1989). *Language and ethnicity in minority sociolinguistic perspective*. Clevedon, England: Multilingual Matters.

Genesee, F. (1976). *Evaluation of the 1975–76 Grade 11 French immersion class: Addendum*. Montreal, Canada: Report submitted to the Protestant School Board of Greater Montreal.

Genesee, F. (1987). *Learning through two languages*. Rowley, MA: Newberry House.

Gfeller, E., & Robinson, C. (1998). Which language for teaching? The cultural messages transmitted by the languages used in education. *Language and Education, 12*(1), 18–32.

Goldstein, H. (1987). *Multilevel models in educational and social research*. New York: Oxford University Press.

González, A. (1998). The language planning situation in the Philippines. *Journal of Multilingual and Multicultural Development, 19*(5 & 6), 487–525.

González, A., & Sibayan, B. P. (1998). *Bilingual education in the Philippines*. Manila, Philippines: Linguistic Society of the Philippines.

Johnson, R. K. (1983). Bilingual switching strategies: A study of the modes of teacher-talk in bilingual secondary school classrooms in Hong Kong. *Language Learning and Communication, 2*(3), 267–285.

Johnson, R. K. (1997). The Hong Kong education system. In R. K. Johnson & M. Swain (Eds.), *Immersion education: International perspectives* (pp. 171-189). Cambridge, England: Cambridge University Press.

Johnson, R. K., & Cheung, Y. S. (1992). Reading literacy in Hong Kong. In N. Bird & J. Morris (Eds.), *QUILT and QUILL: Achieving and maintaining quality in language teaching and learning* (pp. 408–419). Hong Kong: Institute of Language in Education.

Johnson, R. K., & Cheung, Y. S. (1995). *Report on standards of reading literacy in Chinese and English in Hong Kong.* Hong Kong: University of Hong Kong.

Johnson, R. K., & Lee, P. M. (1987). Modes of instruction: Teaching strategies and student responses. In R. Lord & H. N. L. Cheng (Eds.), *Language education in Hong Kong* (pp. 99–122). Hong Kong: Chinese University Press.

Johnson, R. K., & Yau, A. S. N. (1996). Coping with second language tests: The development of lexically-based reading strategies. In D. Watkins & J. Biggs (Eds.), *The Chinese learner: Cultural, psychological and contextual influences* (pp. 123–140). Hong Kong: Comparative Education Research Centre, University of Hong Kong.

Lo, L. F. (1991). Academic achievement, medium of instruction and teacher characteristics. *New Horizons, 32,* 18–26.

Marsh, H. W., Hau, K., & Kong, C. (2000). Late immersion and language of instruction in Hong Kong high schools: Achievement growth in language and nonlanguage subjects. *Harvard Educational Review, 70*(3), 302–346.

Met, M., & Lorenz, E. B. (1997). Lessons from U.S. immersion programs: Two decades of experience. In R. K. Johnson & M. Swain (Eds.), *Immersion education: International perspectives* (pp. 243–264). Cambridge, England: Cambridge University Press.

Pennington, M. C. (1999). Framing bilingual classroom discourse: Lessons from Hong Kong secondary English classes. *International Journal of Bilingual Education and Bilingualism, 2,* 53–73.

Potts, A., Stanisstreet, M., & Boyes, E. (1996). Children's ideas about the ozone layer and opportunities for physics teaching. *School Science Review, 78*(283), 57–62.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. T. (2000). *HLM5: Hierarchical linear and non-linear modeling*. Lincolnwood, IL: Scientific Software International.

Rollnick, M. (2000). Current issues and perspectives on second language learning of science. *Studies in Science Education, 35,* 93–122.

Shek, C. C. W., Johnson, R. K., & Law, E. H. F. (1991). Survey of the language policy and practice in 193 Hong Kong secondary schools. *New Horizons in Education, 32,* 1–10.

Siu, P. K., Chung, S. C., Hinton, A., Cheng, Y. N., Lo, L. F., Luk, H. K., Chung, Y. P., et al. (1979). *The effects of the medium of instruction on student cognitive development and academic achievement*. Hong Kong: School of Education, Chinese University of Hong Kong.

Stanisstreet, M., & Boyes, E. (1996). Young people's ideas about global environmental issues. In G. Harris & C. Blackwell (Eds.), *Environmental issues in education* (pp. 37–52). Aldershot, England: Arena.

Swain, M. (1978). French immersion: Early, late or partial? *Canadian Modern Language Review, 34,* 577–585.

Swain, M., & Barik, H. C. (1977). *Report to the Ottawa Board of Education and Carleton Board of Education re: evaluation of the 1976–77 French immersion program in Grades 4–6*. Toronto, Canada: Ontario Institute for Studies in Education.

Swain, M., & Johnson, R. K. (1997). Immersion education: A category within bilingual education. In R. K. Johnson & M. Swain (Eds.), *Immersion education: International perspectives* (pp. 1–16). Cambridge, England: Cambridge University Press.

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

Willig, A. C. (1985). A meta-analysis of selected studies on the effectiveness of bilingual education. *Review of Educational Research, 55,* 269–318.

Willms, J. D., & Raudenbush, S. W. (1989). A longitudinal hierarchical linear model for estimating school effects and their stability. *Journal of Educational Measurement, 26*(3), 209–232.

Appendix

In order to study the impact of test language on students' understanding or interpretation of written questions, two versions of the S2 SAT were prepared for the EMI students. About one third of the EMI students in each class were given the test in English, while the rest were given a bilingual version.

With the data collected from these two versions of the test, we can split the science achievement scores of the EMI students into two parts, according to the language versions of the test, and rerun the hierarchical linear model, which enables us to compare each EMI group with the students of the three CMI school strata with respect to science achievement. The results of this analysis are summarized in Table A1.

The split-version models on science achievement produce some significant differences in sampling-strata effects. All six coefficients of sampling-strata effects on science achievement are significant and positive in values. When the magnitudes of the coefficients between the two language models are compared, the coefficient in the English-version model is larger than the corresponding value in the bilingual-version model by about .3 of a standard deviation. This means that the students in each of the CMI strata outperformed the EMI students tested with the English test version by a much larger margin (i.e., 0.3 of a standard deviation) than the EMI students tested with the bilingual test. In other words, the EMI students performed more poorly on the English test version than on the bilingual version. This observation suggests that the EMI students may experience problems in understanding certain questions on the achievement test when they are presented in English, and, consequently, their performance is adversely affected. These problems are less serious with the bilingual test, as the students can refer to the questions in Chinese when they fail to understand the meaning of a question written in English.

When the achievement test in English was composed, the general level of English proficiency of the EMI students had been taken into consideration, and the test items were carefully constructed in simple English, avoiding the use of difficult vocabularies and complex sentence structure as much as possible. Nevertheless, the results indicate that some EMI students still lack the basic English proficiency necessary for understanding simple written English. It is very likely that such students will experience even greater difficulties in understanding science subject content when the lessons are conducted in English.

Acknowledgment

This research project was supported by a grant from the Education Department, Hong Kong.

Table A1

Effects of Sampling Strata on S2 SAT Scores—Two Versions of Test Papers for EMI Students

	Bilingual Version	**English Version**
Fixed effects		
Pupil level (Level 1)		
Intercept	.176*	.240*
AAI	.553*	.558
School level (Level 2)		
CHIB effects on mean scores	.522*	.839*
CMID effects on mean scores	.313*	.625*
CLOW effects on mean scores	.521*	.816*
Random effects		
Percent of L1 variance explained	7.76	7.90
Percent of L2 variance explained	71.53	68.89

*Significant at .05 level