Economic Council
of Canada

Conseil économique
du Canada

[redacted]

Working Paper
Document de travail

Working Paper No. 7

Science Achievement in Canadian Schools:

National and International Comparisons

Robert K. Crocker

1990

ISSN 1180-3487

Science Achievement in Canadian Schools

Contents

Foreword

To meet increasing global competition Canada's quest for productivity improvement must focus on technological advance and the skill needs of the knowledge economy. Education and training will thus be a cornerstone of any long-run strategy for economic development. Accordingly, the Economic Council of Canada has decided to undertake a study of this complex and controversial field which will be completed in late 1991. The broad framework of analysis will consider a variety of topics concerned with quality and the efficiency of resource allocation in the education and training system.

One specific question that has attracted considerable public attention has to do with the performance of Canadian students relative to their counterparts in other countries. The present study by Robert Crocker, of Memorial University, examines achievement in science subjects, and a companion study by David Robitaille, of the University of British Columbia, entitled *Canadian Participation in the Second International Mathematics Study*, does the same for achievement in mathematics. The authors interpret the international comparisons from a Canadian standpoint, extract the main conclusions, and indicate the relevant policy implications.

Judith Maxwell
Chairman

Acknowledgment

A project of the scope of the Second International Science Study requires the active participation of a large team of researchers, support staff, and collaborators in schools and school systems. Equally important, such a project can be successful only through the cooperation of many thousands of teachers and students, who took the time to complete the necessary tests and other instruments. Specific acknowledgment of the contributions of participants in the Canadian study has been given in the relevant documents cited throughout this report. Nevertheless, it is appropriate to recognize directly the support of my colleagues Michael Connelly and Heidi Kass of the English Canadian Study and Gilles Dussault of the French Canadian Study. It is their work, as much as my own, which is represented in the pages of this report.

BACKGROUND AND CONTEXT OF THE STUDY

International Organization

The International Association for the Evaluation of Educational Achievement (IEA) is a voluntary organization of educational research centres and similar organizations in participating countries. The aim of the organization is to conduct comparative international studies of educational achievement and factors related to achievement, with a view to stimulating policy debate and action to improve education in member countries.

The first IEA studies were conducted in the 1960s and 1970s. The first international comparisons for science achievement were reported in 1973.[1] The Second International Science Study (SISS) is one of several more comprehensive studies launched in the early 1980s and continuing throughout the decade. Studies in classroom environments and mathematics have been completed. A preliminary international report on the science study was published in 1988.[2] Further reports are in draft form at the time of writing.

Each study organized by IEA establishes an international coordinating centre, based in one of the participating countries. For the first and second science studies, this centre was located at the Australian Council for Educational Research, located in Melbourne. Studies in individual countries are conducted independently, within a broad structure agreed upon at international meetings, and outlined in documents prepared by the international coordinating centre. Each participating country is expected to carry out certain core activities, and particularly to use certain core tests at common population levels, to permit valid international comparisons.

Organization of the Canadian Studies

Canada was not a participant in the first round of IEA studies. In the second round, such studies have proven to be exceedingly difficult to organize in Canada, for reasons having to do with the absence of national coordinating bodies or national research centres in the country. The mathematics and classroom environments studies, for example, were conducted in only two provinces (Ontario and British Columbia for mathematics; Ontario and Quebec for classroom environments). Because the provincial studies were independent, each province had to be treated as a separate country in the international reports.

In organizing the science study, a concerted effort was made to achieve national representation. This was accomplished through establishing a consortium of three regional research centres, located at the University of Alberta, the Ontario Institute for Studies in Education, and Memorial University of Newfoundland. The principal investigators[3] agreed to operate within a single research framework, embodied in a common funding proposal. Once the study had been funded, the principal investigators assumed responsibility for coordinating regional research teams, working from a common agenda. These individuals also represented the Canadian study at international meetings, and participated in various stages of design and execution of the international study.

Although the study began by collecting data from all provinces, it was not possible to maintain this level of representation in the achievement testing phase. Although tests were prepared for use in both English and French, it proved to be difficult to find a way to administer the French tests in Quebec. This, of course, is not an unusual difficulty, and is merely an extreme instance of what frequently happens when any attempt is made to conduct studies in education which cross provincial jurisdictions. It was therefore decided subsequently that an independent parallel francophone achievement study should be conducted.[4] Because it was agreed that this study should not be confined to Quebec, but should encompass francophone students in other provinces, the French language versions of the original tests were dropped, and revised versions produced for use in the francophone study.

To ensure full representatives of English Canada, the original Canadian group subsequently undertook to conduct testing in English language schools in Quebec. In the international reports, Canada continues to be treated as two countries, based on official language groups,[5] following a pattern first used for countries such as Belgium. Throughout this report, references to the Canadian study will be made in discussing the structure of the system and matters of curriculum. A distinction between English and French studies will have to be made in discussing achievement.

The Canadian study was conducted in two major phases. The first involved a comprehensive case study, designed to present a historical and structural account of science education, to describe and compare provincial science curriculums, and to examine perceptions of science teaching. This phase was designed to complement, without duplicating, a policy study being conducted at the same time by the Science Council of Canada,[6] while also fulfilling the requirement in the international study for a case study report from each country. The second phase was more directly related to the goals of the international study, in that the primary objective was to measure achievement levels. The Canadian study went well beyond this goal, however, in conducting detailed interview and observational studies of classroom teaching.

While the francophone study involved only the achievement testing phase, it is important to note that the original study had gathered data on curriculum and related matters from all parts of the country, in both French and English. The report of the first phase of the English study[7] can therefore be considered to represent the country as a whole. Similarly, the original case study submitted to the international centre describes the structure of education in all parts of the country, and refers to science education in both English and French Canada.

Issues in Interpreting Comparative Achievement Results

Reactions to the results of major educational testing programs are often extreme. Some use test results as a weapon in an ongoing battle over what constitutes desirable educational standards. Others disparage testing programs as, at best, capable of measuring only a limited range of important educational outcomes or, at worst, measuring inappropriate or irrelevant outcomes. Test results which purport to compare the performance of different jurisdictions present special problems because a given test may be of widely different validity for different systems, and because the practice of ranking systems often exaggerates any differences which do exist. In order to place the results which follow in proper perspective, it is necessary to address several fundamental points in the interpretation and use of test data.

First, it must be recognized that, strictly speaking, tests are valid only to the extent that they measure outcomes of schooling that are generally considered important in a society. If there is no such agreement, or if statements of desirable outcomes differ widely from one setting to another, test results are likely to be meaningless. In the case of the science tests used in this study, it must be recognized that the content of these tests represent a consensus about a common core of knowledge reached by an international group of science education specialists. On the one hand, this would seem to make a powerful case for their validity. On the other hand, it is also the case that tests arrived at in such a manner are likely to omit areas which are controversial or which represent the views of those repre-senting only some jurisdictions. For example, the decision not to include items on scientific methods or processes in the international core tests would be seen by many as a serious omission, and as an example of failure of the tests to capture an essential feature of science curriculum in some countries. The best argument that can be made for test validity is that there is perhaps little disagreement over what is included in the tests. Rather, the disagreement is likely to be over what has been omitted.

Second, it is important to point out that tests can be developed, and test scores used, in two fundamentally different ways. A test is said to be criterion-referenced if some standard or expectation can be developed, based on a clearly-defined body of knowledge or skill. An example would be a driver's test, in which candidates are expected to possess specific knowledge (rules of the road and the like) and be able to perform specific tasks (such as parallel parking) before a driver's license is issued. Candidates who can meet the established criteria pass, while those who cannot fail. What is important is that no prior judgment is made on the number of candidates who might be expected to pass, and no adjustment of the test criterion is made to ensure that more or fewer people pass.

This seems so obvious that one might ask how there can be any other interpretation of a test result. Nevertheless, achievement test scores are often interpreted from a second point of reference, namely "how well does a student or group of students perform relative to others taking the test?" Such tests may be developed initially so as to highlight differences between students. The term normative or norm-referenced is used to refer to such tests. The essential feature of such tests is that the reference point for judging performance is not the subject matter itself, but rather the performance of others.

Such an interpretation emphasizes rank order rather than score as the basis for comparison.

The distinction between criterion-referenced and norm-referenced tests applies at the system as well as the individual level. It is important to note that comparative achievement results are always based on a normative interpretation. Thus, in comparing Canada's performance relative to other countries, little is being said about the levels of performance relative to a body of knowledge, especially the specific content of a curriculum. What is being said is that others perform better or worse. Of course, some sense of performance criteria can be gained by looking at scores rather than a simple ranking of countries. In particular, this conveys some sense of whether the differences in ranking are small or large. Even more information can be gained by looking at specific test topics or items. However, this requires a level of detail of analysis which is generally unsuitable for public reporting. In any case, it must be kept in mind that when tests are constructed with normative interpretations in mind, they are likely to contain items designed to maximize discrimination between individuals and groups.

A third issue which deserves mention relates to the distinction already made between the intended and the translated curriculum. Jurisdictions which have only a loose connection between the two, or which have many "intended" curriculums (as is true for Canada), are likely to find the results of common achievement tests especially difficult to interpret. There is little doubt, for example, that the SISS tests more closely match the curriculum in some provinces than in others. This has a direct bearing on interprovincial comparisons, and an indirect bearing on international comparisons, as scores for Canada would represent an average from a test with different degrees of validity in different provinces. Some of the differences in validity are reflected in the curriculum analyses and are captured by "opportunity to learn" ratings. However, these at best are rough indicators, and amount to something less than a full analysis of test validity.

Overall, it is necessary to avoid either excessive enthusiasm or excessive caution in looking at the results presented in this report. Test results are only one source of data on which to base policy deliberation. The overview of science education in Canada, presented in the third section of this report, is intended to serve as a reminder of the complexities of the system and the broad range of issues which underlie educational policy. Public enthusiasm for simple straightforward indicators of productivity of the system must be tempered by recognition of the limitations of these or any other tests. At the same time, a strong case can be made for the validity of the tests as indicators of core knowledge in an increasingly crucial facet of education.

DESIGN AND METHODOLOGY

International and National Project Goals

The basic goals of the international study were adopted by the IEA governing body at its annual meeting in 1980. The study had five major goals as follows:

1 to measure, by means of large-scale survey procedures, the current state of school science education across the world;

2 to examine the ways in which science education had changed since the early 1970s;

3 to identify the factors which explain differences in the output of science education programs across countries, and between students within countries, with particular attention to science curriculum as an explanatory factor;

4 to investigate changes in the patterns of relationships between the explanatory factors and outputs between the early 1970s and the early 1980s; and

5 to assist all participating countries, especially less developed countries, to carry out associated national studies of science education in order to investigate issues of particular interest within their own country.

The specific details of each national study are worked out within the country itself. However, to permit valid international comparisons, certain common activities were required. The most obvious of these is the administration of international core tests at particular levels of schooling, using sampling procedures designed to yield representative data for the country. In addition, participating countries were expected to administer school and teacher questionnaires, and to prepare a *national case study* describing the structure of education, curriculum development procedures, and related features of their systems.

More specifically, the international requirements of SISS can be summarized as the analysis of three levels of science education, the *intended*, *translated*, and *achieved* curriculum. The intended curriculum consists of the detailed specification of curriculum objectives and content, as set out in curriculum documents and other statements within the appropriate educational jurisdictions. The notion of translated curriculum derives from the recognition that the intended curriculum may undergo changes at the hands of local jurisdictions or individual teachers. The translated curriculum was analyzed using data on teacher perceptions, beliefs, opinions, and classroom practices. Of particular relevance is the concept of *opportunity to learn*, which is based on teacher ratings of emphasis placed on various curriculum topics and content. Finally, the achieved curriculum is taken to be simply the levels of attainment of students on the tests used in the study.

The main objectives developed by the original Canadian study team were as follows:

1 to cooperate with the IEA study as specified in the international objectives;
2 to generate descriptive knowledge of student achievement, student attitudes, teacher attitudes and teacher perceptions in science education;
3 to examine the relationships among achievement and attitudes and the links between these outcomes and curriculum policies and practices; and
4 to participate in, and influence, world trends in science education and science education research.

The Canadian researchers found it necessary to modify many of the detailed requirements and instruments used in the international study because these generally assumed some form of unitary national educational system. Thus, for example, the Canadian curriculum case study needed to be much more detailed than those for most other countries, because it was necessary to describe provincial curriculums separately. Similarly, questionnaires had to be modified to accommodate privacy and similar considerations which are more at issue in Canada than in some other countries. Finally, sampling schemes had to be developed on a provincial rather than a national basis, because the requirements for access to schools varied from one jurisdiction to another.

International Sampling Requirements

The international study defined three population levels for testing. These were at the upper elementary, junior secondary, and last year of secondary education. Within countries, there were options of choosing age or grade levels in defining these populations. In practice, grade level populations were most often chosen, for reasons of ease in assembling students for test administration. Specifically, Population 1 was defined as the 10-year-old age cohort or the grade level at which most 10-year-olds would be found, Population 2 was defined as 14-year-olds or the corresponding grade group, and Population 3 was defined as those in the last year of secondary education in each country (both age and grade levels varied across countries for this population).

Population sizes at the first two levels corresponded roughly to the total relevant age groups (10- and 14-year-olds) in each country, since in most participating countries almost all children are in school up to this level. The size of Population 3 varied with the proportion of the age group in school and the proportion taking the subjects of interest in the study. Participation rates in science courses at the senior secondary level are, in fact, of substantial importance in interpreting the achievement data at the upper secondary level.

Populations and Samples: English

In Canada, the original population of interest was taken to be all students at the appropriate levels in provincially supported schools in the participating provinces and territories. Excluded were students in private schools, schools for the handicapped, hospital schools, and federally supported schools such as armed forces schools and those on Indian reserves. Subsequently, Francophone schools in all provinces were also excluded from the English study and were sampled separately in the French study.

It was decided to select samples by grade rather than age, for the obvious reason that this is how students in schools are grouped. Grades 5 and 9 were selected for Populations 1 and 2 respectively. By age, this corresponded approximately to 10- and 14-year-olds. Selecting students for Population 3 was a much more complex process. The largest province, Ontario, at the time considered grade 13 as the last year of high school, even though only a relatively small proportion of the total age group (mainly those intending to enter university) continued to this level. Furthermore, since Population 3 involved tests in three separate subjects (biology, chemistry, physics), and most secondary schools are on some form of credit system, there was no assurance that students in their last year would be taking the courses of interest. It was finally decided to define Population 3 as those students taking the highest level high school course in the subjects of interest, whatever their actual grade level. In most cases, except in Ontario, this meant that students were in grade 12, though some students from earlier grade levels may have been found in the courses.

Samples were selected by a two stage random sampling procedure, with schools as the sampling unit at the first stage and classes within schools as the sampling unit at the second. In most cases, a single intact class was selected within each school. The original intent of the study was to produce summaries of results at the regional level. Accordingly, calculations of desired sample sizes were based on estimates of sampling errors for regions. These estimates yielded target samples of 80 schools, or 2,000 students per region, at each population level. Since, in practice, sampling had to be carried out within provinces, the regional samples were divided equally among the provinces within the region, yielding a nominal sample of 20 schools, and 500 students per province in the eastern and western regions. Since Ontario was treated as a region, the sample size for Ontario was the same as that for other regions.

Populations and Samples: French

Target populations were defined in the same way for the French as for the English study. Because of the distribution of the Francophone population of Canada, a different approach to sampling was taken for the French study. The country was first divided into 20 regions, based on clusters of French-speaking population and on large areas of scattered population. Thus, for example, rural New Brunswick was considered one region, while all of Saskatchewan, Alberta, and British Columbia formed another. Various parts of Quebec formed the largest number of regions. Schools were selected at random in proportion to the number of students in each region, and random samples of students were selected within these schools.

Sampling and Measurement Errors

Data derived from studies such as those reported here are subject to errors from several sources. The two major types of error may be described as sampling error and measurement error. Sampling error occurs when the entire population of interest cannot be measured and inferences are made about the population from the sample used. The size of such error depends on sample size and on particular features of the sampling design. In this study, the main feature, other than sample size, which influenced the size of the sampling error was the decision to use intact classes. Samples of this sort are known as cluster samples. In general, the sampling errors for clusters are somewhat larger than for simple random samples of the same size.

Measurement error occurs as a result of our inability to construct perfectly reliable and valid instruments. Measurement error for an individual may be thought of as the difference between the observed score on the test at hand and a hypothetical *true* score which would have been obtained had the test itself been free of error. The size of the measurement error depends on the degree to which the test represents the underlying trait or traits being measured (validity) and the degree to which the test scores give consistent results under the same measurement conditions (reliability). Standard techniques exist for estimating both validity and reliability. However, score comparisons typically take into account the reliability of the test, as this is the main source of random error. Reliabilities of individual SISS tests were generally in the .75 range, which yields errors in individual scores of the order of 8 per cent. Errors for summary statistics such as group means are, however, much smaller, to the point where sampling is a more important source of error than measurement in making comparisons or in estimating population values from sample values.

The actual calculation of sampling errors is a fairly complicated process. Table 1 presents a summary of percentage errors in mean scores calculated for the sampling conditions and sample sizes used in this study. In general, we can see from this table that sample sizes of 400 or so, which are typical of provincial samples at the Population 1 and 2 levels, have errors of the order of 3 percentage points. As it happens, samples of 200 or so, more typical of the Population 3 level, have similar error rates because the standard deviations of scores at this level tend to be smaller. Error rates for data at the national level are much smaller, being in the range of 1 per cent or less.[8]

Because sampling was done by province or region, the sampling fractions[9] were widely different for different subgroups. This is common practice, and is designed to avoid having extremely small samples in smaller subpopulations. For example, a random sample of 1,000 Canadians would yield approximately 300 individuals in Ontario but only about four in Prince Edward Island. Although such a sample would be representative for Canada, it would be impossible to say anything about the population of Prince Edward Island from this sample. Oversampling of small population units solves this problem. However, this requires that weighting factors be calculated when data from different subpopulations are combined to give a national figure. In this study, weights were required for schools, as well as provinces and regions, because different proportions of students were sampled in different size schools.

Table 1

Percentage Sampling Errors for Various Sample Sizes[1]

Sample size	Grade 5 S.D. = 17.7	Grade 9 S.D. = 15.7	Grade 12/13 S.D. = 13.5
200	4.4	3.8	3.2
500	2.8	2.4	2.0
1,000	2.0	1.8	1.4
1,500	1.6	1.4	1.2

1 Calculations are based on a cluster sample of average cluster size 20. The design effect (DEFF) is estimated at 3.0, based on sample calculations of the intraclass correlation coefficient (rho). For specific details of sampling error calculations, the reader is referred to standard texts such as Rossi, Wright, and Anderson (1983).

Structure of the Instruments

All tests used were made up of multiple choice items, and consisted of core items common across all countries as well as national options developed within each country. At the Population 1 level, the tests consisted of two sections, the first containing a set of 24 items common to all students, and the second made up of six rotated forms assembled by combining pairs of four twelve-item sets. The purpose of the rotated forms was to allow a larger set of items to be used than any one student could be expected to answer. A total of 72 items was therefore used, with each student responding to 48 of these. A similar procedure was followed at the Population 2 level, using 90 items with each student responding to 60. Although not all students answered every item, it was possible to gain sufficient data in all items to permit item level analysis, so as to help identify areas of particular strength or weakness. The Population 3 test consisted of a common core of 30 items, along with specialized sets of 30 items in each of biology, chemistry, and physics. Individual students wrote the core section and one of the specialized sections.

In addtion to the tests, students were administered a questionnaire containing sections on general background, attitudes towards school and science, and perceptions of science learning. Teachers were also asked to give ratings of test items in terms of importance and time spent. These responses were intended to give a measure of student opportunity to learn the material contained in the test items. Finally, a school questionnaire was used to gather data on demographic and contextual variables related to science achievement.

At all population levels, the overall testing program, including student background and attitude questionnaires, took approximately three hours to administer.

As pointed out earlier, the researchers viewed the tests as having substantial face validity in representing a common core of scientific knowledge. The most serious weakness of the tests, from the perspective of the objectives of science teaching as generally understood in Canada, was the omission of items on such areas as scientific processes, science and society, and technological applications.

Administration and Data Management

Administrative procedures were similar for both the English and French studies. School packages were prepared at each regional centre by staff experienced in conducting large-scale school surveys. After obtaining the required permissions and assurances of cooperation, packages were mailed to schools with detailed instructions on administration. Materials were returned to the appropriate regional office for checking before forwarding to the central office. Schools were followed up as necessary to ensure that the work was carried out. Machine-scorable answer sheets were used for the tests and student questionnaires, while the teacher and school questionnaires required manual data entry. An array of error checking and editing routines were built into the scoring and data entry procedures to ensure that data were as clean as possible. Master files at both school and student levels were created. Appropriately edited versions of these files were submitted to the international coordinating centre for international analysis, while the national analyses were conducted by members of the regional research teams.

OVERVIEW OF SCIENCE EDUCATION IN CANADA

Science Education and Science Policy

From the outset, it must be recognized that the major structural feature of Canadian education is provincial jurisdiction. While such a feature is common to many federal states, in few instances is this such a strong determinant of educational policy, curriculum, or teaching as is the case in Canada. Most federal states, for example, do have some form of national office of education. Such agencies are typically responsible for national policy, some aspects of funding, research, or other functions. Many states also have national achievement testing programs particularly at the secondary school leaving level. Although in Canada there are several federal agencies with responsibilities which overlap into certain areas of education, and a number of areas in which federal funding is used in education and training, there remains an acute consciousness of provincial responsibility, and a strong sensitivity to any direct national intervention.

In contrast to education policy, science policy in Canada is largely a matter of national jurisdiction. Many areas of federal activity are designed to support scientific research, to ensure an adequate supply of scientific personnel, and to provide specific scientific services to the public and to national agencies. It is widely recognized that there are close links between science policy, economic development, and science education. This may, in part, account for the ability to undertake studies such as SISS or the Science Council of Canada Study, when such activities are not possible in most other areas of education. Further evidence of direct links between scientific and educational policy can be found in the recently announced federal program of undergraduate scholarships for science students. This program complements similar programs which have existed at the graduate level for many years.

Aside from federal involvement in science education, there is evidence of concern with national goals and standards within the Council of Ministers of Education of Canada (CMEC), and interprovincial body which has emerged to promote common interests in education. CMEC has recently embarked on a program designed to develop national indicators in education, including indicators of achievement. While it is not yet clear how this program will evolve, what is evident is an increased concern with interprovincial transferability and with maintaining and improving attainment standards.

Science in the Schools

Almost all schools in Canada include science in the curriculum for all students up to about the end of grade 9. Since attendance at school is almost 100 per cent up to this level, all students can be expected to be exposed to some type of science program. What is less clear is how systematic this exposure is at the elementary level. Nominal inclusion in a provincially mandated curriculum, or even in the timetable of a school, may not be sufficient to ensure systematic, and certainly not uniform, treatment.

At the senior secondary level (in most provinces beginning at grade 10), science is taught mainly as a series of specialist subjects (biology, chemistry, earth science, physics). Courses such as general science or environmental science also exist. Most students in academic programs would take at least some science. Overall, in fact, participation rates in science courses in Canadian senior high schools are quite high relative to those in most other countries. For example, 28 per cent of 18-year-olds in Canada take biology, compared to

12 per cent in Japan, 4 per cent in England, and 6 per cent in the United States. The comparable figures for physics are 19 per cent for Canada and 11, 6, and 1 per cent for the other countries respectively. It should be noted that overall attendance in school at this age level is also substantially higher in Canada than in most other countries, with the exception of the United States. Only the Scandinavian countries approach Canada's participation rates in senior high school science courses.

Science Curriculum

One of the major features of the first phase of SISS/Canada was the conduct of a detailed analysis and comparison of provincial science curriculum documents. An update of this work, including changes up to 1987-88, was carried out by the Newfoundland Task Force on Mathematics and Science Education.[10] Further work, with specific emphasis on textbooks, was undertaken in the Science Council of Canada Study. Detailed accounts of the curriculum work appear in the relevant reports. Only a brief summary will be given here.

The overall curriculum pattern in Canadian schools is one of general science up to about grade 9, with specialized courses being taught thereafter. In some provinces, a general science course may be found up to grade 10. A number of approaches to curriculum can be found in the provincial curriculum documents. These have been labelled in the SISS report as *textbook based*, *content based*, *concept based*, *objective based* and *activity based*. The latter two appear to be most often found in elementary curriculum, while, as might be expected, the first three are characteristic of secondary programs. Most curriculum documents contain various high-minded statements of objectives which go beyond the coverage of content. These have to do with science as a process, the meaning of scientific literacy, science and society issues, and development of scientific attitudes. However, it is not at all obvious that such objectives form an active part of the curriculum as it is finally taught. In particular, since textbooks tend to be content based, there is an absence of concrete materials through which broader objectives can be addressed in the classroom.

Closer examination has shown that science curriculum can be described in terms of three characteristics, which may be labelled commonality, specificity, and prescriptiveness. Commonality refers to the degree to which particular topics can be found in curriculums across the provinces. Specificity refers to the degree of detail in which topics are described in curriculum documents. Finally, prescriptiveness refers to the relative proportions of required versus optional topics. Taken together, these characteristics provide an indication of the degree to which students in different parts of the country and, indeed, in different schools, can be expected to have been exposed to the same material.

The degree of commonality of curriculum varies with level and subject area. In general, programs at the elementary level possess a very low level of commonality across provinces. There is reason to believe that there is even less commonality among schools. Commonality among provinces increases somewhat at the lower secondary level. Because science tends to be treated more formally at this level, greater commonality would also exist across schools, particularly in provinces with highly prescriptive curriculums. At the upper secondary level, commonality is quite high in physics and chemistry, but much lower in biology.

At a more concrete level, commonality is perhaps of most interest in terms of student transfer. What could a student expect to find in moving to a new province, or even to a new school within a province? At the elementary and junior secondary levels, a student up to about grade 6 would find that there is little resemblance between the kind of science taught in different provinces. By the end of grade 9, though not within specific junior secondary grades, a student would have been exposed to somewhat similar content, through rather different underlying curriculum approaches in different provinces. At the end of grade 12, content in physics and chemistry would have been quite similar, but that in biology rather different across provinces.

Turning to specificity, the general pattern again is that the amount of detail specified in curriculum documents increases with grade level. Curriculum at the upper secondary level, in particular, tends to be specified in much greater detail than at the other levels. However, there is no obvious pattern across provinces. As for the subject areas, biology tends to have the most topics specified, followed by physics, and finally by chemistry. While this may reflect the actual diversity of the subject areas, it also suggests that more time would be required for biology than for the other subjects. Alternatively, this might suggest that more optional topics may be found in biology.

The latter point is not strongly supported by the analysis of prescriptiveness. Differences in prescriptiveness are more a function of provinces than of subjects. Overall, the curriculum tends to be more highly prescribed in the eastern provinces, and in Quebec, than in Ontario and the West. Chemistry is relatively highly prescribed in most provinces, while the range of prescriptiveness varies substantially in biology and physics.

Although it is clear that the science curriculum in Canada is far from uniform, it is perhaps surprising that there is any commonality at all, given the independence of provincial jurisdictions, the variation in specificity and prescriptiveness, and the natural tendency towards attenuation of curriculum prescriptions at the level of schools and teachers. One may speculate, of course, that the fundamental nature of scientific knowledge, combined with the codification of such knowledge in textbooks and curriculum documents, contribute to a degree of uniformity of thinking about science curriculum. In any case, there is room for debate on the question of whether increased uniformity of curriculum is necessary or desirable.

Science Teachers and Teaching

In elementary schools, up to about grade 6, science is almost always taught by regular classroom teachers rather than by specialists. Although the overall level of qualification of teachers is high (almost all have at least bachelor's degrees), few elementary school teachers have had any specific training in science at the post-secondary level. Most, in fact, have never taken a science course beyond high school. This is particularly true for women teachers, who dominate the profession at this level. Beginning in the junior high school grades, one finds both increased specialization in teaching assignments and higher educational levels in science subjects on the part of teachers. Male teachers also become dominant at the secondary level.

There are regional differences in the distributions of teachers by sex and by level of qualification. The western provinces generally have the most highly qualified teachers. Ontario has fewer elementary teachers with bachelor's degrees and with specific training in science or science education than other provinces. On the other hand, senior secondary teachers in Ontario are generally the most highly specialized. Most of these differences can be traced to variations in teacher education programs across provinces. As for sex distribution, the difference between elementary and secondary levels is less in the eastern provinces than in the rest of the country. The differences are most extreme in Ontario, where females are clearly dominant in elementary schools, while males are dominant in secondary schools.

Provincial Achievement Testing Programs

Although the SISS studies represent the first attempt to measure science achievement in Canada on a national scale, achievement testing programs are common at the provincial level. Achievement testing within the provinces takes two forms. First, regular public school-leaving examinations are held in some provinces. In addition, most provinces conduct regular testing programs at various grade levels and in a variety of subjects.

Public school-leaving examinations have made a comeback in recent years after being discontinued in most provinces during the late 1960s. Only Newfoundland and Quebec maintained a full program of public examinations throughout the 1970s. These two provinces still operate the most comprehensive systems (over 30 subjects are examined in Newfoundland and more than 100 are examined in Quebec). Since the early 1980s, however, such examinations have been reinstated in the main academic subjects in Alberta and British Columbia. Examinations in mathematics and English have also been introduced in New Brunswick, although these still do not have the status of requirements for high school completion. Saskatchewan maintains a partial public examinations system for students taught by non-accredited teachers.

Public examinations are designed primarily to provide data on individual student attainment and to certify students for high school graduation. Data are generally not compiled in a way which allows claims to be made about the performance of the system as a whole, although levels of success on public examinations may be perceived by the public as a mark of success or failure of the system. In any case, public examinations as currently designed can provide no basis for interprovincial comparisons.

Unlike public examinations, regular provincial assessment programs at other levels are usually intended explicitly to provide system-level rather than student-level data. The particulars of such testing programs vary widely from province to province. In general, the smaller provinces tend to rely on standardized tests (the Canadian Test of Basic Skills is the most widely used test), while larger provinces have developed their own testing instruments. British Columbia has perhaps the most sophisticated system, with various subjects being tested at several grade levels on a cyclical basis. Science assessments, for example, were conducted in 1978, 1982, and 1986 in each of grades 4, 7 and 11. A similar system is in place in Alberta. Ontario has developed a unique system, consisting of *instrument pools* (or item pools) which are made available to local jurisdictions and which have also been used to give provincial estimates based on samples of students. Some provinces, such as Manitoba and Nova Scotia, base their testing on samples of students. This has the advantage of being economical, but allows reporting of results only for large units such as school districts or the province as a whole.

Virtually no comparative data are available from such tests. The main exception is for the Canadian Test of Basic Skills, where national norms are available. Unfortunately, the CTBS has no science component. British Columbia has now had sufficient experience to allow comparisons over time.

These suggest that a substantial improvement in science performance has occurred since the tests were first introduced. British Columbia has also made some comparisons with programs such as the National Assessment of Educational Progress in the United States. Reciprocal testing on a small scale was conducted between Newfoundland and British Columbia in 1988. In this program, a grade 6 test developed in Newfoundland was used in British Columbia, while the British Columbia grade 7 test was used at the beginning of grade 8 in Newfoundland. Results were consistent with data from other sources, including SISS, indicating that Newfoundland students lag somewhat behind those in British Columbia.[11]

Validity and Value of Interprovincial Comparisons

The diversity in curriculum found in Canada leads to a serious question of whether interprovincial achievement comparisons can have any meaning. In fact, the study was not specifically designed to allow such comparisons to be made. Sampling and testing were carried out by region, with provincial samples being designed to fit regional requirements for representativeness. After much debate, the research group was convinced that the content of the SISS tests did represent a common core of science knowledge which could be accepted as representing minimal expectations for attainment of students at the various population levels studied. This is not to say that the tests were fully representative of provincial curriculum descriptions, much less of all important objectives of science education. The major limitations of the tests were, however, considered to be ones of omission, rather than inclusion. Whatever their limitations, therefore, the tests were not considered as unfairly biased. In fact, taken as a whole, the tests can be considered to have substantial face validity in terms of common areas of science knowledge.

Part of the reason for reconsidering the original intention not to produce provincial comparisons is that the level of interest in such comparisons, and in test results in general, has increased in recent years. As described above, a number of provinces now have both provincial school-leaving examinations and testing programs at various levels of schooling. Consideration is being given to a national testing program as a basis for the award of national undergraduate scholarships. In addition, certain reports of the regional comparisons already made were somewhat misleading in terms of the conclusions drawn about the quality of science education in particular provinces. Overall, therefore, it seemed more reasonable to have a full report, complete with appropriate caveats about the value and limitations of such data, than to allow conclusions to be drawn beyond the intentions of the researchers.

In the final analysis, however, the value of interprovincial comparisons must rest on two considerations. First, there must be some sense that education is a matter of national, rather than strictly provincial, concern. Issues of student transfer, post-secondary admissions, national competition for entry level jobs, and the like are all concrete examples of where there is a legitimate national concern. In the case of science, a broader national interest can be identified, based on national science policies and the need for highly qualified scientific and technical personnel. In the latter case, there is clearly a national market, as well as a national policy interest.

The second consideration is that such comparisons are useful only to the extent that they can be used to highlight problems and to stimulate action to solve these problems. Here, it is necessary to reiterate that test performance is only one indicator of the performance of the system, and that test scores must be interpreted carefully in relation to broad goals and specific curriculum content.

OVERVIEW OF INTERNATIONAL RESULTS

International Populations and Samples

As described earlier, the international study defined three population levels for testing. These were at the upper elementary, junior secondary, and last year of secondary education. Within countries, there were options of choosing age or grade levels in defining these populations. In practice, grade level populations were most often chosen for reasons of ease in assembling students for test administration. Specifically, Population 1 was defined as the 10-year-old age cohort or the grade level at which most 10-year-olds would be found; Population 2 was defined as 14-year-olds or the corresponding grade group; and Population 3 was defined as those in the last year of secondary education in each country (both age and grade levels varied across countries for this population).

Population sizes at the first two levels corresponded roughly to the total relevant age groups (10- and 14-year-olds) in each country, since in most participating countries almost all children are in school up to this level. The size of Population 3 varied with the proportion of the age group in school and taking the subjects of interest in the study. Participation rates in science courses at the senior secondary level are, in fact, of substantial importance in interpreting the achievement data, a point which will be taken up in a subsequent section.

Overall sample sizes for the various countries are reported in Table 2. Except at the Population 3 level, these samples are relatively large in comparison, for example, to the sample sizes normally used in public opinion surveys. However, because the samples were generally selected as class groups rather than randomly at the individual level, the sampling errors in the achievement data reported tend to be of the same order as those for national surveys of other kinds.

Table 2

Sample Sizes for Participating Countries

	Population 1	Population 2	Biology	Chemistry	Physics
AUS Australia	4,259	4,917	1,631	1,177	1,073
CAE Canada (English)	5,151	5,636	3,409	3,110	2,892
CAF Canada (French)	2,738	2,350	249	1,187	944
ENG England	3,748	3,118	884	892	917
FIN Finland	1,600	2,546	1,652	871	810
HKG Hong Kong	5,342	4,973	3,614	3,670	3,679
HUN Hungary	2,590	2,515	301	143	398
ITA Italy	5,156	3,228	147	217	1,766
JAP Japan	7,924	7,610	1,212	1,468	1,187
KOR Korea	3,489	4,522	-	-	-
NET Netherlands	-	5,025	-	-	-
NOR Norway	1,305	1,420	276	283	443
PHI Phillipines	16,851	10,874	-	-	-
POL Poland	4,390	4,520	764	765	1,716
SIN Singapore	5,547	4,430	902	945	1,071
SWE Sweden	1,449	1,461	1,232	1,172	1,156
THA Thailand	-	2,519	-	-	-
USA U.S.A.	2,822	3,780	659	537	485

Summary of Achievement Results

Chart 1 shows the Population 1 ranking of countries for which data were available for the preliminary international report, with the Canada (English) and Canada (French) positions highlighted by the black bars. As can be seen, Canada's position varies substantially with population level. For Population 1, French and English Canada are ranked fifth and seventh out of the 16 countries represented. Only 4 percentage points separate French Canada from Japan, the top ranked country. The comparable difference for English Canada is seven points.

At the Population 2 level, shown in Chart 2, somewhat larger differences are observed. Below the top three countries lies a group of eight others, including both English and French Canada, with very similar scores. Although English Canada is ranked fourth and French Canada ninth, the score difference between the two is, in fact, smaller than for Population 1. The spread between both Canadian populations and Hungary, the top-ranked country, has increased to 10 percentage points.

Chart 1

Achievement Levels in SISS Countries, Population 1

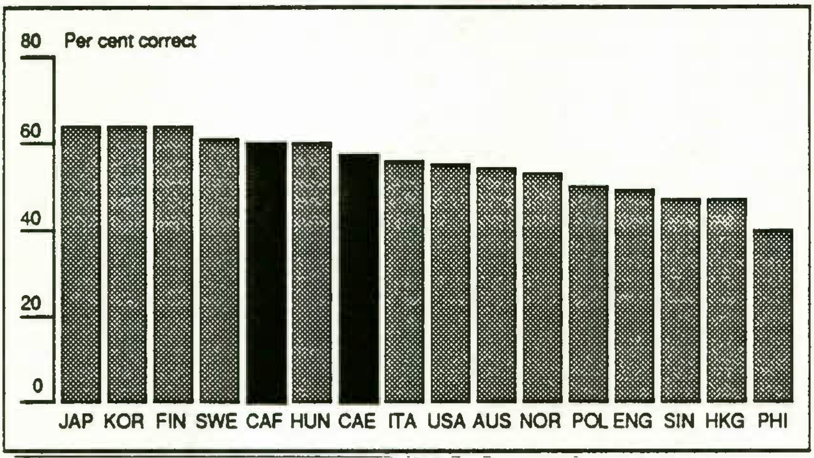
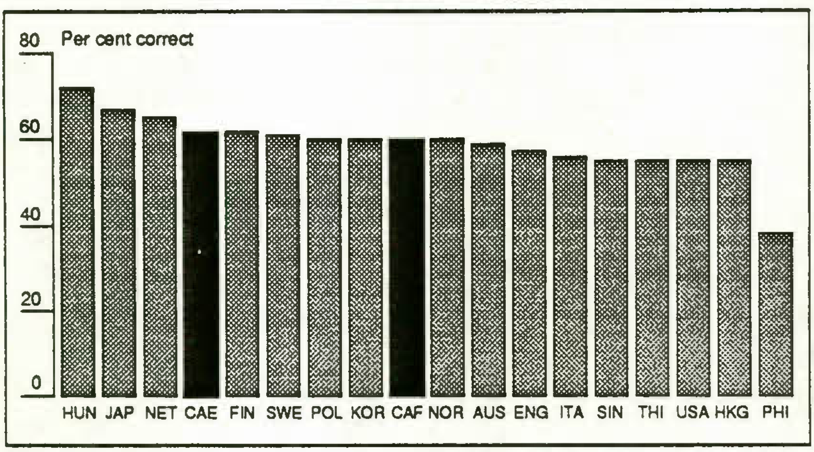


Chart 2

Achievement Levels in SISS Countries, Population 2



Four separate tests were used at the Population 3 level. A 30 item core test was administered to all students, while specialized tests in biology, chemistry, and physics were used for students taking these subjects. As shown in Charts 3 through 6, a marked shift in rankings of countries occurs at the Population 3 level. Three countries, namely Hong Kong, Singapore, and England, which had been near the lower end of the scale for Populations 1 and 2, appear at the top for almost all tests in Population 3. Furthermore, the differences between these and other countries are much larger at the Population 3 level, particularly in chemistry. Overall, the relative positions of countries are similar for all measures, with one or two notable exceptions (such as the relatively low performance of Japan on the biology test). For both Canadian populations, the ranking drops to near the bottom of the scale on all measures, with the differences between Canada and the highest performing countries being much larger at this level than for the earlier populations. The performance of French Canada is also substantially below that for English Canada on all measures.

Age, Participation, and Achievement

The most striking feature of the achievement results is the dramatic shift in the rankings of countries, and in the magnitude of score differences, for Population 3 compared to Populations 1 and 2. In particular, the position of both Canadian groups shows a decline from near the top of the rankings to very near the bottom, with more than 30 percentage points separating both Canadian populations from the highest scoring countries. This leads to the question of why the differences are so large and, in particular, why Canadian students at the senior secondary appear to fare so much worse than their counterparts in other countries.

The complexity of the definitions for Population 3 makes it difficult to arrive at a simple explanation for the drastic shifts in performance observed. More detailed analysis currently being undertaken internationally may eventually shed considerable light on differences between educational systems at this level. The information currently available allows us to point to one or two obvious features which might be expected to influence achievement levels. The first is age, which might be expected to be related to exposure to schooling, intellectual maturity, or other variables which might affect achievement. The second is participation rates, which, in turn, is related to specialization in the various subjects and to selectivity in access to particular courses.

Chart 3

Achievement Levels in SISS Countries, Population 3, Core Test

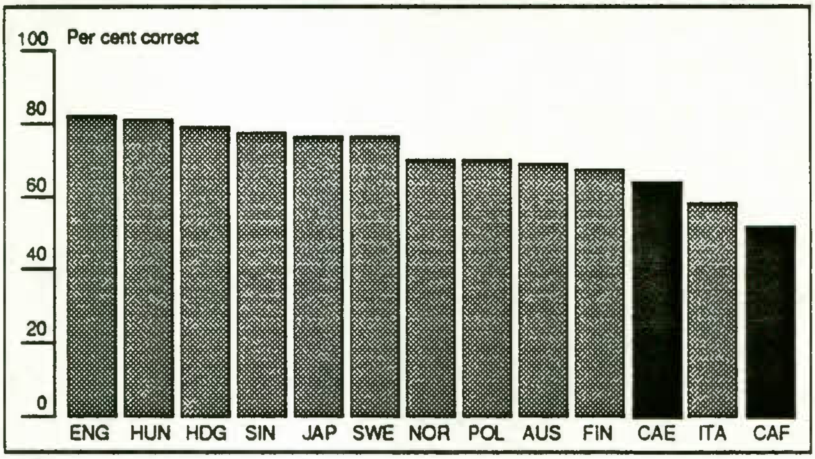


Chart 4

Achievement Levels in SISS Countries, Population 3, Biology

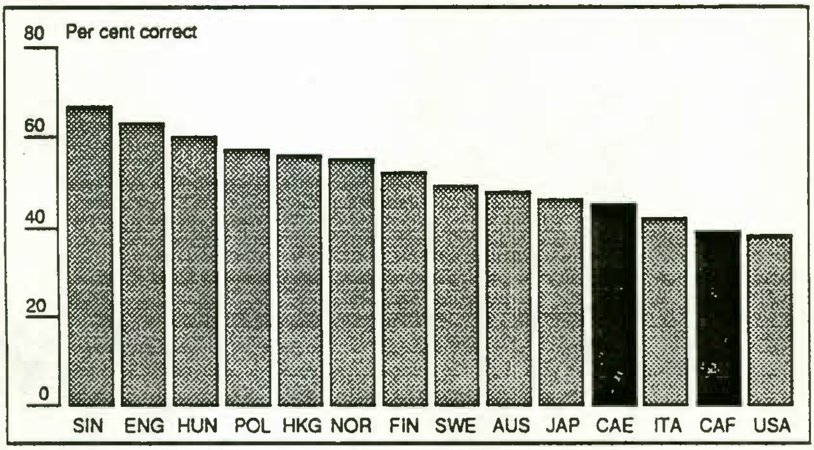


Chart 5

Achievement Levels in SISS Countries, Population 3, Chemistry

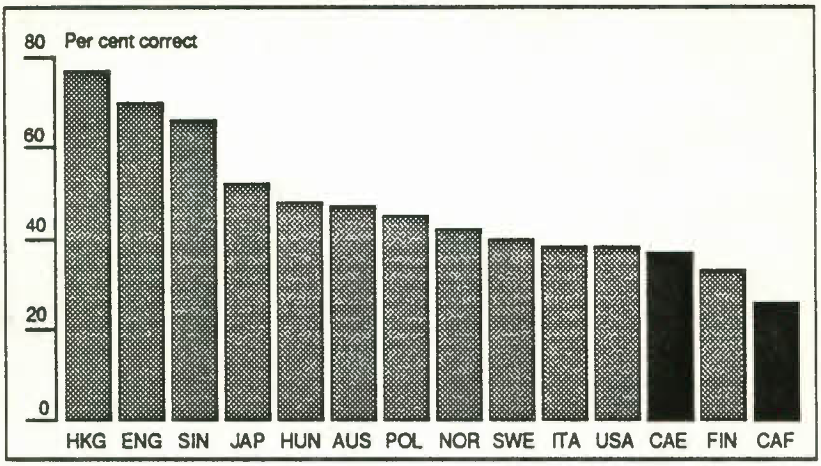
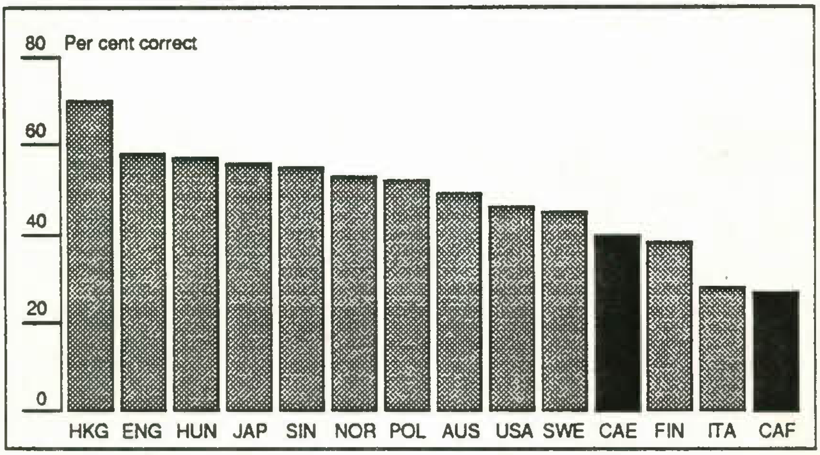


Chart 6

Achievement Levels in SISS Countries, Population 3, Physics



Data on ages of students and participation rates in the various subjects for Population 3 are given in Table 3. As can be seen, ages of students sampled varied by nearly two years, from 17 years 2 months in French Canada to 19 years 0 months in Italy and Sweden. English Canadian students were in the middle of the range, at 18 years 3 months. Overall participation in schooling varied widely across countries. However, it is notable that only a few countries have participation rates higher than 50 per cent at this level. The highest overall rates are in the United States and English Canada, with 90 and 71 per cent respectively. Several countries have rates lower than 20 per cent. As for enrolments in the specific subjects, these are generally much lower than overall participation rates, reflecting the existence of options in the curriculum of almost all countries at this level. Canadian participation rates are among the highest in all sub-jects, most notably in chemistry, where the rate is substantially higher than in any other country. It is interesting to note that despite having the highest overall participation rate, the United States ranks very low in the proportions of students taking the specialized science subjects.[12]

Table 3

Ages and Participation Rates for Population 3

	Percentage in school	Mean age	Percentage of age group taking course		
			Biology	Chemistry	Physics
		(Years:months)			
AUS Australia	39	17:3	18	12	11
CAE Canada (English)	71	18:3	28	25	19
CAF Canada (French)	67(79)*	17:2	7	37	35
ENG England	20	18:0	4	5	6
FIN Finland	45(63)*	18:7	45	14	14
HKG Hong Kong	20	18:4	7	14	14
HUN Hungary	18(40)*	18:0	3	1	4
ITA Italy	52	19:0	14	2	19
JAP Japan	63	18:2	12	16	11
NOR Norway	40	18:11	10	15	24
POL Poland	28	18:7	9	9	9
SIN Singapore	17	18:1	3	5	7
SWE Sweden	15(30)*	19:0	15	15	15
USA U.S.A.	90	17:7	6	1	1

* The figures in parentheses include students in vocational or similar streams which were not sampled.

How does all of this relate to achievement? Correlations between age and achievement were essentially zero, indicating that age is probably unrelated to exposure to the relevant subject matter. The fact that students in the French Canada sample were the youngest of all the participating groups does reflect, however, the important fact that most of these students were in grade 11 in Quebec, where secondary education ends at this level. In this instance, grade level rather than age is likely an important factor in accounting for the difference between the English and French Canadian results.

Correlations between participation rates and achievement were generally negative (-.28, -.35, and -.48 for biology, chemistry, and physics respectively),[13] suggesting that the higher the participation rates the lower the average achievement scores. If one associates low participation rates with a high level of selectivity in the system, such correlations are what might be expected. It is quite likely, for example, that the high achievement scores for England, Singapore, and Hong Kong, and the major change in ranking of these countries from the lower levels to the Population 3 level, is associated with the high level of selectivity of the British "A" level courses and their counterparts in Singapore and Hong Kong.

Obviously, however, the same line of reasoning cannot be applied to the United States. In fact, the diversity of American secondary programs suggests that choice of course is more a matter of interest or similar factors than of the sort of selectivity practised in some other countries. Were it not for the United States data, the negative relationship between participation and performance would be much stronger.

As for Canada, the situation is similar to the United States, in the sense that programs are not designed to be highly selective. Given Canada's[14] high participation rates, low average performance compared to that for countries with selective programs is not surprising. One might speculate that the major difference between Canada and the United States is that students within a fairly wide range of abilities elect to take the specialized science courses, but that many more such students are attracted to these courses in Canada than in the United States.

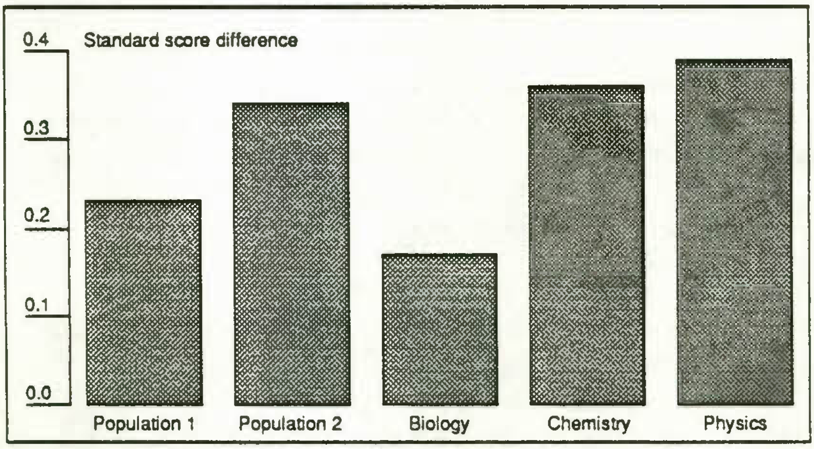
Gender Differences in Achievement

Studies have consistently shown that boys outperform girls on science tests. For example, the First International Science Study[15] reported standard score differences[16] from .23 at Population 1 to .46 at Population 2 to .69 at Population 3. Not only were gender differences consistent across grade levels and countries, but also these differences increased with level of schooling.

The results for SISS were no exception to the general pattern. Chart 7 shows the mean standard score percentage differences between boys and girls for the main tests at all population levels. As for the first international study, the differences are observed to increase from Population 1 to Population 2, although the Population 2 differences are somewhat less here than in the first study. At the Population 3 level, the differences depend strongly on subject, being greatest for physics and least for biology. At the same time, none of the observed differences are as large as those reported in the first international study.

Chart 7

Gender Differences in Achievement, International



It is important to note that the observed differences varied substantially from country to country, although there were only two cases out of about 60 comparisons for which girls outperformed boys. These were for biology in Hong Kong, form 7 (0.18), and Sweden (0.23). For Population 1, the differences ranged from nearly zero in Japan and the Phillipines to 0.44 in Norway. Both Canadian groups were near the middle of the range, at 0.27. At the Population 2 level, most countries showed significant differences, with Canadian students being at the high end of the range, at 0.37 for English and 0.47 for French Canada.

Curriculum and Achievement

It is reasonable to expect that achievement would be associated with the degree to which the subject matter covered by the achievement test is actually included in the curriculum of each country. Three separate indices were used internationally to measure the relationship between the content of the tests and the content of the curriculum. The *Curriculum Relevance Index* measured the match between test items and curriculum content in a particular country. The *Test Relevance Index* was a variation of the first index, based on the number of test items representing each curriculum area within a country. Finally, the *Curriculum Coverage Index* gave a measure of the extent to which the curriculum of a country covered the complete content grid used in constructing the tests. Only preliminary relationships between these indices and achievement can be given here, as the necessary multivariate analysis has not yet been completed.

Each of the curriculum indices was correlated with achievement test scores across countries. As might be expected, a fairly complex pattern of relationships emerged. These are summarized in Table 4. At the Population 1 level, the curriculum indices were relatively low, and the correlations with achievement were negative. This suggests two things: first, that curriculum coverage is somewhat erratic, and second, that student knowledge at this level may be acquired from sources other than school.

The Population 2 curriculum indices were slightly higher and somewhat less variable across countries than those for Population 1. The correlations were positive, though not particularly high. This likely reflects the beginning of more systematic formal instruction at this level. It also indicates that higher levels of curriculum emphasis are associated with higher achievement. It is important to note, however, that these results may reflect no more than the limitations of the tests in capturing important elements of curriculum in some countries. For example, it is clear that the science curriculum at this level in several countries shows an emphasis on science as a process, rather than science as a body of facts and concepts, as reflected in the SISS curriculum grid and test items.

For Population 3, the curriculum indices were generally much higher, exceeding 90 per cent on average for the three specialized subjects. The variation among countries was also much smaller. This clearly suggests increased standardization of curriculum at this level, and increased congruence of national curriculums with the SISS grid. Correlations between the curriculum indices and achievement were, however, generally no higher than for Population 2. In part, this is a reflection of the lack of variability in curriculum coverage. However, as already indicated, the picture at this level is complicated by variations in participation rates, levels of specialization, and no doubt other factors, which may mask the effects of curriculum.

Table 4

Correlations of Curriculum Indices with Achievement

	Curriculum relevance		Test relevance		Curriculum coverage	
	Percentage	Correlation	Percentage	Correlation	Percentage	Correlation
Population 1	56	-.19	76	-.37	43	-.43
Population 2	58	.38	87	.38	68	.30
Population 3 Biology	76	.43	99	–*	97	.43
Population 3 Chemistry	89	.34	91	.33	87	.21
Population 3 Physics	89	.39	97	.42	91	.13

* Coefficient could not be computed because of lack of variance in test relevance data.

CANADIAN PROVINCIAL COMPARISONS

Sample Summary

A summary of final sample sizes achieved in each province for the English study and for the total French study are given in Table 5. Several points should be noted. First, it is clear that sample sizes at the provincial level are relatively small. The impact of this on sampling error has been discussed in an earlier section. Second, it was possible to more closely meet the sampling targets for Populations 1 and 2 than for Population 3. The reasons for this are not entirely clear. Absenteeism, smaller class sizes than anticipated, the existence of external examinations, and other factors more prevalent in senior high schools than at other levels were the most likely contributors to sample loss. It should also be noted that the wide differences in sampling fractions made it necessary to compute compensating weights at the school, provincial, and regional levels, when presenting data aggregated to these various levels. Finally, it is important to point out that the Francophone samples were too small in each province to permit provincial breakdowns of the kind provided for the Anglophone samples. Following the procedure used in the report of the French study, the achievement results are given for Quebec and Francophones outside of Quebec. Achieved sample sizes are not known for these two subgroups.

Table 5

Summary of Achieved Sample Sizes by Province

Province/Area	Population 1	Population 2	Biology	Chemistry	Physics
Territories	77	46	33	22	18
British Columbia	518	844	273	320	252
Alberta	615	676	353	319	277
Saskatchewan	414	298	212	227	175
Manitoba	218	236	247	217	262
Ontario	1,749	1,993	1,297	1,277	1,209
Quebec (English)	448	426	305	299	380
Newfoundland	361	441	197	88	168
Nova Scotia	360	423	295	249	206
Prince Edward Island	428	418	266	182	128
New Brunswick	411	264	236	204	197
French	2,378	2,350	249	1,187	944

Summary of Achievement Results

Charts 8 through 13 show the mean scores by province[17] for each population level. For comparison, the international mean is also shown.

For Population 1, the results show fairly small interprovincial differences. For the core test, the differences between the highest and lowest scores are of the order of 11 percentage points, which is quite sufficient for statistical significance.[18] The overall pattern for the rotated forms is similar, though the differences are generally somewhat smaller. In general, the western provinces attained consistently higher scores than either Ontario or the eastern provinces. In fact, considered on a regional basis, where sampling errors are much smaller, these differences are more distinct than the provincial differences themselves.

Turning to Population 2, the differences between provinces are generally slightly larger than for Population 1. What is most important to note is that the pattern of the western provinces having the highest scores, followed by Ontario and then the eastern provinces, generally holds at this level as well as for Population 1. A notable exception is the ranking of Nova Scotia, which approaches that of Alberta, the highest province, and which is not in accord with the general pattern of lower scores for the Eastern region. It is again important to stress that the observed differences, though seemingly fairly large, do not permit a clear ranking of provinces in order of performance, because the differences between adjacent pairs of scores are usually smaller than required for statistical significance. At the same time, a generally consistent pattern across the two levels strengthens substantially the conclusion that the regional differences are real, rather than simply chance sampling errors.

The argument that there are distinct patterns of regional difference is further strengthened by the Population 3 results shown in Charts 10 through 13. In general, the observed differences for Population 3 are found to be substantially larger than those for the two lower levels. The observed patterns are also consistent across the various forms. Nevertheless, a couple of shifts in the pattern deserve to be noted. First, Ontario is closer to the top here than at the other levels, in all subjects except physics. Saskatchewan also occupies an anomalous position on the specialized tests, particularly chemistry and physics, compared to its position on the other measures. Regional differences are particularly notable in physics, where the eastern provinces score consistently well below the others.

Of course, none of this answers the question whether the observed levels of performance are satisfactory. Average scores in the 60 per cent range on multiple choice tests are fairly typical, particularly when the item selection process is based at least partly on normative considerations of item difficulty (i.e., items which are too easy or too difficult when tried on a representative sample are commonly discarded). Some estimate of adequacy can be obtained from the international results. Using this as a guideline, performance at the first two population levels can be considered reasonable, while that at the Population 3 level must be judged highly unsatisfactory.

Chart 8

Achievement Levels by Province, Population 1, Core Test

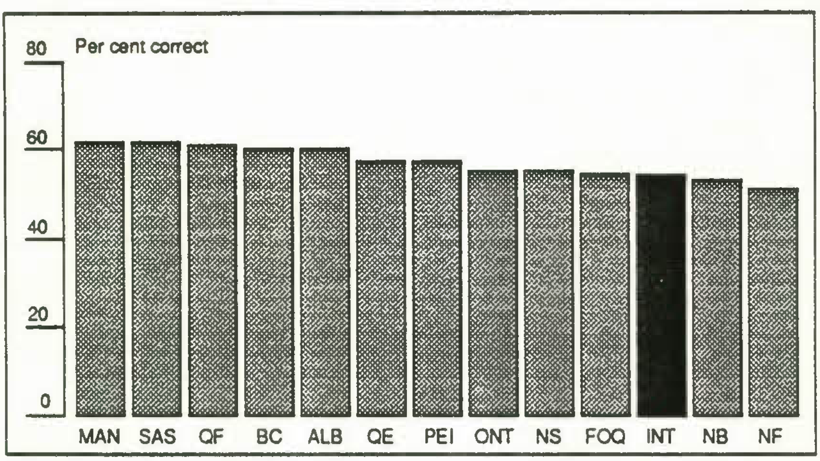


Chart 9

Achievement Levels by Province, Population 2, Core Test

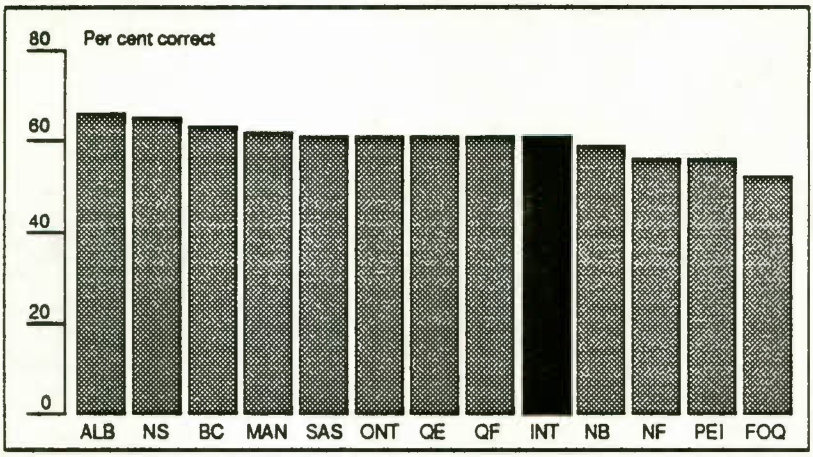


Chart 10

Achievement Levels by Province, Population 3, Core Test

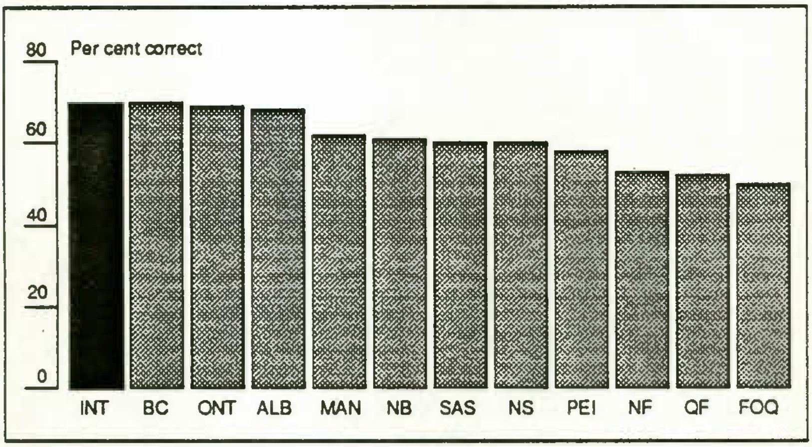


Chart 11

Achievement Levels by Province, Population 3, Biology

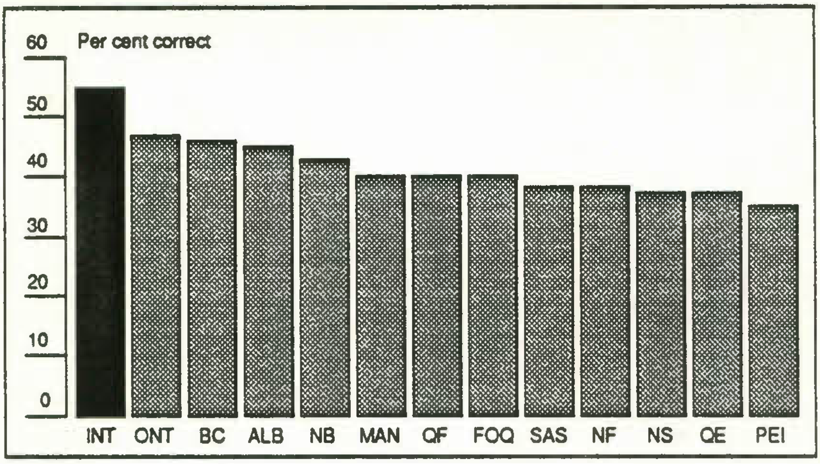


Chart 12

Achievement Levels by Province, Population 3, Chemistry

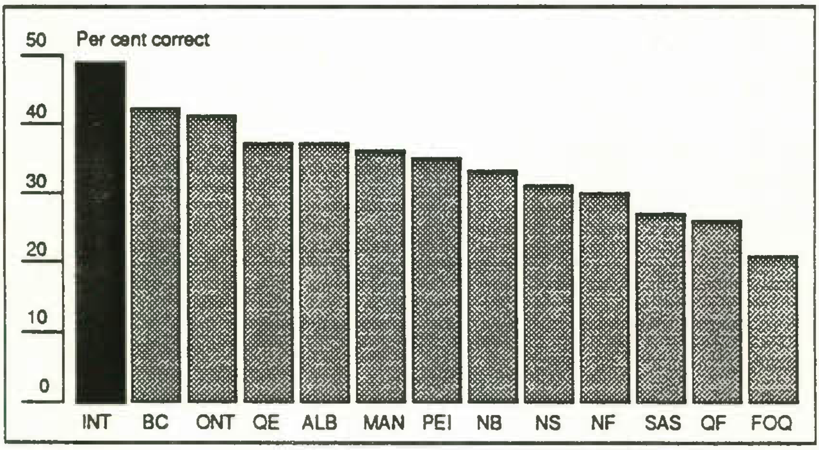
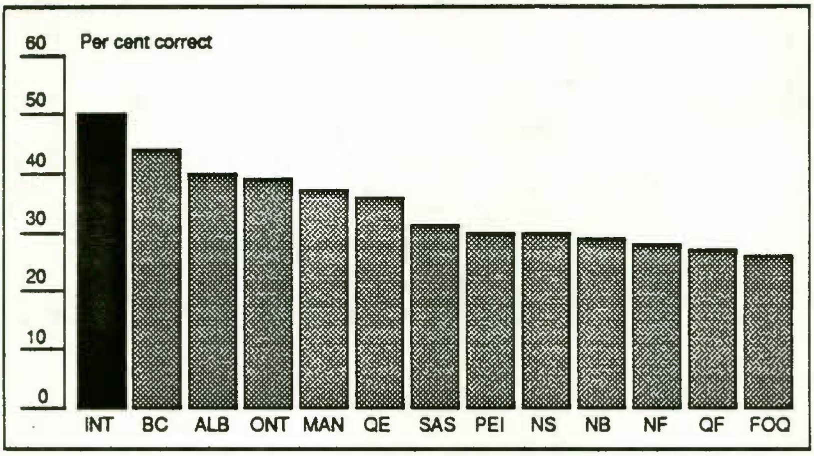


Chart 13

Achievement Levels by Province, Population 3, Physics



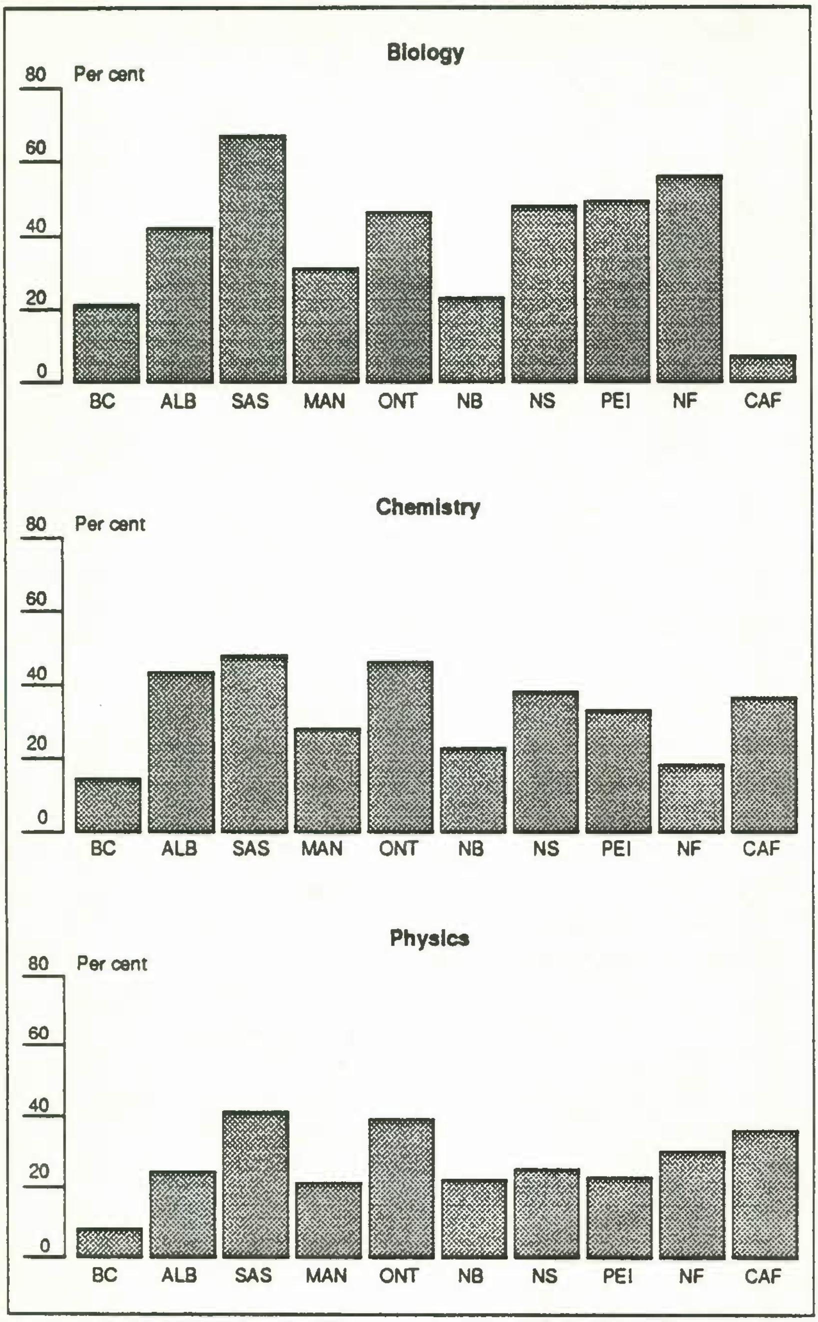
Participation and Achievement

It has already been suggested that the high participation rates already noted for Canada can account for the country's relatively poor Population 3 performance internationally. If this explanation is valid, one would expect that variations in participation rates across the provinces would also account for provincial differences. As Chart 14 shows, the percentage of grade 12/13 students taking the specialized science courses varies dramatically from province to province. Correlations of participation rate with mean achievement scores by province were -.39, -.20, and -.40 for biology, chemistry, and physics respectively. Although these coefficients are not statistically significant (since they are based on just 10 pairs of points), the consistency of direction, and the comparability with international findings, lends some credence to the hypothesis that high achievement is associated with low participation.

Nevertheless, it is clear that the relationship between participation and performance is not a straightforward one. For example, British Columbia and Alberta show similarly high levels of achievement, even though their participation rates vary substantially. Similarly, Newfoundland and Nova Scotia show similar levels of performance in chemistry, even though the participation rates for these two subjects are widely different. Ontario has generally high participation

Chart 14

Participation Rates in Specialized Science Courses by Province



CAF: French Canada, all provinces combined.

and high achievement throughout. On the other hand, Saskatchewan is distinguished by the highest participation rates throughout, but shows relatively low performance.

The hypothesis that differences in participation rates can help account for achievement differences is meaningful, of course, only if one assumes that low participation is associated with high selectivity, such that those taking the courses are the ones most likely to do well. Unfortunately, no direct information is available on this point. In most provinces, course selection is a matter of student or parent choice, with advice from the school. An exception is Ontario, which practises grouping from grade 9. While ability is likely to be one consideration in course selection, there is little evidence that schools engage in any extreme form of ability grouping. At the same time, it is easy to see how the ways in which decisions are made would differ from province to province, in such a way that there might be substantial differences in the degree of selectivity.

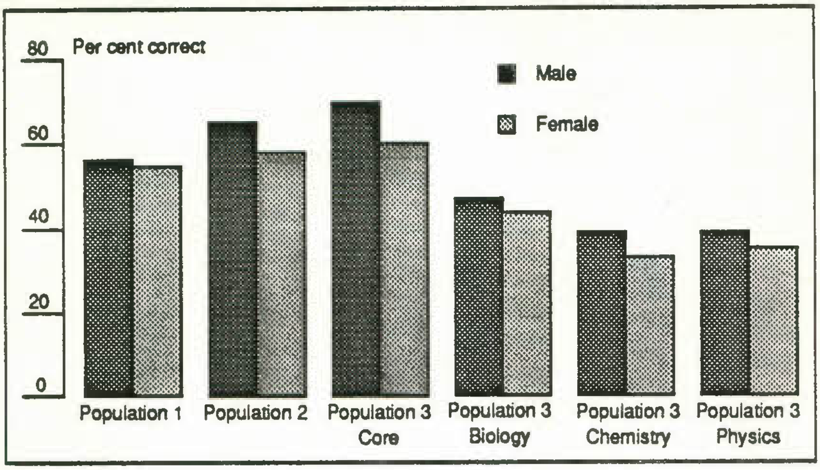
Gender Differences in Achievement

As shown in the previous section, the SISS international results showed differences favouring boys at all population levels and in all countries, with only one or two exceptions, most notably the superior performance of girls in Population 3 biology for a couple of countries.

The Canadian data revealed no exception to this trend. The pattern is illustrated by Chart 15. Several points are worth noting about these differences. First, although provincial differences are not shown, the overall pattern is strikingly consistent across provinces, and across English and French studies, with only minor differences in magnitudes of the differences. Although the differences are relatively small, the pattern makes it extremely improbable that the differences are the result of sampling error. Second, the appearance of such differences at the grade 5 and 9 levels suggests that neither course selection factors nor differential exposure to science can account for the differences. Third, the overall differences increase with grade level, suggesting some differentiation of interest as students progress through the system. At the senior secondary level, of course, one would expect that differential course selection would become a factor. It is not clear, however, what part this factor plays in the observed effects.

Chart 15

Gender Differences in Achievement, Canada



There is some evidence of gender differences in responses to items, based on obvious cases of items which might be differentially attractive to boys or girls. For example, an item at the Population 2 level having to do with measuring ingredients in a recipe was answered better by girls, while a conceptually similar item on a problem in mechanics was answered better by boys. More generally, those items on which girls outperformed boys tended to be in the life sciences areas. Overall differences may thus be related to subtle biases in the tests. At the same time, it is still not clear why item responses would be biased in this way. The most obvious explanation has to do with interest and attitudes. This explanation is reinforced by the positive relationships between attitudes and achievement reported in the next section. Nevertheless, gender differences in attitudes present a much more complex picture than those for achievement, as will be seen in the next section.

It is possible, even likely, that the observed gender effects are associated with subtle behaviours and attitudes within the classroom itself. For example, although it has been argued that there is little differential exposure to science in Canada up to grade 9, this does not mean that girls are not treated differently from boys within the classroom. Classroom process studies provide considerable evidence of teacher differential treatment of boys and girls. While none of this evidence is specific to what happens in science classes, it is not difficult to imagine that the same stereotypes found in society at large are at play in such classrooms. Substantial research is required in this area if we are to develop effective strategies by which the classroom setting can be used to overcome rather than possibly reinforce these stereotypes.

Curriculum, Time, and Achievement

In the previous section, the relationship between curriculum coverage and achievement was discussed in terms of three indicators of curriculum coverage and relevance, compiled for each participating country.

Because of the variation in curriculum across jurisdictions in Canada, the notion of a national indicator of curriculum coverage was not a particularly meaningful one. Within the English Canada study, a more direct and comprehensive approach was taken to analyzing such relationships. This approach was based more directly on the underlying concept of *opportunity to learn*. This concept was defined in terms of teacher ratings of the test items on the following dimensions:

- time spent in teaching the content of the item;
- conceptual level of the item compared to the level of content taught;
- percentage of students who could be expected to answer the item correctly;
- importance of the item to course objectives; and
- level of school at which the content of the item is first introduced.

Because the analysis of the relationships of these dimensions to achievement was carried out at the item level, lengthy summary tables would be required to convey the detailed results.[19] The following points convey the main findings:

- For Population 1, teachers were able to predict with some accuracy student levels of success on items, but the pattern of relationships between achievement and other ratings was no better than expected by chance.

- For Population 2, whether the topic was taught, and the rated importance of content correlated positively with achievement for 13 and 14 items respectively, out of a total of 90 items. This is more than expected by chance, but still not a particularly large number. Again, teachers were able to predict student success with greater accuracy, with positive correlations for 29 items.
- Ratings for Population 3 biology were somewhat better predictors of achievement, with significant positive correlations for 30 per cent of items for time and for importance of content, 37 per cent of items for whether or not taught, and 53 per cent for predicted student success. Rated conceptual level of items was a poor predictor of achievement.
- For Population 3 chemistry, relationships were analyzed only for time devoted to topic. Significant relationships with achievement were found for only 3 of 30 items.
- Population 3 physics showed the strongest relationships between opportunity to learn and achievement. Significant positive correlations were yielded by 26 per cent of items for content taught or not taught, 44 per cent for importance of content, and 59 per cent for teacher ratings of student success. Conceptual level was negatively related to achievement for 35 per cent of items (the negative correlation means that students performed more poorly on items rated by teachers as of higher conceptual level than the content as taught).

In general, the pattern of relationships was consistent with other research findings which suggest that time and curriculum coverage are important variables in achievement. The relative weakness of the relationships may be due, in part, to the low reliability of teacher ratings and to restriction of range of some of the variables (in the extreme, correlations will always be zero when all of the values of one variable are the same, such as when all teachers rate an item as very important). Because of the weakness of the overall relationships, there seemed to be little point in making provincial comparisons, which would result in even less variability in the ratings. A strong argument can be made, however, that more refined measures of time and curriculum variables are needed, so that the reliability of these measures can be comparable to that for achievement. Such refinements can be attained, however, only through direct classroom observations, detailed timetable analysis, or other more intensive (and expensive) means.

STUDENT ATTITUDES

The Importance of Attitudes

Common sense suggests that it is important to approach any area of endeavour with a positive outlook. What is less clear is the effect of attitudes on success. In fact, the relationship between attitudes and achievement is one of these areas in which it is almost impossible to distinguish cause from effect. It may be argued that a positive attitude is a necessary condition for high achievement. However, it is just as plausible to suggest that success in an area is likely to engender positive attitudes. Going a step further, it may be argued that having a positive attitude is desirable in itself. Science teachers frequently use the argument that one of their main goals is to promote positive attitudes towards science. Many statements of the broad objectives of schooling are more likely to be concerned with positive attitudes than about high achievement. Similarly, it is common to find statements about scientific attitudes in science curri-culum guidelines. Whatever the specific reasons, it is clear that attitudes are generally regarded as important. Although the measure-ment of attitudes presents some problems, it is common to include such measures in any comprehensive survey of the outcomes of schooling.

Attitude Scales[20]

Recognizing the importance of attitudes, students at all population levels were administered an attitude scale. Scales were in the commonly-used Likert format, in which students were asked to respond to particular propositions using a conventional five point scale, from strongly disagree to strongly agree.

Analysis of the attitude items revealed that the items could be conveniently grouped into four scales corresponding to 1) attitudes towards school, 2) importance of science, 3) science in school, and 4) science as a career. The latter scale was not included in the Population 1 instrument. Otherwise, all scales appeared at all levels, although the items were not exactly the same in all cases.

In interpreting the bar graphs shown for the attitude scales, it should be kept in mind that the scale values shown represent averages over several items and several hundred students. Thus, only small differences are required for statistical significance. Any such differences highlighted in the discussion are significant beyond the .99 level of confidence.[21] Furthermore, such differences occur in clearly defined patterns across scales and grade levels, thus adding strength to the conclusion that these are not merely a result of sampling errors. Nevertheless, it is important to recognize that the overall differences are not large, and that the general pattern is one of reasonably positive attitudes throughout.

Attitudes Towards School

The results for *Attitude Towards School* for the three population levels appear in Chart 16. Gender differences are also reported in this chart. The most obvious feature of these graphs is that overall attitudes are relatively positive (a mean value of 3.0 on the scale represents a neutral attitude). A slight drop appears between grade 5 and grade 9. However, this drop is reversed by the senior secondary level. Interprovincial differences are, in general, fairly small. Analyzed at the regional level, however, statistically significant differences, favouring those in the eastern region, occur for grade 9 only. Sex differences are somewhat more pronounced, and are consistently in favour of girls at all levels.

Importance of Science

Chart 17 shows the patterns of response for the *Importance of Science* scale. The overall picture is one of more positive attitudes here than for the previous scale. The differences between provinces remain quite small. Unlike the attitude to school scale, there is no drop in the overall mean scale values at the Population 2 level. Sex differences are not as strong nor as consistent here as for the previous scale. Except perhaps for grade 12, where there are small differences favouring boys, no clear pattern of sex differences emerges. Regional differences again favour those in the eastern region, at both the grade 9 and grade 12 levels.

Chart 16

Attitudes Towards School by Province and Gender

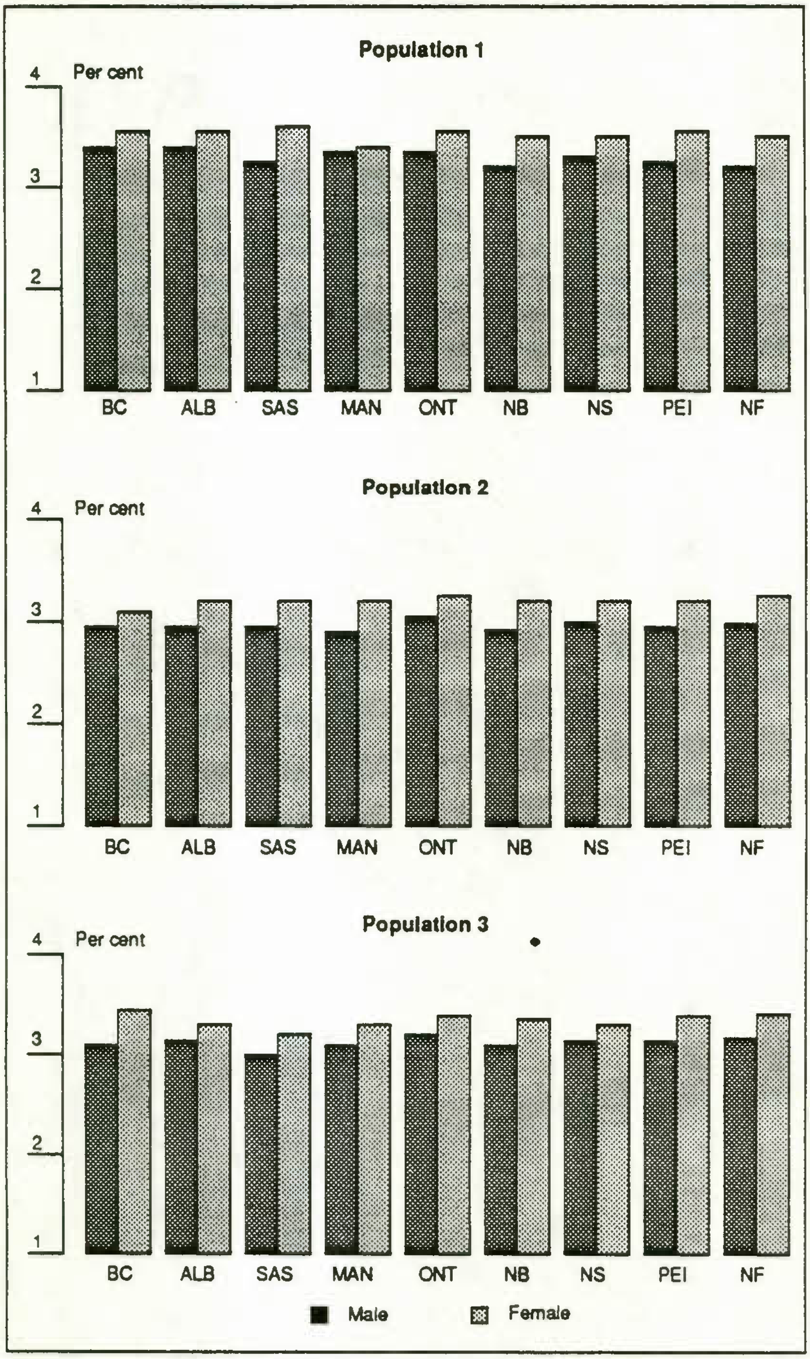
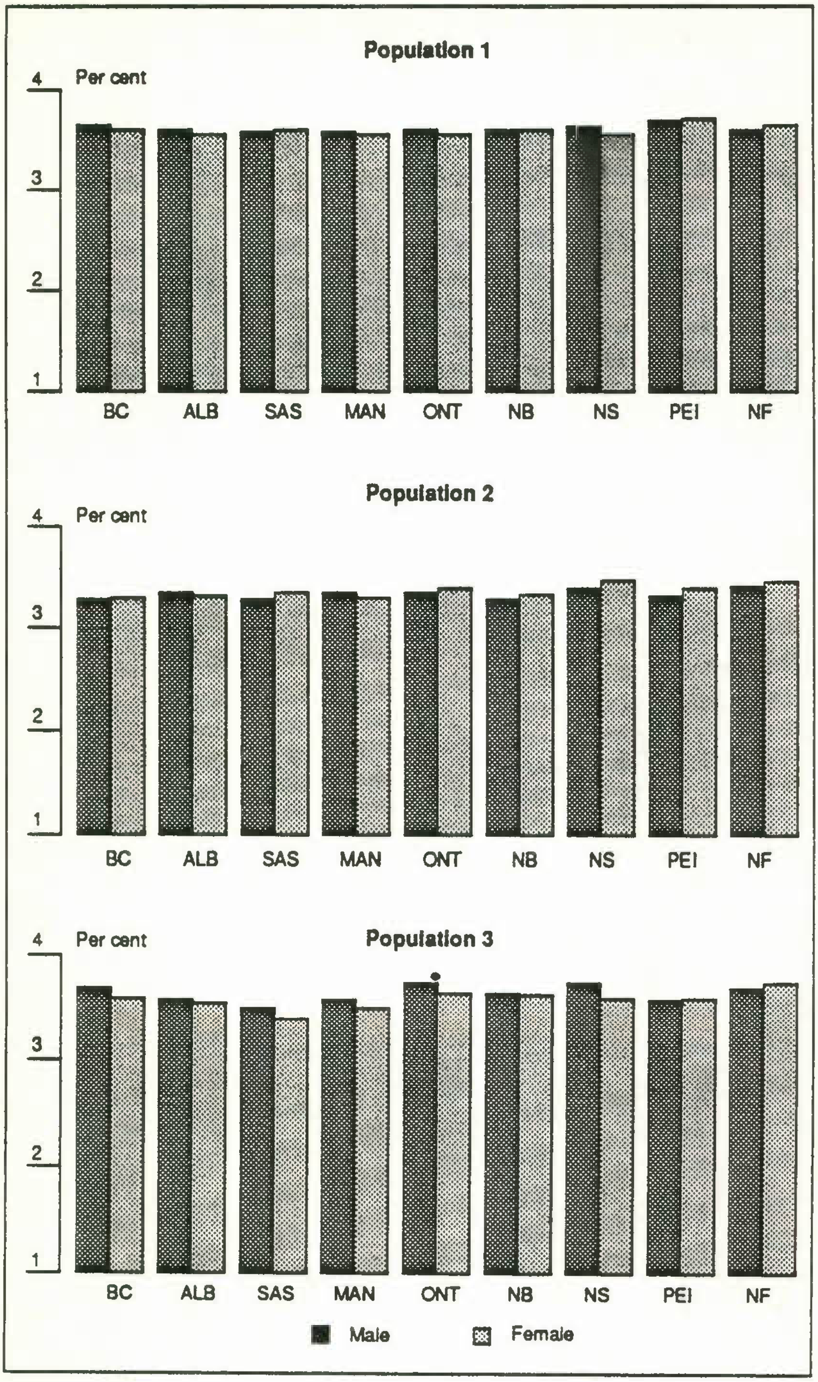


Chart 17

Importance of Science by Province and Gender



Science in School

Chart 18 gives the results for the *Science in School* scale. Overall, attitudes towards science in school closely resemble those for importance of science. The main difference is that gender differences for the science in school scale tend to favour males throughout. Although these differences are small, their consistency across levels and provinces is striking. The results indicate that boys see themselves as having more positive school science experiences than girls. While there is no clear causal link between these perceptions and the similar pattern for achievement, the results do suggest that sex differences are quite pervasive. Small regional differences, again favouring the eastern region, appeared for grades 9 and 12.

Science as a Career

The final attitude scale had to do with student views on *Science as a Career*. This scale was administered only in grades 9 and 12. As Chart 19 indicates, the pattern of positive views and of more favourable responses for boys compared to girls also holds for this scale. Regional differences are also consistent with those on previous scales.

Attitudes and Achievement

Table 6 shows the correlations between the four attitude scales and achievement at the various grade levels. (Provincial breakdowns are not given here because the concept of provincial differences in such relationships is not a particularly meaningful one.) As can be seen, the pattern is one of fairly small but consistently positive relationships. It is important to reiterate here that correlational relationships, in themselves, are insufficient to allow us to answer the question of whether positive attitudes are a prerequisite for higher achievement or whether the opposite is the case. The relationships between attitudes and achievement shown here are comparable in magnitude to other similar relationships between educational variables (for example, similar size correlations are typically found between time and achievement, curriculum coverage and achievement, and so on). However, such relationships are smaller in magnitude than those usually found for ability or socio-economic status and achievement.

Chart 18

Science in School by Province and Gender

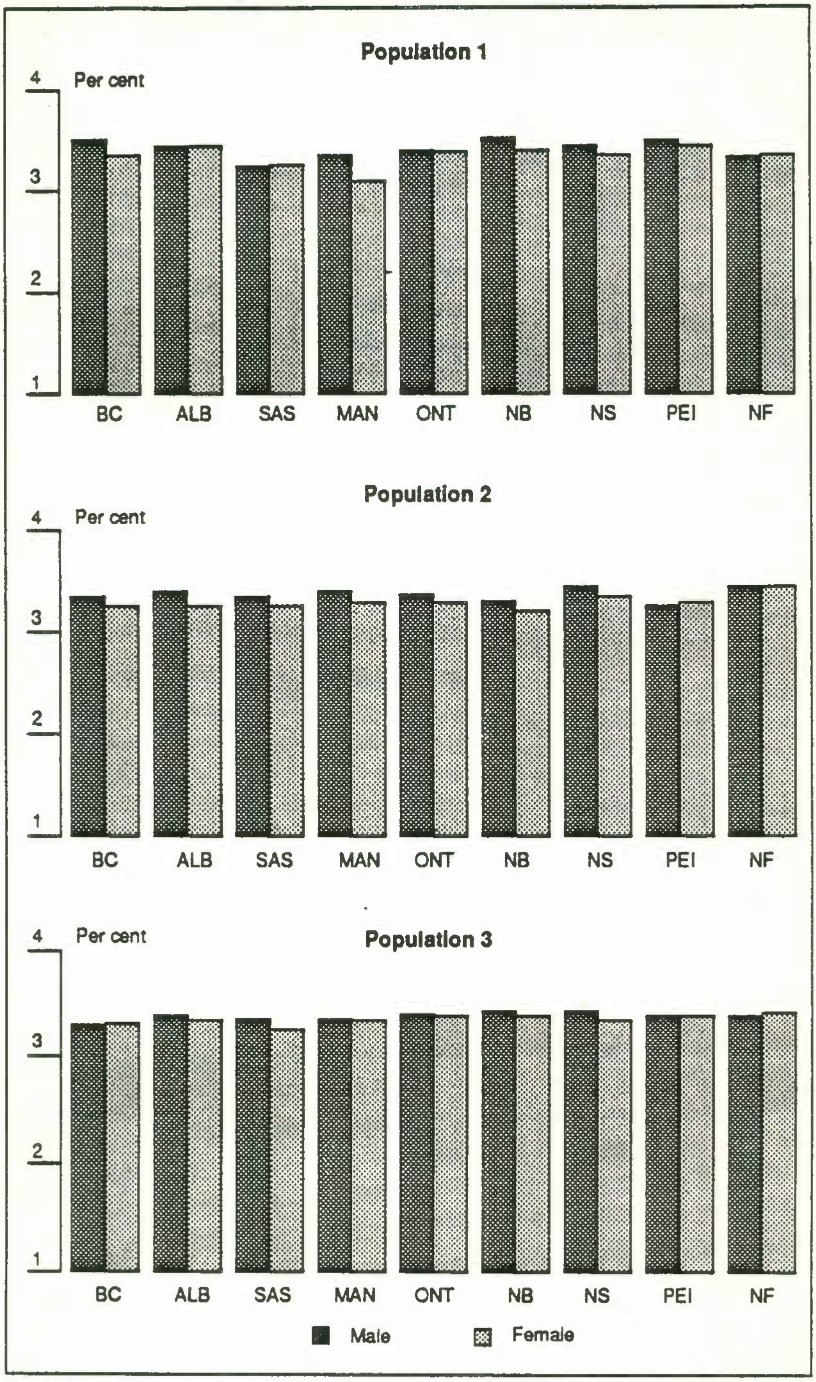


Chart 19

Science as a Career by Province and Gender

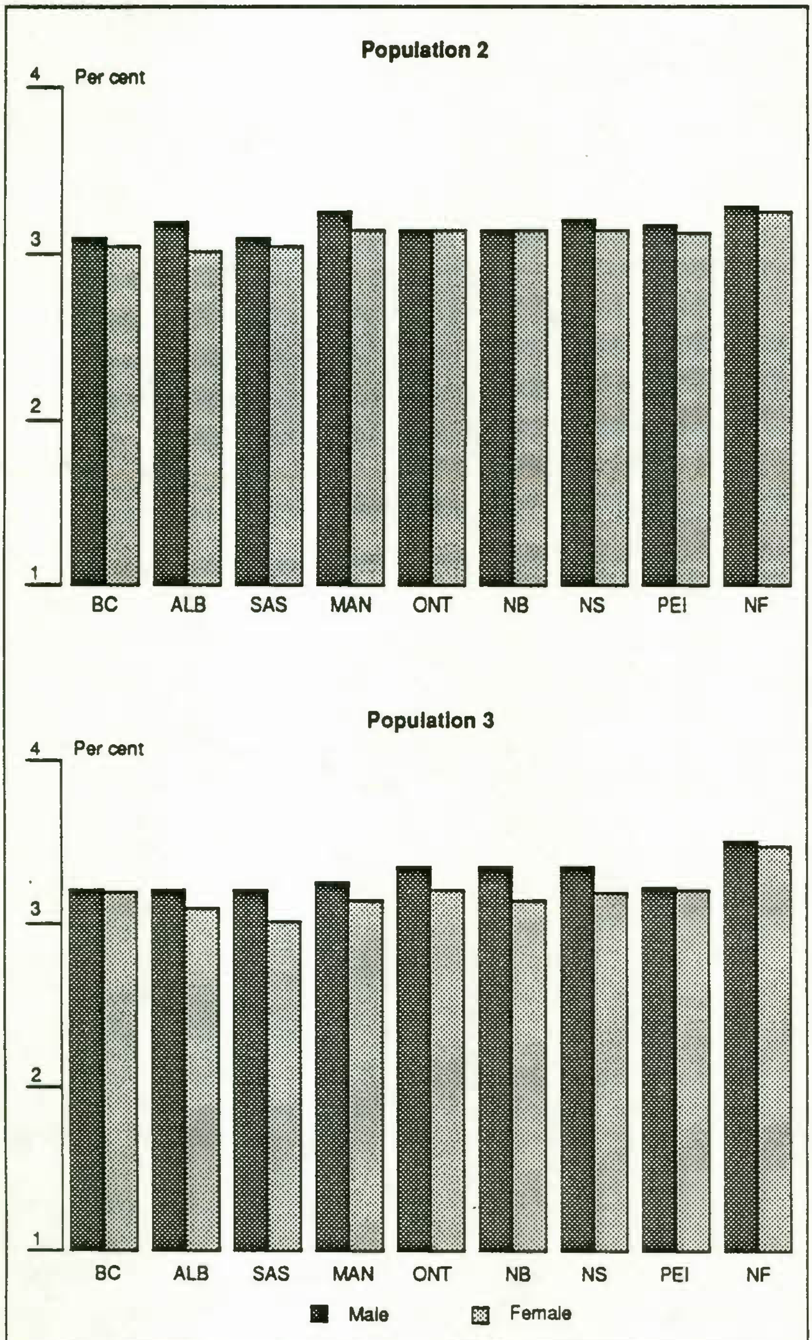


Table 6

Correlations of Attitudes with Achievement

Attitude scale	Grade 5	Grade 9	Grade 12/13 Biology	Grade 12/13 Chemistry	Grade 12/13 Physics
Attitudes towards school	.15**	.11	.09	.25***	.08
Importance of science	.10	.27***	.20**	.36***	.24**
Science in school	.04	.07	.02	.19**	.04
Science as a career		.29***	.30***	.24***	.04

* $P < .05$
** $P < .01$
*** $P < .001$

SUMMARY, CONCLUSIONS, AND IMPLICATIONS

Purposes

The main purpose of this report has been to present an overview of the English and French Canadian contributions to an international study of science achievement, sponsored by the International Association for the Evaluation of Educational Achievement. Emphasis has been on Canada's performance relative to other participating countries, and to more detailed comparisons across provinces and between Anglophone and Francophone groups. The purpose of these comparisons is not one of congratulation or condemnation, but rather to use such data to help identify factors which contribute to achievement, and ultimately to find ways in which achievement can be improved. While large-scale surveys are limited in their ability to pinpoint underlying causes, they can serve as a starting point for more intensive investigations or for policy deliberations.

Results of the study were reported under four main headings: 1) The Canadian Context of Science Education, 2) International Achievement Comparisons, 3) Provincial Achievement Results, and 4) Attitudes. The major findings are summarized briefly in the following sections.

The Canadian Context

- The exclusive jurisdiction of the provinces over education, and the inherent diversity which this creates, is perhaps the essential characteristic of education in Canada. This diversity makes it extremely difficult to present a coherent picture of Canadian science education, especially for an international audience. The need for separate Anglophone and Francophone studies was the most obvious example of how diversity influences the conduct of research.

- Science is taught to almost all Canadian students up to grade 9 or 10. Beyond this level, science is typically taught as a series of specialized subjects, with biology, physics, and chemistry being the most common. The number of students taking such courses is much lower than at the earlier grade levels. Nevertheless, Canada has among the highest participation rates in science courses of any of the IEA countries.

- Science programs at the elementary level possess very little commonality across provinces. Commonality of content increases with grade level, to the point where the senior secondary courses are relatively similar. Variations exist among provinces at all levels in the degree to which the curriculum is prescribed and in the amount of detail given in the provincial curriculum guidelines.

- At the elementary level, science is almost always taught by regular classroom teachers, with no specific training in science. The degree of specialization increases with grade level, in terms of both teacher qualifications and teacher assignments. There are some regional variations in specialization, with Ontario showing greater extremes than other areas.

- Female teachers dominate at the elementary level, while secondary teachers are predominately male. Substantial regional variations exist, however, with the male/female distribution being most unbalanced in Ontario and least unbalanced in Eastern Canada.

International Comparisons

- At the grade 5 level, French Canada ranked fifth and English

Canada seventh in achievement out of 16 participating countries, with scores 4 and 7 percentage points respectively below that of Japan, the highest ranked country.

- At the grade 9 level, the ranking for English Canada improved to fourth and that for French Canada declined to ninth out of 18 countries. The percentage difference between English Canada and the highest ranked country (Hungary) increased to 10 points. A number of countries were clustered at about the same level as Canada, as only two points separated English and French Canada, despite the difference in ranking.

- Both Canadian groups ranked near the bottom among the 14 participating countries in all subjects at the senior secondary level.

- It is important to recall that wide differences exist in the definition of who is included in the population at the senior secondary level. In Canada, a high proportion of students stay in school and a high proportion take science. In contrast, the highest ranked countries, such as England and Hong Kong, are highly selective at the senior secondary level, with low proportions taking science.

Provincial Achievement Results

- About 10 percentage points separated the highest and lowest performing provinces at the grade 5 level. More generally, the western provinces tended to have the highest scores, followed by Ontario, and finally the eastern provinces.

- A similar pattern exists for the grade 9 scores, with the exception that Nova Scotia departs from the general pattern of lower scores for the eastern provinces. Indeed, Nova Scotia ranked with Alberta in having the highest achievement at this level.

- At the grade 12/13 level, the differences between provinces were somewhat greater than for the earlier grades. As much as 17 per cent separated the highest and lowest performing provinces. Regional differences were similar to those at the earlier levels, except that Ontario fared somewhat better relative to the western provinces at this level. The lowest performance was again found in the eastern provinces.

- Participation rates in senior secondary science courses varied widely from province to province. Correlations between participation rates and achievement were consistently negative. However, the statistics are not sufficiently stable to allow a firm conclusion to be drawn on the relationship between participation and achievement.

- Boys consistently outperformed girls at all levels and in all provinces.

- Opportunity to learn, as measured by teacher ratings, was not strongly related to achievement. Teachers were generally able to predict levels of student success more accurately than other variables related to opportunity to learn.

Attitudes

- In general, attitudes towards school and science were quite positive at all levels. Small regional differences were found, in a direction which tended to be opposite to the achievement differences. That is, students in Eastern Canada tended to have more positive attitudes than their counterparts in other parts of the country.

- Gender differences in attitude varied with the specific attitude scale. Girls showed consistently more positive attitudes towards school than boys. On the other hand, on scales more directly related to school science experiences and science as a career, the attitude differences tended to favour boys.

- Correlations between attitudes and achievement were generally small but positive.

Selected International Comparisons: Normative Interpretation

The basic underlying question in examining the international results is whether or not the performance of Canadian students is satisfactory. As discussed at the beginning of this report, several issues enter the picture in interpreting the results of any testing program. Taking a purely normative stance, concentrating on rankings, it might be argued that the results show Canada's performance to be reasonably satisfactory at the elementary and junior high school levels, but highly unsatisfactory at the senior high school level. In fact, given the diversity which exists within Canada, it is not at all surprising that average achievement levels are below those of the top-ranked countries.

It would be a mistake to place too much emphasis on overall rankings, especially at the senior secondary level. It is somewhat more useful to make specific comparisons, using countries with which Canada is frequently compared or those which have educational systems fairly close to those found in Canada. Some comment on Canada's performance relative to the United States is obviously warranted. Other useful comparisons might be made with Sweden and Australia. Finally, since Japan seems to be the country currently used as a model of strong economic performance, and which is reputed to have a highly intensive approach to education, it is difficult to avoid comparisons with Japan.

At the Population 1 level, the differences among the countries in question are relatively small.[22] Both Canadian populations are somewhat behind Japan (by 4 and 7 percentage points for French and English respectively). Both are a comparable amount ahead of the United States and Australia. Sweden is in about the same position as French Canada, slightly ahead of English Canada. In general terms, it can be argued that the countries which have performed best at this level are those with relatively centralized and intensive educational systems. Among the countries with decentralized systems, particularly the United States and Australia, Canada's performance is quite high.

In the general sense, participation is not an issue here, since all of the countries in question have close to universal education at this level. One area which needs to be considered, however, is the placement of students with learning disabilities of various sorts. In the United States, and increasingly in Canada, such students are largely integrated into regular classrooms, and hence would have been included in the defined populations. Whether this is also true for other countries is not known. Differences in such placement would be a variety of selectivity which could substantially influence scores at the lower end.

There are many other possible explanations for differences in performance at this level. An example would be the intense difficulties experienced in inner-city education in the United States and, on a different level (commonly associated with non-native speakers of English or French), in Canada. Given the relative independence of curriculum indices and achievement at this level, it is also possible that differential exposure to influences outside the school system could be significant. As suggested elsewhere in this report, time is often cited as a significant factor in achievement. Unfortunately data on time have not yet been incorporated into the international analyses. While countries such as Japan do have longer school years and days than Canada, it is by no means clear that this translates into increased exposure to science at the elementary level.

At the Population 2 level, the performance of both Canadian groups is close to that of several other countries, including Sweden and Australia. Canadian students achieve distinctly better than their American counterparts, but notably worse than Japanese students. One interesting point about Population 2 is that the overall international mean score is higher than for Population 1, suggesting that the test was somewhat easier in general. At the same time, the range of scores across countries was greater, with Japan along with a couple of other countries, including Australia, gaining somewhat relative to Population 1. The scores for Canadian students, along with their counterparts in the United States and Sweden, remained relatively unchanged. This suggests that greater growth in achievement has occurred for some countries relative to others. It is possible that time, intensity of programs, or other school related factors are at work here. It is important to note that by the end of grade 9, Canadian students have generally been exposed to three years of systematic study of science, beyond the more haphazard approach of the earlier years. Many such programs are, however, based on a process approach to science, rather than the facts and concepts approach which would match the test more closely. There is, in fact, evidence that students in at least two provinces (Newfoundland and British Columbia) perform better on test items dealing with scientific processes than on content-based items.[23] More information is obviously needed on this point.[24]

Obviously, the picture at the Population 3 level is much more complex than at the other two levels. Issues of participation and curriculum relevance have already been discussed. In terms of the comparisons being made here, the differences between Canada and Japan have widened, although Japan has been displaced from the top rank by countries which are obviously more selective at the senior secondary level. Sweden and Australia also outperform Canada at this level, while the performance of American students is slightly higher for chemistry and physics, and slightly lower for biology. As already indicated, however, the American results are entirely anomalous, because low achievement is combined with extremely low participation rates.

A Criterion-Based Interpretation

Unless we believe that the only relevant question in performance is whether Canadian students are as good as their counterparts in other countries on the specific limited set of objectives measured by the tests at hand, normative comparisons can give only a limited answer to the question of whether levels of achievement are satisfactory. Unfortunately, moving to a criterion-based approach would require an examination of performance on specific items and some judgement on test validity relative to important objectives of science teaching. Such an analysis is clearly beyond the scope of this report.[25] By considering certain features of the tests, however, it is possible to make some inferences about performance criteria.

As already pointed out, the tests were validated with respect to a list of topics on which an international panel of experts could agree as representing a common minimum core of scientific knowledge expected of students at the relevant levels. The tests were also constructed so as to be of average difficulty and to have broad power to discriminate high-performing from low-performing students. Under such conditions, experience suggests that the average score over a large sample of students would be in the range of 60 per cent, and that students in all countries should perform at close to the same levels.

By these criteria, scores for students in Canada, as well as for those in Australia and Sweden, may be considered satisfactory at the Population 1 and 2 levels. The same is true for English-Canadian, but not for French-Canadian, students on core knowledge at the senior secondary level. This is true despite the relatively low rankings, since the rankings at this level are obviously linked to selectivity. The performance of Japanese students, especially at the Population 2 level, would have to be considered outstanding. Students in the United States, on the other hand, can be said to be performing at unsatisfactory levels.

Issues of selectivity aside, it is difficult to make a case that performance is satisfactory when students achieve at less than the 50 per cent level on a test made up of items on the specific subject matter of courses they were just completing at the time of testing. Looked at in another way, the failure rate would have been unacceptably high had this test been used as a final examination in the courses concerned. The same argument would apply, of course, to most countries, with the exception of those with highly selective programs. This raises the important question of selectivity as an educational policy, a question which will be addressed in the final section of this report.

Observations on Provincial Results

The two broad interpretations of the results can be applied equally to the provincial comparisons. In fact, an argument can be made that a normative interpretation is more appropriate here than for the international data, if we assume that provinces would be expected to share common goals, and that students across Canada will ultimately have to face post-secondary and job market expectations established in a national setting. In reality, although there is no national coordinating body for elementary and secondary education in Canada, a good deal of commonality is achieved through such devices as the textbook market and exchanges of information at the policy level.

Under these conditions, it is not surprising to find that differences between provinces are somewhat smaller than those between countries. At the same time, differences ranging from 10 percentage points at the Population 1 level to 20 or more percentage points in the specialized subjects are larger than might be expected. For Population 1, only four provinces have scores at or above the 60 per cent criterion. The best overall performances are for Population 2 and the Population 3 core test, where only four groups are below this level. All groups are below the 60 per cent criterion in the specialized subjects. In fact, no province even comes close to the international mean in these subjects.

Two groups stand out as particularly disadvantaged. These are students in Newfoundland and Francophone students outside Quebec. Francophone Quebec students also fare particularly poorly at the senior secondary level. It is important to keep in mind, however, that students in Quebec were in grade 11 rather than grade 12. Newfoundland students were also the first group following the introduction of grade 12 in that province. With few exceptions, students in Eastern Canada are noticeably below those in Western Canada in performance. In general, the highest levels of achievement are found in Alberta and British Columbia.

Again, this analysis, in itself, cannot account adequately for the observed differences in achievement. One might hypothesize that the poor performance of Francophones outside Quebec is related to the relative isolation of such groups within English-speaking provinces. This point obviously needs detailed examination, as it has important national policy implications. As for the performance of students in Newfoundland, this has recently been the subject of an intensive investigation by a provincial task force.[26] One of the primary sources of difficulty was found to lie in low expectations. These, in turn, are manifested in curriculum, teaching practices, evaluation standards, and a variety of other aspects of education in Newfoundland. It is likely that some of the same factors are sources of poor performance in other provinces as well.

Educational Policy Implications[27]

The most obvious policy question raised by the study is "how can the achievement levels of Canadian students be improved?" The conclusion that scores are generally satisfactory at the elementary and junior secondary levels gives little reason for complacency, when one considers that other countries, with no obvious advantages over Canada in state of development, financial resources, or other characteristics important to attainment, perform at substantially higher levels. More disturbing is the fact that a generally satisfactory national score can mask substantial differences between provinces. Added to this, scores are clearly unsatisfactory at the level at which students are beginning to make important decisions about further education and careers.

One means of improving scores at the senior secondary level would be to use more stringent selection procedures, so that only the most talented have access to the specialized courses at this level. Aside from the ethical and political considerations of such a course of action in an educational system based on egalitarian principles, there are sound pragmatic reasons for rejecting this approach. The Science Council of Canada report, as well as other studies, has made a strong case that a high level of scientific literacy is required for all citizens.[28] This is incompatible with restrictive access, and particularly with ending formal instruction in science for most students at the grade 10 level. At a much broader level, it is clear that education beyond the secondary level will be required in the future for increasingly larger proportions of the population. Restricting access to the basic prerequisites for post-secondary education would obviously work against this goal.

At the same time, the Newfoundland Task Force found an important link between low achievement levels and attempts to improve retention rates in high school and participation rates in science courses. The link appears to lie in a lowering of expectations which accompanied a general broadening of the high school program. The basic policy underlying recent revisions of secondary education was to find programs suitable for students with a limited range of academic capability and interest, on the assumption that is was better to keep such students in school than to have them terminate their education on the grounds of academic failure. While few would dispute the goal of keeping more students in school, it is just as plausible to argue that this should be accomplished by measures which would improve their chances of success in basic academic subjects, rather than offering alternatives to these which have the effect of lowering expectations and reducing opportunities for further education.

An important characteristic of the Canadian federation has been the attempt to overcome regional disparities through various kinds of national initiatives. Such initiatives are evident in health, social services, and especially in attempts at economic equalization. Perhaps the overriding national educational policy issue in Canada is whether such initiatives should be extended to education. Without making a broad argument on this point, there are clear examples of areas in which federal initiatives have occurred in education, when aspects judged to be in the national interest are at stake (vocational education and training is perhaps the best such example). As it happens, there are many aspects of national scientific and economic policy which relate directly to science education, a point which has been recognized in such initiatives as the Canada Scholarships Program.[29] A case can be made that the time has come to reconsider the historic relationship between federal and provincial governments in education, especially in areas of strategic national importance.

Turning to more specific issues, the Science Council study and the Newfoundland Task Force have both made recommendations for strengthening science education in Canada. While differing in detail, the two sets of recommendations are closely parallel in principle. While it is beyond the scope of this report to present the detailed recommendations or to elaborate on their rationale, it is useful to comment briefly on some of the main areas in which improvement can be made. Here we concentrate on areas common to the two reports and on issues of broad concern, rather than on the specifics of programs within a single province.

A first major step is to ensure a more systematic treatment of science in elementary schools. This requires the allocation and use of time for science in all classrooms, improvements in curriculum and support materials, and more intensive scientific preparation for teachers at this level. More specifically, elementary science programs should focus on the natural environment and on the world of technology, rather than on the more formal concepts from the traditional scientific disciplines.

Both reports referred to the inadequate view of science presented in existing textbooks. Several problems are evident. Texts rarely present science from a Canadian perspective or present Canadian examples. A substantial gap exists between the high-minded statements of curriculum guidelines and the detailed material found in textbooks. The interaction of science and society, and the relationship between science and technology, are poorly treated. All of this points to a need for a major thrust in curriculum development. Given the cost of producing high-quality curriculum material, the small markets for such materials in most provinces, and the need for an increase in Canadian rather then American materials, a case can be made for national curriculum initiatives, rather than repeated efforts in each province.

The education of science teachers is a crucial element in any plan for improvement. The Newfoundland Task Force noted with dismay that more than two decades of intensive teacher upgrading effort in that province has produced a plentiful supply of highly qualified teachers, but a serious imbalance in areas of specialization, with almost all teachers concentrated in English and social studies and very few in science, especially physics and chemistry. The Science Council study concentrated on the need for continuing teacher education and for more adequate preparation of elementary teachers.

A further area of concern is that of monitoring and assessment. There is good evidence, especially from British Columbia, that a systematic approach to student assessment yields substantial payoff in achievement. For this to occur, however, it is necessary to go beyond the simple reporting of test results. Detailed analysis of areas of strength and weakness is required, and efforts made to overcome weaknesses. It is particularly important to go beyond normative interpretations. The Newfoundland study found, for example, that public examinations in that province have been marked and scaled so as to yield similar high pass rates from year to year, thus creating an illusion of high achievement, while losing the ability to determine whether actual achievement levels are changing over time.

Again, one can make a case for a national approach to student assessment. The SISS project has shown that it is feasible to carry out such testing. What is required, however, is a test development initiative, which would produce more adequate instruments than those used in the international study. Several provinces have already embarked on major test development projects. Given the commonality of science goals and content, and the developmental costs involved, a strong argument can be made for coordinating these efforts on a national scale. Devices such as common item banks, national validation and interpretation panels, anchor items for national comparison, coordinated schedules of test administration, and so on could yield substantial improvement in test validity, efficiency, and comparability. As it happens, a project currently being initiated by the Council of Ministers of Education is designed to accomplish some of these goals. The Council has rejected the use of an existing standardized test, and is proceeding with the development of new national instruments. Such efforts are to be encouraged.

A final point made by both the Science Council and the Newfoundland studies is the need to identify and encourage exceptional talent in science, and the associated problem of encouraging higher levels of participation and achievement among young women. This leads us back to the question of whether an egalitarian system, which encourages high levels of participation, diminishes the performance of the most talented students. The existence of a negative correlation between participation and achievement across countries tells us little about the performance of the best students. Unfortunately, within Canada, the provincial samples were much too small to allow any conclusions to be drawn on this point. International analyses now in the preliminary stages suggest that bright students do very well on the tests in all countries. However, this result may be an artifact of the SISS instruments, which were not designed for fine discriminations at the top end. That is, most of the very best students would be expected to get nearly perfect scores on these tests. Further study is required on this issue.

Both the Science Council and the Newfoundland Task Force studies addressed this problem. Both recommended the development of special programs and experiments with special schools for talented students. The Newfoundland study recommended a program of provincial scholarships comparable to the Canada Scholarships Program. The use of role models for women, and the avoidance of gender bias in course selection and in career guidance were also recommended. Finally, following the example of the Canada Scholarships Program, the Newfoundland Task Force argued that all new programs and incentive systems be designed so that half the rewards would go to women.

Economic and Scientific Policy Implications

It has become almost an article of faith in our society to equate economic strength with scientific and technological strength. Canada's low level of expenditure on research and development has frequently been cited as one of the detriments to improved economic performance and increased competitiveness. A variety of government initiatives, ranging from higher levels of funding for the research granting councils to incentives for increased private sector investment in research and development, have been used in an attempt to improve this situation. Long-standing programs such as fellowships for graduate studies, along with more recent initiatives such as the Canada Scholarships Program, have been based on the further assumption that education is a key element in improving the nation's scientific and technological capability.

One thing is clear. The developed world is now in the midst of a technologically-driven economic revolution as profound as any which has ever occurred. It is no exaggeration to compare the changes now underway to those of the industrial revolution of the nineteenth century or the communications revolution of the early part of this century. The current revolution is characterized by several features, including a decline in the importance of resource extraction and of manufacturing, an increase in the importance of service industries, and increased internationalization of commerce. Aside from the host of new products and services engendered by such developments as microchip technology and biotechnology, one of the strongest areas of growth is in areas concerned with knowledge processing itself.

Part of the shift can be traced to continuing increases in the efficiency with which most basic goods can be produced. Basic food production, for example, can be accomplished with fewer and fewer people. In contrast, food processing and service has become a major area of growth. Manufacturing has suffered essentially the same fate, as automation of the means of production has led to a reduced demand for labour, and as increased mobility of capital and production facilities has led to the relocation of much manufacturing to developing countries, where whatever demand continues to exist for labour can be met at much lower cost than in the developed world.

The growth in service occupations presents a fundamental dilemma for the analysis of the demand in scientific and technological fields. Current trends and projections[30] show that demand is increasing in a range of low level service occupations, characterized most directly by the fast-food industry. The other main area of growth is at the opposite end of the occupational spectrum, in the professional and management categories. Traditional middle-range occupations, such as the skilled trades, have been on a long-term decline. At the same time, many of these occupations (auto mechanics and typists being primary examples) have been profoundly affected by the technological revolution.

To the extent that educational policies are directed to the production of workers, it may be argued at one level that current trends would allow us to settle for the modest requirements of areas such as the food service or hospitality industries. On the other hand, this does little to enhance the nation's ability to compete in a world increasingly characterized by technological innovation and knowledge production and use. The leading edge of most areas of economic activity, from forestry to computer software development, from fishing to telecommunications, involve the application of science and technology to improve productivity, develop new products and services, or solve problems. It is now widely recognized that knowledge is becoming the most prized commodity in the emerging global economy. Knowledge resides in people, and people are much less mobile than other elements in economic activity. Those societies which are most successful in educating their people to a level which is attractive in a global economy are much more likely to be successful than those which continue to rely on natural resources, manufacturing, or services built upon unsophisticated labour.

Beyond the question of a sophisticated workforce, there is, of course, the issue of the broader function of education in producing citizens who are able to comprehend the complex issues which emerge from living in a technological society. On a global scale the multiple threats of nuclear war, environmental degradation, maldistribution of wealth, overpopulation, and so on are ever present. Each of these has strong scientific and technological overtones. It may be argued that for every problem solved by science and technology, new ones are created which demand ever more of the scientific and technological knowledge. Increased sophistication on the part of the public is obviously required to understand the issues of society and to respond socially and politically.

All of this has taken us well beyond the immediate problem of science achievement. Many important questions are raised which require further research and deliberation. The link between science education and economic development needs to be much more fully developed before we can move from broad statements or articles of faith to prescriptions for improvement. Studies such as SISS tell us something about the status quo. Educational policy studies, such as the Science Council investigation or the Newfoundland Task Force, examine directly the problems of curriculum, teaching, and other elements of how children are introduced to scientific knowledge. Economic studies can perhaps point to areas in which Canada's economic performance is not at an optimal level, and to relationships between these problems and such factors as expenditures on research and development or worker training. Studies of a rather different sort are required to establish more closely how education is related to national economic or social goals. Examples of areas requiring investigation are the relationships between science attainment and occupational or economic success, the factors which influence student decisions to enter or not enter scientific or technological careers, or the degree to which specific programs, such as scholarships or research funding, have met their objectives. Also required are more precise projections of requirements for scientific and technological personnel. All of these require that there be closer linkages between those engaged in educational research and policy studies and those concerned with matters of economic development.

Notes

1 L. C. Coomber and J. R. Keeves, *Science Education in Nineteen Countries: International Studies in Evaluation I*, Stockholm, Almkvist and Wiksell, 1973.

2 International Association for the Evaluation of Education Achievement, *Science Achievement in Seventeen Countries: A Preliminary Report*, Oxford, Pergamon Press, 1988.

3 Professors Heidi Kass, at the University of Alberta, Professor Michael Connelly, at OISE, and Professor Robert Crocker, at Memorial University.

4 The Francophone study was directed by Professor Gilles Dussault of l'Université du Québec à Hull.

5 International reports refer to the two studies as Canada, English (CAE), and Canada, French (CAF). Whenever possible, the comparisons made in this report are based on data from both studies.

6 Science Council of Canada, *Science for Every Student: Educating Canadians for Tomorrow's World*, Ottawa, Supply and Services Canada, 1984.

7 F. M. Connelly, R. K. Crocker, and H. Kass, *Science Education in Canada: Volume I, Policies, Practices, and Perceptions*, Toronto, OISE Press, 1985.

8 It is important to stress that this discussion does not do justice to the issue of error analysis. In fact, the complexity of the sampling designs and instruments makes it virtually impossible to calculate errors with any degree of precision. The figures given here are almost certainly underestimates of total error.

9 The *sampling fraction* is defined as the ratio of sample size to its corresponding population size. It is important to note that the large populations sample size rather than sampling fraction is the key determinant of sampling error.

10 This work is summarized in the following research reports: *Mathematics and Science Curricula in Canada: Comparative Analyses* (Report No. 4), *First Year Mathematics and Science*

Programs in Selected Canadian Universities (Report No. 5), St. John's, Government of Newfoundland and Labrador, 1989.

11 Detailed results of this program appear in *Achievement in Mathematics and Science*, Background Report No. 1 of the Newfoundland Task Force on Mathematics/Science Education, St. John's, Government of Newfoundland and Labrador, 1989.

12 These figures are so low as to arouse suspicion as to their accuracy. Nevertheless, colleagues in the United States have given assurances that this does represent an accurate picture.

13 The correlation coefficient is a measure of the degree of association between two variables. The coefficient has a value of 1.0 when the order of numbers for both variables is exactly the same (the highest number for the first variable corresponds to the highest for the second, and so on throughout the range). Its value is -1.0 when the order of numbers is reversed (the highest value of one variable corresponds to the lowest of the other, etc.).

14 These comments apply to English Canada, as participation rates for French Canada were not available.

15 Coomber and Keeves, *Science Education in Nineteen Countries*, Stockholm, Almkvist and Wiksell, 1974.

16 The standard score difference is computed by subtracting the mean score for boys from the mean score for girls and dividing the result by the mean standard deviation over all scores.

17 Small changes have been made to the usual provincial abbreviations to accommodate the space available in the charts. The following additional abbreviations have been used in the charts:
INT International Mean Score (unweighted)
QE Quebec English
QF Quebec French
FOQ French outside Quebec.

18 That is, it is unlikely that the pattern of differences between provinces can be attributed to sampling errors. Whether or not differences are statistically significant depends on the size of the sampling and measurement errors, and the level of confidence we wish to attach to a particular conclusion. For example, depending on population level and subject, differences in mean scores

ranging from 3 to 5 per cent are required before we can say with 95 per cent confidence that these are not due to sampling errors.

19 These results appear in various chapters in F. M. Connelly, R. K. Crocker, and H. Kass, *Science Education in Canada: Volume II, Achievement and its Correlates*, Toronto, OISE Press, 1989.

20 Attitude data are not available for Quebec (English) or for the French study.

21 For samples of 250, which are typical of the smallest groups being compared across provinces, a mean difference of about .03 on the 5 point scale is required for significance at the .99 confidence level.

22 Because of the large sample sizes, differences between countries are generally statistically significant. However, the discussion here is more concerned with practical than with statistical significance.

23 Robert K. Crocker, *Achievement in Mathematics and Science*, Background Report No. 1, Mathematics/Science Task Force, St. John's, Government of Newfoundland and Labrador, 1989.

24 Several SISS countries did administer a process test. No international results for this test have been published to date.

25 Substantial detail on item performance is given in Connelly, Crocker, and Kass, op. cit., 1989.

26 References have already been made to background reports of this Task Force. A full account is found in *Towards an Achieving Society*, (Final Report of the Task Force on Mathematics/Science Education), St. John's, Newfoundland, Queen's Printer, 1989.

27 This section, and those which follow, draw heavily on the 1984 Science Council of Canada report and on the report of the Newfoundland Mathematics/Science Task Force.

28 A detailed discussion of the meaning of the concept of scientific literacy is beyond the scope of this report. Interested readers are referred to the Science Council of Canada report and the report of the Newfoundland Mathematics/Science Task Force for elaboration of this concept.

29 The Canada Scholarships Program is a new federal program which offers four-year undergraduate scholarships to promising high school graduates enroled in university programs in the basic sciences.

30 Canadian Occupational Projection System, *Job Futures: An Occupational Outlook to 1995*, Ottawa, Employment and Immigration Canada, 1988.

List of Tables and Charts

Tables

Charts

References

Connelly, F. M., et al., *Science Education in Canada: Volume I, Policies, Practices, and Perceptions*, Toronto, OISE Press, 1985.

______, *Science Education in Canada: Volume II, Achievement and its Correlates*, Toronto, OISE Press, 1989.

Coomber, L. C. and J. R. Keeves, *Science Education in Nineteen Countries: International Studies in Evaluation I*, Stockholm, Almkvist and Wiksell, 1973.

International Association for the Evaluation of Educational Achievement, *Science Education in Seventeen Countries: A Preliminary Report*, Oxford, Pergamon Press, 1988.

Newfoundland Task Force on Mathematics/Science Education, Research Reports:

No. 1 *Achievement in Mathematics and Science*

No. 2 *Science Achievement in Canada: Interprovincial Comparisons*

No. 4 *Mathematics and Science Curricula in Canada: Comparative Analyses*

No. 5 *First Year Mathematics and Science Programs in Selected Canadian Universities*

Final Report:

Towards an Achieving Society, St. John's, Queen's Printer, 1989.

Science Council of Canada, *Science for Every Student: Educating Canadians for Tomorrow's World*, Ottawa, Supply and Services Canada, 1984.