

by John Bishop

HIGH SCHOOL DIPLOMA EXAMI- NATIONS: DO STUDENTS LEARN MORE? WHY?

Certaines provinces ont remis à l'honneur les examens provinciaux en vue de l'octroi d'un diplôme. Les autres devraient-elles leur emboîter le pas? Un examen des résultats d'études effectuées au Canada et ailleurs semble suggérer qu'elles le devraient. D'après certains indices, en effet, les étudiants habitant dans les pays ou les provinces qui se servent d'examens de ce genre ont tendance à avoir des meilleurs résultats que ceux qui habitent ailleurs.

Many have expressed dissatisfaction with the academic achievement of Canadian youth. The Economic Council of Canada, for example, observed in 1992 that: "International test results for science and mathematics show that Canadian students in Grade 4 perform at least at the international average; by Grade 8, they are outperformed by students in many other countries; and by the end of secondary school, Canadian results are poor."¹

Comparisons over time are apparently unfavourable as well. Researchers working on the Economic Council report compared the results of Canadian Test of Basic Skills norming studies in 1966, 1973, 1980 and 1991 and concluded that: "On the whole, they indicate a deterio-

ration between 1966 and 1973, a minor improvement between 1973 and 1980 and then another decline (to about the 1973 level) between 1980 and 1991."²

Political leaders have responded to the concerns in a variety of ways. One of the most prominent responses has been to reverse decisions made in the late 1960s and early 1970s to eliminate provincial diploma examinations. British Columbia reestablished its Provincial Examination system in 1983 and Alberta followed one year later. Manitoba started administering rotating provincial examinations in 1991. They examined mathematics in 1991 and 1992, biology and physics in 1993, chemistry in 1994 and social studies in 1995. In 1996 Manitoba stopped rotating its exams and began administering provincial examinations in mathematics and language arts every year.

What effects do provincial diploma examinations have on school policies, teaching and student achievement? Time series data on the student achievement by province is not available, so two types of cross-section evidence will be examined: comparisons of countries with and without curriculum-based external exit examinations and comparisons of Canadian provinces with and without such exams.

International cross section evidence

International data on achievement in mathematics and science of 13 year old students is available from the 1991 International Assessment of Educational Progress (IAEP) and the 1994-95 Third International Math and Science Study (TIMSS). Comparative education studies, government documents and education ministry officials, embassy personnel and Cornell graduate students from the country were interviewed to determine which of the nations have curriculum-based externally set exit examinations (CBEEES) similar to the diploma examinations of Alberta, BC and Quebec.³

The median science and mathematics test scores of 13 year olds were then regressed on per capita gross domestic product for 1987 and 1990 deflated by a purchasing power parity price index, a dummy for East Asian nations and a dummy for a CBEEES. Students in East Asian nations did better than students in Europe and North America, especially in mathematics. National wealth was also associated with higher achievement levels.⁴

The analysis of IAEP data found that CBEEES had large (nearly 2 grade level equivalents) statistically significant effects on mathematics but small statistically insignificant effects on science achievement. The analysis of TIMSS data found that CBEEES raised performance by about 1.1 Canadian grade level equivalents in science and about two-thirds of a Canadian grade level equivalent in mathematics.⁵

Comparing Canadian provinces

When the Educational Testing Service canvassed countries about participating in the 1991 IAEP, Canada

decided to collect sufficient data to allow reliable comparisons between provinces and between the Anglophone and Francophone school systems of the five provinces with dual systems.⁶ At the time, Alberta, BC, Newfoundland, Quebec and Francophone New Brunswick had curriculum-based provincial examinations in English, French, mathematics, biology, chemistry and physics during the senior year of high school (U.S. GAO 1993).⁷ These exams accounted for 50 percent of that year's final grade in Alberta, Newfoundland and Quebec and 40 percent in BC. These provincial exams are medium stakes, not high stakes tests. They influence grades and appear on transcripts but passing the examination is not essential for graduation. Employers appear uninterested in exam scores. Job application forms do not request that applicants report exam scores or grades.

The other provinces did not have curriculum-based provincial external exit examinations in 1990-91. Ontario eliminated them in 1967, Manitoba in 1970 and Nova Scotia in 1972. Nova Scotia substituted multiple-choice norm-referenced achievement tests in reading, language usage, proof-reading, mathematics, science and social studies that do not influence student grades. Anglophone New Brunswick had provincial exams in language arts and mathematics but exam grades were not reported on transcripts or counted in final course grades. The absence of any stakes attached to performance resulted in many students blowing off the test.

The effects of curriculum-based provincial exit exams taken by 12th graders on achievement and the behaviour of Canadian 13 year olds, their parents, teachers and school administrators were examined by estimating models predicting these behaviours using schools as observations. The data set comprises 1338 Canadian schools. The model contained 11 variables: logarithm of the mean number of books in the home, the mean number of siblings, the proportion of the school's students whose home language was different from the language of instruction, logarithm of the number of students per grade in the school and zero-one indicator variables for secular independent schools and for non-secular independent schools, schools with primary grades, schools that include all grades in one building, French-speaking schools and location in a

province with provincial exit exams. Twenty percent of the schools in the study were run by locally elected Catholic (or Protestant) school boards in the provinces of Alberta, Newfoundland, Ontario or Saskatchewan. A separate dummy variable was created for these schools.

The metric of the primary dependent variable, achievement, is percent correct with adjustments for guessing. The mean of the mathematics test was .47 and its standard deviation across individuals was .248. For the science test the mean was .541 and the individual standard deviation was .19.

Table 1 presents regression results predicting student achievement at age 13 and various indicators of student, parent, teacher and school administrator attitudes and behaviours. The means and standard deviations across schools of each dependent variable are presented in columns 1 and 2. The estimated impact of

Table 1
Effects of Provincial Diploma Examinations in Canada

Achievement	Dep Var Mean	School StDev	Diploma Exam	Relig. SchBd	French Speaking	AdjR2
Mathematics	.470	.135	.051***	-.048***	.074***	.329
Science	.541	.096	.026***	-.036***	.021***	.323
Discipline Problems	.765	.720	-.017	-.13**	.19***	.080
Absenteeism Problems	.822	.766	.140***	.00	-.16***	.131
School Administrator Behaviour						
Math Specialist Teachers	.45	.50	.18***	-.20***	.08**	.280
Science Specialist Teachers	.46	.50	.15***	-.10***	-.03	.279
Math Tchrs majored in Math	.64	.39	.19***	-.12***	-.06*	.127
Sci. Tchrs majored in Sci.	.69	.38	.19***	-.17***	-.21***	.199
Math Class Hours/week	3.98	.88	.33***	-.06	.31***	.124
Science Class Hours/week	2.93	.79	.16***	-.37***	-.06	.132
Good Science Labs	1.95	.95	.28***	-.10	.04	.274
Computers per student	.051	.043	.001	-.009***	-.006*	.195
Teacher Behaviour						
Total Homework (hrs/wk)	4.41	1.62	.66***	.62***	-.48***	.149
Emph. Whole Num. Computation	1.68	.49	-.09***	-.01	.10**	.035
Students do Sci. Experiments	1.52	.63	.28***	.14***	.36***	.145
Teachers do Sci. Experiments	2.42	.47	.15***	.10***	.23***	.111
Home Behaviour & Attitudes						
TV Hours/week	14.7	2.85	-.68***	.63***	1.70***	.255
Read for Fun Index	1.85	.28	.05***	.03	.08***	.115
Watch NOVA, Nature	.97	.38	.06**	.07**	.21***	.091
Parents Talk about Math Class	.62	.17	.04***	.04***	.02	.046
Parents Talk about Science Class	.47	.17	.06***	.01	-.01	.056
P. want me to do well in Math	3.54	.22	.05***	.09***	-.01	.104
Parents are interested in Science	2.18	.34	.06***	.11***	.12***	.071
Science Useful in Everyday Life	2.46	.31	.06**	.14***	.18***	.095

Source: Analysis of IAEP data on 1338 Canadian schools. Control variables not shown included: logarithm of the mean number of books in the home, the mean number of siblings, the proportion of the school's students whose home language was different from the language of instruction, logarithm of the number of students per grade in the school and dummies for independent secular and non-secular schools, schools with primary grades and schools that include K through 11th grade in one building.

*** p < .01 on a two tail test

** p < .05 on a two tail test

* p < .10 on a two tail test

provincial exams systems is given in column 3. The estimated impacts of control by a religious school board is given in column 4. Estimates of the effect of attending a French-speaking school are in column 5. The R2 corrected for degrees of freedom is reported in column 6. Three stars (***) signifies statistical significance at the 1 percent level on a two tail test (*i.e.*, there is less than a one percent chance that, controlling for the other variables in the model, there is no true relationship between provincial exams and the dependent variable.) Two stars (**) signifies significance at the 5 percent level. Let us begin by examining how French-speaking schools and schools run by elected religious school boards differ from other schools.

How do schools run by elected religious school boards differ?

Even though students and parents had more positive attitudes about mathematics and science, homework assignments were greater and discipline was better, schools run by religious school boards had significantly lower math and science achievement levels than schools run by secular school boards.⁸ Why? Could it be that math and science teaching gets lower priority at schools run by elected religious school boards? These schools are less likely to hire math and science teachers who majored in their subject and who specialize in teaching it. They also schedule fewer hours of instruction in science.

How are French-speaking schools different?

Students at French-speaking schools did considerably better than students at English-speaking schools, particularly in mathematics. Principals report fewer absenteeism problems but more discipline problems. Parents and students have more positive attitudes toward math and science, but they do less homework and watch a lot more television. Experiments are more common in science class. Administrators appear to give priority to math teaching. More time is devoted to math instruction but not to science instruction. Specialist math teachers are more common, specialist science teachers are less common.

Effect of diploma examinations on behaviour of students, teachers and administrators

Provincial diploma exams had significant effects on achievement: 20.6 percent of a Canadian standard deviation (about three-quarters of a grade level equivalent) in mathematics and 13.7 percent of a standard deviation (about a half a grade level equivalent) in science. Diploma exams also apparently affected the behaviour of parents, teachers and school administrators. Schools in diploma exam provinces scheduled more hours of math and science instruction, assigned more homework, had better science labs, were significantly more likely to use specialist teachers for math and science and more likely to hire math and science teachers who had studied the subject in college. Eighth-grade teach-

ers in diploma exam provinces gave tests and quizzes more frequently. Hours in the school year, class size and teacher prep time were not significantly affected by diploma exams.

One possible skeptical response to these findings is to point out that the correlation between diploma exams and other outcomes may not be causal. Maybe the people of Alberta, BC, Newfoundland, Quebec and Francophone New Brunswick — the provinces with exam systems — place higher priority on education than the rest of the nation. Maybe this trait also results in greater political support for examination systems. If so, we would expect that schools in the diploma exam provinces should be better than schools in other provinces along other dimensions such as discipline and absenteeism, not just by academic criteria. In a 1996 article in the *International Journal of Education Research*,⁹ I predict, to the contrary, that exam systems induce students and schools to *redirect* resources and attention toward learning/teaching exam subjects and away from the achievement of other goals such as low absenteeism, good discipline and teaching computer skills. These competing hypotheses are evaluated in the 3rd, 4th and 12th rows of Table 1. Contrary to the “provincial taste for education” hypothesis, principals in diploma exam provinces had not purchased additional computers, did not report significantly fewer discipline problems and were significantly more likely to report absenteeism problems.

Do CBEEES distort teaching?

Opponents of externally set curriculum-based examinations predict that they will cause students to avoid learning activities that do not enhance exam scores. This hypothesis was tested by determining whether 13 year old students were less likely to read for pleasure and watch science programs like NOVA and Nature when they lived in provinces with a diploma exam. Neither of these hypotheses was supported. Indeed students in exam provinces spent significantly more time reading for pleasure, more time watching science programs on TV, while watching significantly less TV overall (see row 17-19). Parents in these provinces were more likely to talk to their children about their math and science classes and their children were more likely to report that their parents “are interested in science” or “want me to do well in math” (see rows 20-23).

Do CBEEEs skew teaching in undesirable ways? Madeus has pointed out that “preparation for high stakes tests often emphasizes rote memorization and cramming of students and drill and practice teaching methods” and that “some kinds of teaching to the test permit students to do well in examinations without recourse to higher levels of cognitive activity.”¹⁰ Contrary to this hypothesis, however, students did more (not fewer) experiments in science class, the teacher did more experiments in front of the class and emphasis on computation using whole numbers — a skill that should be learned by the end of 5th grade — declined signifi-

cantly (see rows 14-16). Apparently, teachers subject to the subtle pressure of a provincial exam four years in the future adopt strategies that are conventionally viewed as "best practice," not strategies designed to maximize scores on multiple choice tests. Students responded to the improved teaching by becoming more likely to report that science was "useful in everyday life" (see row 24).

The argument cited by Madeus can also be challenged on *a priori* grounds. It assumes that examinations developed by the committees of teachers working for provincial Ministries of Education are worse than the tests developed by individual teachers. In fact, the tests that teachers develop for themselves are generally of very low quality. Fleming and Chambers study of tests developed by high school teachers found that four-fifths of the items on teachers' tests tapped the lowest of [Bloom's] taxonomic categories, knowledge (of terms, facts or principles).¹¹ Rowher and Thomas found that only 18 percent of history test items developed by junior high teachers and 14 percent of items developed by senior high teachers required the integration of ideas.¹² University instructors, by contrast, required such integration in 99 percent of their test items. Secondary school teachers test low-level competencies because that is what they teach. This should be no surprise when one realizes that many are teaching subjects that they neither majored in nor minored in at university. Provincial diploma examinations, by contrast, get a great deal of high-level professional scrutiny. Item writers are generally drawn from the ranks of the most outstanding teachers and all items are pre-tested and checked for bias. Well-designed external examinations will induce improvements in instructional practice. Sherman Tinkelman, New York State's Assistant Commissioner for Examinations and Scholarships, describes one such instance:

*For years our foreign language specialists went up and down the state beating the drums for curriculum reform in modern language teaching, for change in emphasis from formal grammar to conversation skills and reading skills. There was not very great impact until we introduced, after notice and with numerous sample exercises, oral comprehension and reading comprehension into our Regents examinations. Promptly thereafter, most schools adopted the new curricular objectives.*¹³

A further benefit of CBEEES is the professional development that teachers receive when they are brought to centralized locations to grade the extended answer portions of the examinations. In May 1996 I interviewed a number of Alberta teachers union leaders about their experience serving on grading committees. Even though they and their union have opposed the examination system for years, they all said that having to discuss and agree with their colleagues about what constituted an excellent, good, adequate, poor and failing response to essay questions had been "a wonderful professional development activity."¹⁴

Conclusions

This review of the evidence suggests that the claim that reintroducing provincial diploma examinations will increase student achievement is probably correct. Students from countries with medium and high-stakes systems outperform students from other countries at a comparable level of economic development. Not only did 13 year olds from Canadian provinces with such systems know more science and mathematics than students in other provinces, they watched less TV and talked with their parents more about school work. Schools in provinces with external exams were more likely to:

- employ specialist teachers of math and science
- hire math and science teachers who had studied the subject in college
- have high quality science laboratories
- schedule more hours of math and science instruction
- assign more homework in math, in science and in other subjects
- have students do and watch experiments in science class.

Diploma exams have clearly not lowered the quality of instruction, they appear to have enhanced it.

Endnotes

1. Economic Council of Canada, *A Lot to Learn: Summary* (Ottawa: Canada Communication Group, 1993).

2. Keith Newton, Patrice de Broucker, Gilles McDougall, Kathryn McMullen, Thomas Schweitzer and Tom Siedule, *Education and Training in Canada* (Ottawa: Canada Communication Group, 1992).

3. John Bishop, "Do Curriculum-Based External Exit Exam Systems Enhance Student Achievement," CPRE Research Report RR-40, Consortium for Policy Research in Education, Vol. 1, no. 32, University of Pennsylvania (1998). Twenty-two national school systems were classified as having a CBEEES for both science and mathematics in all parts of the country: Austria, Bulgaria, Columbia, Czech Republic, Denmark, England, Hong Kong, Hungary, Ireland, Iran, Israel, Japan, Korea, Lithuania, the Netherlands, New Zealand, Russia, Scotland, Singapore, Slovak Republic, Slovenia and Thailand. Three countries — France, Iceland and Romania — had a CBEEES in mathematics but not in science. Five countries — Australia, Canada, Germany, Switzerland and the United States — had CBEEES in some provinces but not in others. The countries classified as not having a CBEEES in either subject were Belgium (both Flemish and French speaking systems), Cyprus, Greece, Philippines, Portugal, Spain and Sweden. The university entrance examinations of Greece, Portugal Spain, Cyprus and the ACT and SAT in the US were not considered to be a CBEEES. University entrance exams should have smaller effects on student and teacher behaviour because those headed into the job market do not take them and teachers can avoid responsibility for their students' exam results by arguing that not everyone is college material or that examiners have set an unreasonably high standard to limit enrollment in higher education.

4. Bishop, "Do Curriculum-Based External Exit Exam Systems Enhance Student Achievement."

5. Bishop, "Do Curriculum-Based External Exit Exam Systems Enhance Student Achievement."

6. All French speaking schools in New Brunswick, Saskatchewan and Manitoba were invited to participate and stratified random samples of 105 to 128 secondary schools were selected from the French speaking school systems of Ontario and Quebec and the English speaking school systems in all provinces with the exception of Prince Edward Island.

7. US General Accounting Office, *Educational Testing: The Canadian Experience with Standards, Examinations and Assessments* (written by Kathleen D. White) GAO/PEMD-93-11, Washington, DC, Vol. 1, no. 74 (April 1993).

8. School characteristics not shown in Table 1 also had substantial effects. Holding family background constant, students at independent schools (whether run by a secular or religious organizations) did better in mathematics than students at schools run by elected secular school boards. In science, however, independent schools had no advantage. Schools that combined elementary and middle grades did worse than schools that had no elementary grades in the building or that included all grades under one roof. When the grade structure of schools was controlled, size (the number of students per grade level) had no significant relationship with student achievement (Bishop, "Do Curriculum-Based External Exit Exam Systems Enhance Student Achievement").

9. John Bishop, "The Impact of Curriculum-Based External Examinations on School Priorities and Student Learning," *International Journal of Education Research*, Vol. 21, no. 2 (1996), p. 6.

10. George Madeus, "The Effects of Important Tests on Students: Implications for a National Examination or System of Examinations," *American Educational Research Association Invitational Conference on Accountability as a State Reform Instrument*, Washington, DC, June 1991, 1-19.

11. M. Fleming and B. Chambers, *Teacher-made Tests: Windows on the Classroom* (San Francisco: Jossey Bass, 1983).

12. W. D. Rohwer and J.W. Thomas, "Domain Specific Knowledge, Cognitive Strategies, and Impediments to Educational Reform," in M. Pressley (ed.), *Cognitive Strategy Research* (New York: Springer-Verlag, 1987).

13. Sherman N. Tinkelman, "Regents Examinations in New York State after 100 Years" (Albany, NY: The University of the State of New York, The State Education Department, 1996), pp. 1-15.

14. Bob, Interview conducted in Calgary Alberta in May 1996.

John Bishop is Chair, Human Resource Studies Department, Cornell University, Ithaca, New York.

by Helen Raham

BUILDING SCHOOL SUCCESS THROUGH ACCOUNTABILITY

Les Canadiens s'attendent à ce que leur système d'éducation soit soumis à des examens rigoureux afin de déterminer si les milliards de dollars qu'ils y investissent produisent les résultats les plus probants du point de vue de la qualité de l'enseignement. Or, les méthodes actuelles d'évaluation du système n'aident guère à démontrer dans quelle mesure les étudiants ont atteint les objectifs envisagés ou quel est l'impact des dépenses sur les programmes. Il faut recentrer le tir pour axer le système sur les résultats. L'auteure se penche sur certaines nouvelles stratégies.

Measuring school and system performance presents a major challenge and opportunity for governments. In May 1996, the Council of Ministers of Education, Canada (CMEC) held a national consultation on Accountability. There was little consensus on key indicators for which the system must be held accountable, and many of the education providers present rejected the measurement of results as "narrowing the purposes of education."

Canada is not alone. *Education Week's* comprehensive 1997 report card on the condition of public education in the US noted a troubling lack of useful hard data on performance: "Public education is a vast enterprise... Its success is clearly linked to the welfare of the nation and the future of our children. Yet we do not know in any but the crudest way, how well our education system is performing."