12-14-1994

The Impact of Curriculum-Based Examinations on Learning in Canadian Secondary Schools

John H. Bishop
Cornell University

The Impact of Curriculum-Based Examinations on Learning in Canadian Secondary Schools

Abstract

Externally set curriculum-based examinations at the end of high school apparently have pervasive backwash effects on middle school students, their parents, teachers and school administrators. Holding the social class background of students constant, students from Canadian provinces with examination systems were substantially (23 percent of a standard deviation) better prepared in mathematics and 18 percent of a standard deviation better prepared in science than students from provinces lacking such exams. The effect of an exam system on mathematics achievement of 13 year olds is larger in a standard deviation metric than the decline in math SAT scores between 1969 and 1980 that has been such a focus of public concern. Other natural experiments yield similar findings. When adjustments are made for ethnicity, gender and social class of SAT test takers, New York State ranks higher on the SAT than any of the other 38 states where the test is taken by large numbers of students. The mathematics and science achievement of Swedish high school seniors declined in the years following the elimination of high/medium stakes curriculum-based exams. The analysis also found that examination systems had pervasive effects on school administrators, teachers and parents. In the provinces with external exams, schools were more likely to: -- employ specialist teachers of mathematics and science -- employ teachers who had studied the subject in college, -- have high quality science laboratories -- schedule extra hours of math and science instruction -- assign more homework in math, in science and in other subjects -- have students do or watch experiments in science class and -- schedule frequent tests in math and science class. At home students watch less TV, spend more time reading for fun, and are more likely to report their parents want them to do well in math and science. In addition, parents are more likely to talk to their child about what they are learning at school.

Keywords

examination, school, student, learning, achievement, college, academic, New York, SAT, United States, curriculum, test, Canadian, math, science, grade, exam

Disciplines

Curriculum and Instruction

Comments

This paper was presented at the NSF/Review of Economics and Statistics conference on School Quality and Educational Outcomes, at the Kennedy School of Government, Harvard University on December 14-15, 1994. It's preparation was possible because of support from the Pew Charitable Trust, The Center for Advanced Human Resource Studies and the Center on the Educational Quality of the Workforce (agreement number R117Q00011-91, as administered by the Office of Educational Research and Improvement, U.S. Department of Education).

Suggested Citation

Bishop, J. (1994). *The impact of curriculum-based examinations on learning in Canadian secondary schools* (CAHRS Working Paper #94-30). Ithaca, NY: Cornell University, School of Industrial and Labor Relations, Center for Advanced Human Resource Studies.
http://digitalcommons.ilr.cornell.edu/cahrswp/255

CAHRS / Cornell University
187 Ives Hall
Ithaca, NY 14853-3901 USA
Tel. 607 255-9358
www.ilr.cornell.edu/depts/CAHRS/

CAHRS
Center for Advanced Human Resource Studies

WORKING PAPER SERIES

The Impact of Curriculum-Based Examinations on Learning in Canadian Secondary Schools

John H. Bishop

Working Paper 94-30

The Impact of Curriculum-Based Examinations on Learning in Canadian Secondary Schools

John Bishop
School of Industrial and Labor Relations
Center for Advanced Human Resource Studies
Cornell University
Ithaca, NY 14853-3901

Working Paper 94-30

http://www.ilr.cornell.edu/depts/cahrs

This paper was presented at the NSF/Review of Economics and Statistics conference on School Quality and Educational Outcomes, at the Kennedy School of Government, Harvard University on December 14-15, 1994. It's preparation was possible because of support from the Pew Charitable Trust, The Center for Advanced Human Resource Studies and the Center on the Educational Quality of the Workforce (agreement number R117Q00011-91, as administered by the Office of Educational Research and Improvement, U.S. Department of Education). The findings and opinions expressed in this report do not reflect the position or policies of the Office of Educational Research and Improvement or the U.S. Department of Education. This paper has not undergone formal review or approval of the faculty of the ILR School. It is intended to make results of Center research, conferences, and projects available to others interested in human resource management in preliminary form to encourage discussion and suggestions.

Abstract

Externally set curriculum-based examinations at the end of high school apparently have pervasive backwash effects on middle school students, their parents, teachers and school administrators. Holding the social class background of students constant, students from Canadian provinces with examination systems were substantially (23 percent of a standard deviation) better prepared in mathematics and 18 percent of a standard deviation better prepared in science than students from provinces lacking such exams. The effect of an exam system on mathematics achievement of 13 year olds is larger in a standard deviation metric than the decline in math SAT scores between 1969 and 1980 that has been such a focus of public concern. Other natural experiments yield similar findings. When adjustments are made for ethnicity, gender and social class of SAT test takers, New York State ranks higher on the SAT than any of the other 38 states where the test is taken by large numbers of students. The mathematics and science achievement of Swedish high school seniors declined in the years following the elimination of high/medium stakes curriculum-based exams.

The analysis also found that examination systems had pervasive effects on school administrators, teachers and parents. In the provinces with external exams, schools were more likely to:

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

At home students watch less TV, spend more time reading for fun, and are more likely to report their parents want them to do well in math and science. In addition, parents are more likely to talk to their child about what they are learning at school.

The National Assessment of Educational Progress (NAEP) reports that only 7.5 percent of 17 year old students in the U.S. can "integrate specialized scientific information" (NAEP 1989 p.56) and 6.4 percent "demonstrated the capacity to apply mathematical operations in a variety of problem settings" (NAEP 1988b p. 42). There is a large gap between the science and math competence of American secondary school students and their counterparts in other industrialized nations.

Educational researchers and policy makers have proposed a host of different explanations for these phenomena: the diversity of student backgrounds, insufficient funding, overly large classes, insufficient teacher preparation time, tracking by ability, long school vacations, insufficient time devoted to core academic subjects, classrooms not managed to maximize time on task, too little homework, too much TV, high absenteeism, poor discipline, unsafe schools, poorly trained teachers, too much emphasis on non-academic goals such as self esteem and sports, unmotivated students, peer pressure against studying, school cultures that do not support learning, dumbed down textbooks, unsupportive parents, latch key children, school districts that are too large, the dead hand of public management/bureaucracy, lack of performance incentives for teachers and lack of school choice and competition among schools. What almost all of these proposed explanations have in common is that key actors in the learning enterprise (students, parents, teachers, administrators and/or school boards) are being accused of giving insufficient priority to the goal of academic achievement. Some other goal—e.g. leisure, avoiding controversy, low taxes, equity—is taking precedence over the academic achievement of students.

Regardless of which of these proposed causes of poor academic performance are the most important, a more fundamental question remains: "Why do American students, parents, teachers, administrators and school boards apparently place a lower priority on the goal of academic achievement than their counterparts in Europe and East Asia." In a previous paper I have proposed the following answer to this question:

> The fundamental cause of the low effort level of American students, parents, and voters in school elections is the absence of good signals of effort and learning in high school and a consequent lack of rewards for effort and learning ....In most other advanced countries mastery of the curriculum taught in high school is assessed by ...examinations which are set and graded at the national or regional level. Grades on these exams signal the student's achievement to colleges and employers and influence the jobs that graduates get and the universities and programs to which they are admitted. How well the graduating seniors do on these exams influences the reputation of the school and in some countries the number of students applying for admission to the school. In the United States, by contrast, students take aptitude tests that are not intended to assess the learning that has occurred in most of the classes taken in high school. The primary signals of academic achievement are grades and rank

> in class—criteria which assess achievement relative to other students in the school or classroom, not relative to an external standard (Bishop 1990b).

Thus, the claim is that systems of state or national curriculum-based examinations in the final years of high school transform the incentive environment of students, parents, teachers and school administrators and the result is higher academic achievement.[1]

The paper will proceed as follows. Section 1 elaborates on the argument quoted above and develops 9 specific hypotheses about the impacts of curriculum-based exams. Section 2 discusses how the exam as incentive hypothesis may be best tested and examines existing comparative evidence on the issue. As predicted the U.S. and Sweden lag behind countries with exam systems and students from the only state in the nation with curriculum based exams, New York, do better than students from other states when demographics of the student population are controlled. This paper presents an additional test of the hypothesis using data on 39,000 Canadian 13 year olds. Section 3 contrasts U.S. and Canadian achievement levels. After describing the data in section 4, an analysis of the determinants of math and science test scores is presented in section 5 and an analysis of the impact of provincial exam systems on the behavior of parents, students, teachers and school administrators is presented in section 6. The paper concludes with a summary and some observations about the policy implications.

I. Hypotheses

Student and Parent Behavior: External assessments of achievement in specific high school subjects increase the students' rewards for learning and this should induce the student to choose more demanding courses and work harder in them. When such exams are absent, many students choose courses that have the reputation of being fun and not requiring much work to get a good grade. As one student who had avoided the harder courses even though she was sure she could do the work explained her decision: "Why *should I do it,* [the extra work], if *I don't have to?"* (Ward 1994) Teachers know this and adjust their style of teaching and their homework assignments with an eye to maintaining enrollment levels:

> *An angry math teacher [who remembering] the elimination of a carefully planned program in technical mathematics for vocational students simply because not enough signed up for it,...[said] 'Its easy to see who really makes decisions about what schools teach: the kids do.'* (Powell, Farrar and Cohen 1985, p. 9)

External assessments also have pervasive effects on the **structure** of student rewards. When signals of achievement assess performance relative to fellow students (e.g. grades and class rank) rather than relative to an absolute standard, students have a personal interest in persuading each other not to study. The studious are called nerds, in part, because they are making it more difficult for others to get good grades. Since devoting time to studying for an exam is costly, the welfare of the entire class is maximized if no one studies for exams which are graded on a strict curve. The cooperative solution is "no one studies more than the minimum." Participants are generally able to tell who has broken the "minimize studying" code and reward those who conform and punish those who do not. Side payments and punishments are made in a currency of friendship, respect and ridicule that is not limited in supply. For most students the benefits that might result from studying for the exam are less important than the very certain costs of being considered a "brain geek", "grade grubber" or "acting White," so most students abide by the "minimize studying" "don't raise your hand too much" norm. Most American students are part of friendship circles in which the following norms prevail: *It is OK to be smart. You cannot help that. But, it is definitely not OK to spend a lot of time studying. Instead, use your free time to socialize, participate in athletics or earn money.*

When learning is assessed relative to an outside standard, students no longer have a personal interest in getting the teacher off track or persuading each other to refrain from studying. Consequently, our theory predicts that:

H.1—Educational systems employing external assessment will have higher achievement, even when student characteristics, school resources, curriculum, teacher qualifications and teaching techniques are held constant. The effects should be strongest in 11th and 12th grade, but they should reach down to lower grades as well.

Student attitudes toward math and science might also be expected to change. In many European countries, job applications ask students to report their final grades in examination subjects (Raffe 1984). Resumes also conventionally include this information. In countries where such customs prevail one would expect students to be more likely to report that knowledge of examination subjects is important for getting a good job. Therefore:

H.2—If the existence of an examination system (and the consequent greater reliability and validity of student grades) increases the proportion of employers that use grades in examinations subjects to screen job applicants, students should be more likely to report that an examination subject is important for getting a good job.

But the effects should not stop there. External exams also transform the incentives faced by the adults in the system. I would expect parents to set stricter limits on the amount of TV their children can watch and to be more likely to talk to their child about what they are learning in examination subjects. Our next set of hypotheses, therefore, are that:

H.3—External exams will result in students spending less time watching TV.

H.4—External Exams will induce parents to spend more time talking with their children about school and result in student's perceiving their parents to be more interested in their doing well in examination subjects.

Those who oppose external exams argue that they will cause students to lose their desire to learn for its own sake and concentrate solely on learning what is likely to be on the exam. If they are right, students in systems with external exams should be less likely to read for pleasure or watch science programs like NOVA and Nature. Therefore, Hypothesis #4 is that:

H.5--Students will spend
- ***** less time watching science documentaries on TV and
- ***** less time reading for fun.

Teacher Behavior: Teachers are inevitably held somewhat accountable for how many of their students pass their courses. In systems without external exams, a teacher can lower class failure rates by lowering the passing standard. Indeed, American teachers often report being pressured by administrators to pass students. When asked "Is there pressure on your teachers to pass students who don't earn a passing grade?", 58 percent of local teacher union leaders in New York State responded YES.

> *Some school districts have a policy that each teacher's student grades are printed out and distributed among the faculty. Those teachers with the lowest student averages are singled out for "discussions" about why their students are lower.* (New York State United Teachers, March 3, 1994, p. 2).

Students also pressure teachers to set low standards. Theodore Sizer's description of Ms. Shiffe's biology class, illustrates the difficulties that some teachers get into:

> *She wanted the students to know these names. They did not want to know them and were not going to learn them. Apparently no outside threat-flunking, for example--affected the students. Shiffe did her thing, the students chattered on, even in the presence of a visitor ...Their common front of disinterest probably made examinations moot. Shiffe could not flunk them all, and, if their performance was uniformly shoddy, she would have to pass them all. Her desperation was as obvious as the students' cruelty toward her.* (1984 p. 157-158)

Parents often push for lower standards as well. In one school:

> *Students were given class time to read The Scarlet Letter, The Red Badge of Courage, Huckleberry Finn, and The Great Gatsby because many would not read the books if they were assigned as homework Parents had complained that such homework was excessive. Pressure from them might even bring the teaching of the books to a halt ....[As one teacher put it] '7f you can't get them to read at home, you do the next best thing. It has to be done ....I'm trying to be optimistic and say we're building up their expectations in school."* (Powell, Farrar and Cohen 1985, p.81)

Some exceptional teachers are able to induce students outside the honors track to undertake tough learning tasks. But too often academic demands are compromised because the bulk of the class sees no need to accept them as reasonable and legitimate. Under a system of external exams, teachers and local school administrators lose this option. Their response will be to strive to prepare their students for the external exam. And because of the exam, they will find they have more cooperative students. Therefore, Hypothesis #6 is that:

H.6—External exams will induce teachers
(a) to set higher standards,
(b) to assign more homework,
(c) to increase the number of experiments that students do in science class
(d) to have students solve mathematics problems alone rather than in groups,
(e) to give more quizzes and tests,
(f) to increase their use of other teaching strategies which they believe improve exam performance.

On the further assumption that teachers know which teaching strategies are most effective in improving exam results, our seventh Hypothesis is that:

H.7—Teaching strategies that regression analysis finds are effective should be more common in educational systems with external exams.

School Administrator Behavior: The behavior of school administrators and local school boards is also influenced by how student achievement is signaled to others. In the U.S., academic achievement must compete with the other school goals. American schools are also expected to foster self-esteem, to provide counseling, supervised extra-curricular activities, musical training, health services, community entertainment (e.g. interscholastic sports), drivers education and to do all this in a racially integrated setting. These other goals require additional staff and different kinds of staff. They may not be served by hiring teachers with a strong background in calculus or chemistry.

Some American school administrators focus on lowering the failure rate rather than raising achievement. A principal who had recently fired a teacher for failing too many of her students justified his decision with the following:

> *'I have made it very clear that one of my goals is to decrease the failure rate, to make sure the kids feel good about learning, stay in class, stay in school and do well .... Math is just a big body of knowledge; what is Algebra II across the nation anyway?"* he asks. When he taught band, he adds, he certainly didn't expect kids to finish the year as musicians—but he did want them to know more about music than they did before ....All the talk about preparing students for college struck him as *"ludicrous."* Instead the goal should be to keep students studying math (Bradley, Sept 19, 1993 p. 19, 20).

When there is no external assessment of academic achievement, students and their parents benefit little from administrative decisions that opt for higher standards, more qualified teachers or a heavier student work load. The immediate consequences of such decisions—higher taxes, more homework, having to repeat courses, lower GPA's, complaining parents, a greater risk of being denied a diploma—are all negative. Since college admission decisions are based on rank in class, GPA and aptitude tests, not externally assessed achievement in high school courses, upgraded standards will not improve the college admission prospects of next year's graduates. Graduates will probably do better in difficult college courses and will be more likely to get a degree, but that benefit is uncertain and far in the future. Maybe over time the school's reputation and, with it, the admission prospects of graduates will improve because the current graduates are more successful in local colleges. That, however, is even more uncertain and postponed. As a result, school reputations are determined largely by things that teachers and administrators have little control over: the socio-economic status of the student body and the proportion of graduates going to college.

The Scholastic Aptitude Test is no substitute for curriculum based exams because it does not assess knowledge and understanding of science, history, social science, statistics and calculus or the ability to write (Jencks and Crouse 1982). Consequently, parents can see that improving the teaching of these subjects will have only minor effects on how their children do on the SAT, so why worry about standards? In any case, doing well on the SAT matters only for those who aspire to attend a selective college. Most American students plan to attend open entry public colleges which admit all high school graduates from the state with the requisite courses. External exams in high school subjects can be expected to transform the signaling environment. There is now a very visible payoff to hiring better teachers and improving the school's science laboratories. Larger numbers of students pass the external exams and this in turn influences college admissions decisions. School reputations will now tend to reflect student academic performance rather than the family background of the community or the success of football and basketball teams. If additionally parents and students can choose which high school to attend and aid from higher levels of government is based on enrollment, the stakes for the school administrators become very high indeed. Poor student performance on the external exams might force layoffs of school staff. Hypothesis #8, therefore, is that:

H. 8—External exams will cause priorities to shift in favor of achievement in examination subjects. Administrators and school boards will be induced:

(a) to improve the school's science laboratories (if science is an examination subject).
(b) to increase the share of the school week devoted to examination subjects,
(c) to lengthen the school day and school year,

(d) to use specialist teachers with a thorough background in the field to teach examination subjects.
(e) to pay higher salaries.
(f) to recruit and retain more experienced teachers.
(g) to reduce class size in examination subjects.
(h) to give teachers additional preparation time.

While these effects should be most pronounced in senior high school, effects on the policies of middle schools are anticipated as well. On the further assumption that administrators know how to allocate school resources to improve student achievement when that becomes a higher priority goal, our ninth Hypothesis is that:

H. 9—School resources and policies that regression analysis finds are effective should be more common in educational systems with external exams.

II. Testing Hypotheses about the Effects of Examination Systems

Only a few countries (the U.S., Sweden and Latin America are the major examples) lack curriculum-based external exams at the end of secondary school (Noah and Eckstein 1988, Madeus and Kellaghan 1991). There is a positive correlation between such examination systems and national mean achievement levels, particularly when standards of living are controlled (Bishop 1990). This is consistent with the causal hypotheses presented above. Causation is not proved, however, because other explanations for the U.S. lag are available—e.g. lack of a Confucian culture, lower standards for entry into secondary school teaching—that fit the aggregate data just as well. If comparisons between the U.S. and the rest of the industrialized world cannot yield decisive evidence that exam systems cause performance differentials, other sources of variation in curriculum based exams need to be analyzed. Best of all would be studies which hold national culture constant.

Effects of Changes in Examination Systems

One approach would be to study the effects of changes in examination systems. There are, however, three problems with this approach. The first problem is that changes in examination systems will influence student achievement outcomes only with a long lag. An 18-year-old's skills in mathematics are the product of 12 years of education, not just what happened during the previous year. But even more important is the conservative nature of schools as institutions. School cultures are very slow to change. Teachers who have adapted to a system of external exams by setting high standards for their students will not quickly lower their standards when they are given more authority over final senior year grades. Parental and public attitudes about learning and the proper goals and priorities for schools also shift at a glacial pace. Consequently it will probably take more than a decade for most of the effects of changes in examination regime to be felt. Over that time frame schools and societies change in so many other ways, it will be hard to distinguish the effects of changes in examination regime from the effects of other phenomena.

The second problem is the marginal character of most of the changes that have been made. Many countries have tinkered with the mechanism for deciding who is admitted to different kinds of post-secondary education or have lowered the weight of the external exam when assigning course grades (Noah and Eckstein 1988). Case studies I recently conducted of France, the Netherlands, England and Scotland suggest, however, that the weight of the external exam in the final grade or college admissions decisions is not critical (Bishop 1993). What appears to make the difference is first, the very existence of the exam system and the publicity attached to the results and second, the proportion of the age cohort that expects to take an external exam. Exams for an elite segment of an educational system influence standards in that segment, but may have limited effects on average performance levels in the system as a whole. It is not essential that everyone take the same exam. In most countries the type of school one attends and the specialty one has chosen determine which exams a student takes. Indeed requiring everyone to take the same exam appears to force a lowering of the passing standard. The passing standard of the exams, of course, also matter, but the passing standard is determined endogenously with average performance levels, so one cannot treat the passing standard as an exogenous determinant of performance.

The third problem is the limited number of countries for which it is possible to track performance levels over time. Sweden appears to be the only country for which such data is available which has made a non-marginal change in its examination system. Curriculum-based external examinations which influence student grades were eliminated during the 1970s.[2] As predicted, this has coincided with a decline in the numbers of Swedish students taking rigorous mathematics and science courses.

Eight countries participated in both the First International Mathematics Study in 1964 and its replication in 1982 (see figure 1 and 2). **The proportion of Swedish 18 year olds taking college-prep mathematics fell from 16 percent to 12 percent. This more selected group of students scored slightly higher in algebra and lower in geometry on the anchor items that appeared in both assessments.** Finns, by contrast, simultaneously increased the proportion of the age cohort taking college prep mathematics from 7 to 15 percent and significantly improved their mean scores (Husen 1967; Robitaille and Garden 1989).

INSERT FIGURE 1 and 2

Figure 1

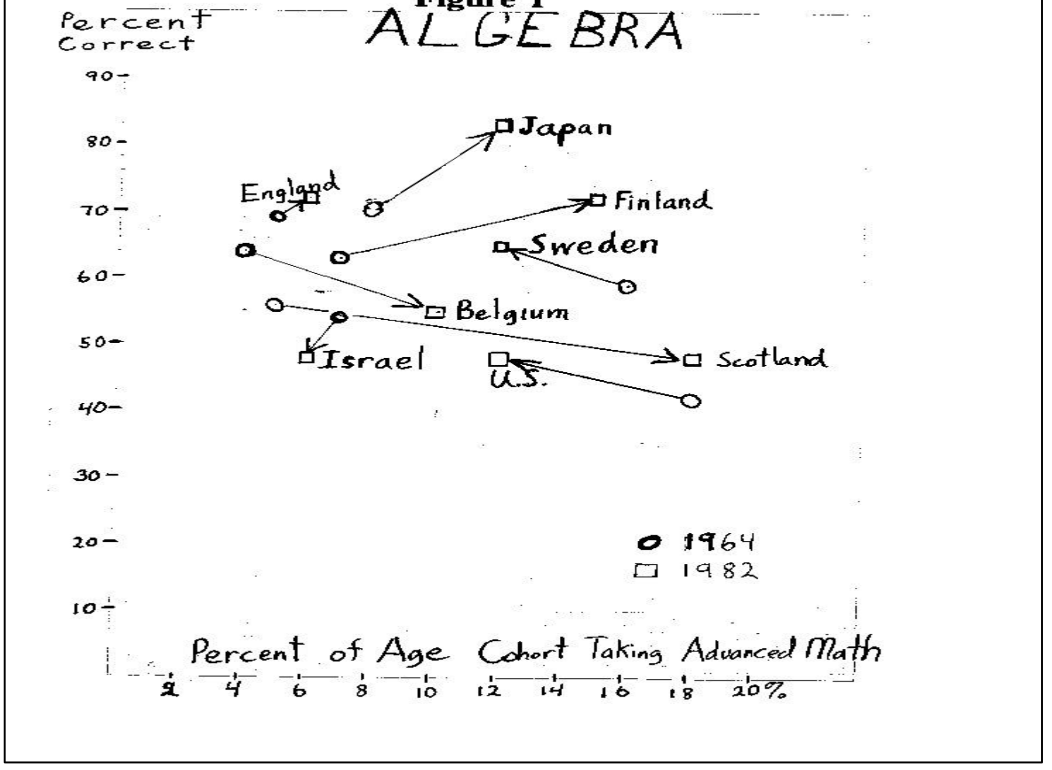
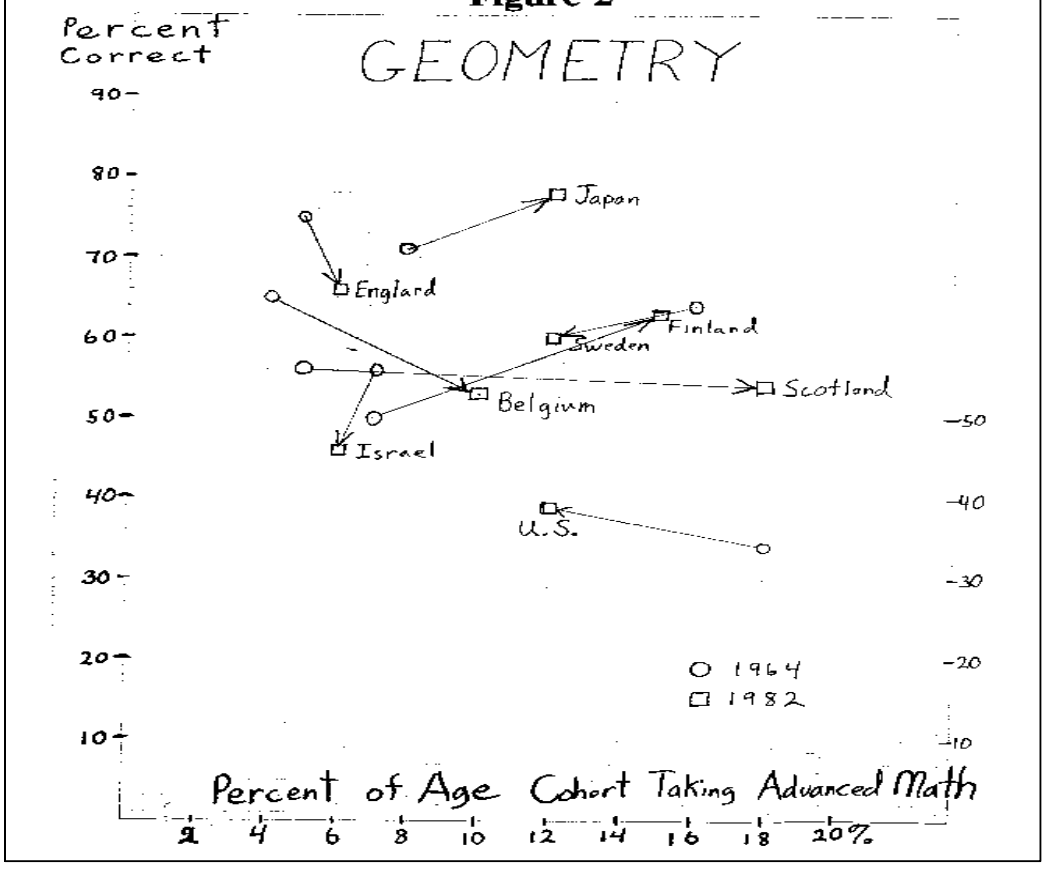


Figure 2

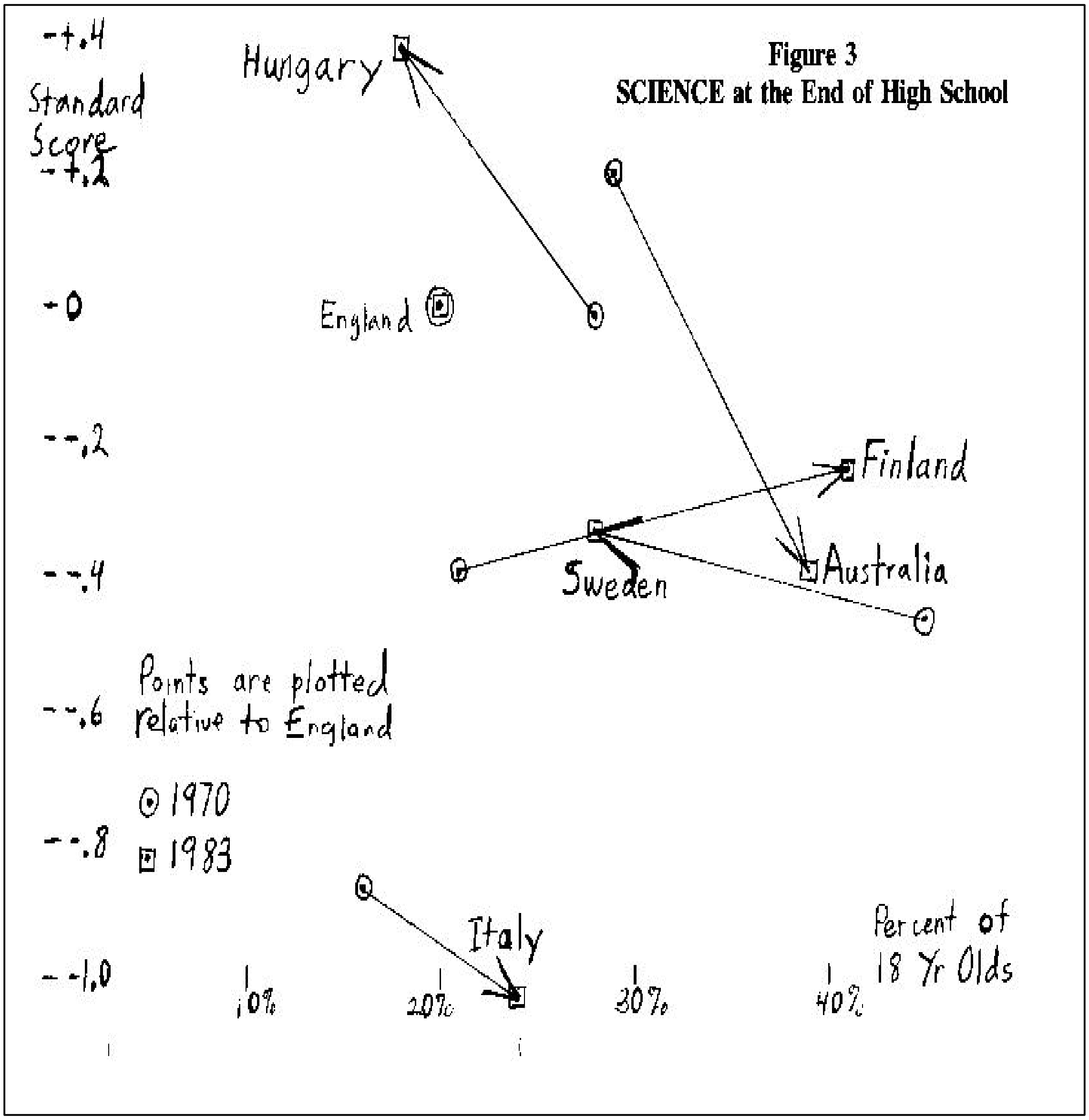


Simultaneously, **the proportion of Swedish 18 year olds in academic lines of study that were given the IEA science exam fell from 45 to 28 percent between 1970 and 1983.** This much smaller and surely more able group of Swedish students only slightly improved their position relative to England (see Figure 3)[3] Finland, by contrast, increased the proportion of the age cohort assessed in science from 21 to 41 percent and simultaneously improved its score relative to England. Other countries—Australia, Italy—had declines in relative test scores, but they were associated with large increases in the proportion of 18 year olds in the academic lines of study that were tested. Hungary was the only other country to reduce the share of 18 year olds that participated in the IEA science exam and it experienced a substantial increase in relative test scores (Comber and Keeves 1973 p. 168; Postlethwaite and Wiley 1994 p. 74).[4]

While supportive of the hypothesis, the pattern of changes in Swedish test scores is once again not decisive evidence. Additional tests of the hypotheses are needed.

INSERT FIGURE 3

Figure 3
SCIENCE at the End of High School



Cross-Sections within a Country

Probably the best way to test these hypotheses is to limit the study to just one country and compare jurisdictions with different kinds of examinations systems. The jurisdictions must be reasonably large, however, for otherwise colleges and employers are not likely to use grades on the curriculum-based exams in their selection decisions, so the rewards for doing well may be quite limited.

New York vs. the Rest of the United States: New York State is reasonably large and has a Regents Examination system which reaches over half of the state's high school students.[5] It is, indeed, the only state in the U.S. with a curriculum based examination system covering the majority of high school graduates.[6] California is currently trying to introduce one. Consistent, with our hypothesis, the Regents exams (or something else unique to New York State) appear to have raised statewide achievement levels. When the family income, parental education, race and gender of SAT test takers are controlled, New York State has the highest adjusted mean Scholastic Aptitude Test score of the sample of 38 states with adequate numbers of test takers to be included in the study (Graham and Husted, 1993).[7] This occurs despite that fact that Regents exam grades account for less than half of the course grade and influence only the type of diploma received. A passing score on Regents exams is not necessary for admission to non-university higher education and employers ignore exams results when they make hiring decisions. Scholarships awarded by the state government have been based on aptitude test scores, not Regents exam results.

Comparing Canadian Provinces: Probably the best place to test hypotheses about the impact of curriculum-based external examinations is Canada. Some Canadian provinces have curriculum-based exams; others do not. At the time the data used in this study was collected, 1990-91, Alberta, British Columbia, Newfoundland and Quebec had curriculum-based criterion-referenced examinations in English, French, mathematics, biology, chemistry, physics and many other subjects. These exams accounted for 50 percent of the final grade in Alberta, Newfoundland and Quebec and 40 percent in British Columbia. New Brunswick had such exams in language arts and mathematics but not in science or other subjects. Local school districts decided how much weight the New Brunswick Provincial Achievement Exam was to be given in assigning final grades.

The other provinces had no curriculum-based provincial 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, proofreading, mathematics, science and social studies which do not influence student grades.

Many of these provinces are now in the process of reintroducing curriculum-based provincial exams (GAO 1993).[8] Canadian data is, therefore, used to test the hypotheses presented above.

III. Mathematics and Science Achievement—Canada vs. the U.S.

In terms of governance and finance, the Canadian system of elementary and secondary education is more similar to America's system than almost any other nation. As in the U.S., education is a provincial/state responsibility. Localities administer schools and use the property tax to raise their share of the funding. In 1980 localities accounted for 43 percent of the funding in the United States and 28.5 percent in Canada.[9] Comprehensive schools predominate in both countries.

Despite these similarities, Canadian secondary schools appear to do a significantly better job of teaching mathematics and science than American secondary schools. Table 1 presents comparative data on the performance of primary school and lower secondary school students. On the International Association for the Evaluation of Educational Achievement (IEA) mathematics exam given in the early 1980s American 13-14 year olds scored 6 percentage points below their counterparts from British Columbia and 3.4 percentage points below students from Ontario (McKnight et al. 1987). In science English speaking Canadians were ranked 4th, French speaking Canadians ranked 9th and the U.S. ranked 16th among 17 industrialized countries. After a rough adjustment for age differences, American students lagged about 44 percent of a U.S. standard deviation (about one and one-third grade level equivalents) behind Canadian students.

Table 1
Gains in Achievement in Lower Secondary School

	1991 IAEP Science		1991 IAEP Math		1991 IEA Reading (Age Adjusted)		1983 IEA Science		1982 Math
	Gain 9-13	Level at Age 9	Gain 9-13	Level at Age 9	Age 9	Age13	Gain 10-14	Level at Age 10	Level Age 13
Canada	20.2	58.9	31.0	45.1	514	540	1.31	58.3	50.9
United States	17.2	60.5	25.4	45.6	543	528	1.13	54.8	46.2
		(16.2)			(21.1)	(74)	(85)		(14.0)

Columns 1 & 3 are the increase in the percent correct on items common to the tests given to 9 and 13 year olds. Columns 2 and 4 are the mean percent correct on these common items at age 9 (IAEP, 1992a 1992b). Columns 5 and 6 are the age adjusted mean reading scale scores for ages 9 and 13. Canada was represented by British Columbia in this study (Elley 1992, 108-109). Column 7 is the differential in standard score units between students 10 years old and students 14 years old on the test given in the Second International Science Study sponsored by the International Association for the Evaluation of Educational Achievement. Column 8 is the percent correct for students aged 10 years old (Postlethwaite and Wiley, 1992, p. 55 & 74). Column 9 is a weighted mean of percent correct for eighth graders in the Second International Mathematics Study. Canada was represented by the provinces of Ontario and British Columbia in this study (McKnight et al 1987, p. 124). The numbers in parenthesis in the third row are standard deviations.

At the end of high school, significantly larger numbers of Canadian's were pursuing mathematics and science at an advanced level. Data on performance differences at the end of secondary school are given in Table 2. Americans who participated in the Second International Math Study were high school seniors in college preparatory math courses like trigonometry, pre-calculus and calculus. This very select group, representing 13 percent of American 17-18 year olds, got 39.8 percent of the questions correct. Thirty percent of seniors in British Columbia high schools study mathematics at this level and they got almost as high a score. The 19 percent of the 13th graders studying mathematics in Ontario got 51.2 percent of the questions correct (McKnight et al. 1987).

Table 2
Achievement at the End of Upper Secondary School

	1982 IEA Mathematics Final Yr. of Sec. Sch.			1983 IEA Science—Final Year of Upper Secondary School: Physics			Chemistry			Biology		
	Percent Correct	%Age Group	Mean Age	Percent Correct	%Age Group	Mean Age	Percent Correct	%Age Group	Mean Age	Percent Correct	%Age Group	Mean Age
Canada												
English Spk				41.7	18	18:4	39.6	25	18:4	43.7	28	18:2
French Spk				26.7	35	17:1	27.9	37	17:1	39.3	7	17:2
British Col.	37.8	30	17:9									
Ontario	51.2	19	18:5									
United States	39.8	12	17:8	45.3	1	17:10	37.7	2	17:8	38.1	12	17:5

Column 1 is a weighted mean percent correct for students taking a college prepatory mathematics in the final year of secondary school from the Second International Mathematics Study (McKnight et al 1987, p. 124). The mean ages were the same in the US and Canada. Column 2 is the share of the age cohort in the college preparatory mathematics courses included in the study. Column 3 is the mean age of these students. Students in Ontario Canada are older because they were tested in their 13th year of schooling. Columns 4, 7 & 10 give the percent correct for students studying each science subject in the final year of secondary school. Column 5, 8 and 11 are the proportions of the age cohort taking each science subject in the final year of secondary school [for the U.S. it is the share of students taking their second year of the subject]. Columns 6, 9 and 12 give the mean age of the students taking the IEA test. (Postlethwaite and Wiley, 1992, p. 6, 36, 37 & 69).

The twenty-five percent of Canadian 18 year olds in English speaking schools who were studying chemistry in their final year of high school did almost as well as the elite 2% of American high school graduates taking their second year of chemistry, most of whom were in Advanced Placement classes (Postlethwaite and Wiley 1994). The 28 percent of English speaking Canadians taking biology were much more knowledgeable than the 12 percent of American youth who were taking a second year of biology during their senior year. The Francophone high school seniors tested in chemistry and physics represented 35 to 37 percent of the age cohort. About half of them were from Quebec and were therefore tested in their 11th year of school. The age difference and the lack of selectivity explains the lower scores of the Francophone high school seniors.

A Canadian advantage is also apparent in studies conducted in the early 1990s. In the 1990/91 IEA study of reading achievement, age adjusted scores indicate that Canadian fourteen year olds were reading about 15 percent of a U.S. standard deviation (about one-half of a grade level equivalent) better than American fourteen year olds (Elley 1992 p. 109). Science achievement of Canadian 13 year olds was 11 percent of a U.S. standard deviation (about one-third of a grade level equivalent) ahead of American 13 year olds.[10] Mathematics achievement of Canadian 13 year olds on the 1991 IAEP was one-third of a U.S. standard deviation (about one grade level equivalent) above the American mean (IAEP 1992a).

It is Canada's superior system of secondary education that produces these results, not its primary schools. Indeed, in the most recent studies by the IAEP and the IEA, Canadian 9 and 10 year olds lag slightly behind their American counterparts in reading, mathematics and science (see Table 1). Thus, Canadian students enter lower secondary school behind American students, but by age 13-14 Canadians have taken a substantial lead in mathematics and a small lead in science. What is it about Canadian secondary education that produces these positive outcomes? Is the system of provincial exams that prevails in more than half of Canadian provinces responsible for the Canadian advantage in mathematics and science? This hypothesis is tested below.

IV. The IAEP Data

The hypotheses outlined in section I and III were tested in data on the mathematics and science competence of 42,241 Canadian and American 13 year olds from the International Assessment of Educational Progress (IAEP). When the Educational Testing Service canvassed countries about participating in the 1991 International Assessment of Educational Progress, Canada decided to collect sufficient data to allow valid comparisons between provinces and between the Anglophone and Francophone school systems of the five provinces with dual systems. The Yukon, the Northwest Territories and Prince Edward Island did not to participate in the study. Stratified random samples of 105 to 128 secondary schools were selected from the French speaking school systems of Ontario and Quebec and from the English speaking school systems of Alberta, British Columbia, Manitoba, Saskatchewan, Ontario, Quebec, New Brunswick, Nova Scotia and Newfoundland. The United States sample contained 106 schools. A school's likelihood of selection was roughly in proportion to its estimated number of 13 year olds. All French speaking schools in New Brunswick, Saskatchewan and Manitoba were invited to participate.

In the United States 21 percent of the schools invited to participate declined. In Canada the highest school refusal rates were for the English speaking schools in Quebec (15 %), the English speaking schools in Saskatchewan (12 %) and the French speaking schools in New Brunswick (12 %). In the rest of the provinces refusal rates were below 7 percent and in many provinces all invited schools participated. When sampled schools declined to participate an alternate was selected from the same stratum (IAEP 1992c).

Random samples of 30 to 34 thirteen year olds were selected from each school. Half were assigned to the mathematics assessment and half assigned to the science assessment. Students also completed a brief questionnaire that asked about books in the home, number of siblings, language usually spoken at home, hours watching TV, hours doing homework, pleasure reading, watching science programs on TV, home availability of mathematics and science resources, and attitudes towards math and science. Students were also asked to describe how frequently they took tests or quizzes, listened to the teacher give a ______ lesson, solved mathematics problems in groups, worked alone on mathematics problems, did experiments, watched experiments and watched science films. The principals of participating schools also completed questionnaires describing school policies, school resources and the qualifications of 8th grade mathematics and science teachers.

V. Determinants of Achievement in Math and Science at Age 13

In this first stage of the analysis the dependent variable is the percent correct with adjustments for guessing. It is defined as the (number of correct answers minus .25 times the number of answered questions) all divided by (.75 times the number of items on the test). Adjusted for guessing, the students examined here got an average of 47.2 percent in math (SD= 24 points) and 57.3 percent in science (SD =20.2 points). The independent variable of greatest interest is EXAM, a dummy variable that equals 1 if one's province has a curriculum-based provincial exam in the subject at the end of high school and zero otherwise. In the model predicting math score, students in Alberta, British Columbia, New Brunswick, Newfoundland and Quebec are assigned a 1 on EXAM. Students in Ontario, Saskatchewan, Manitoba, Nova Scotia and the United States have a 0 on EXAM. When the science score is predicted, New Brunswick students get switched to 0 on EXAM.

Table 3 presents our estimates of the impact of curriculum-based exams taken in 12th grade on test scores in 8th grade. Odd numbered columns present results for math. Even numbered columns present results for science. The first three rows present the results of simple regressions containing no controls for personal or school characteristics. Consistent with the findings of section III, American students score 8.6 percentage point below the Canadian sample in mathematics and 1.9 points below them in science. Row 2 tells us that students in provinces with exam systems score 6.3 points higher in math and 5.6 points higher in science. Adding controls for USA and for Francophone school (see row 3) slightly reduces the estimated impact of the exam to 5.5 points for math and 5.2 points for science. The coefficient on the USA dummy falls by about 35-40 percent, suggesting that exam systems are one of the reasons why Canadian students outperform American students. But the significant negative coefficients on the USA dummy in the mathematics equation suggest that there are also other reasons for the American lag in mathematics.

Table 3
Impact of Curriculum-Based Exams on Math and Science Achievement

Controls in Model	SUBJECT EXAM Math	SUBJECT EXAM Science	UNITED STATES Math	UNITED STATES Science	FRENCH SPK SCHOOL Math	FRENCH SPK SCHOOL Science	Adj R^2 Math	Adj R^2 Science
1. None			-.086 (12.5)	-.019 (3.3)			.0076	.0005
2. None	.063 (18.6)	.056 (19.0)					.0160	.0180
3. None	.055 (15.6)	.052 (17.1)	-.056 (7.8)	-.012 (1.9)	.007 (1.9)	-.049 (14.6)	.0192	.0284
4. Family Demog-Bks in home, sibs, gender, lang	.054 (15.4)	.044 (15.4)	-.035 (5.1)	.002 (.4)	.036 (8.5)	-.019 (5.7)	.1052	.1307
5. Add Mean School Demog. Books, Languages	.056 (15.6)	.037 (12.8)	-.004 (.5)	.015 (2.6)	.100 (19.4)	.025 (6.2)	.1292	.1431
6. Add TV & Sch mean TV Reading, MathSci Resources	.051 (14.3)	.030 (10.5)	-.003 (.4)	.018 (3.2)	.073 (13.7)	-.002 (.5)	.1672	.1887
7. Add Student Attitudes	.051 (14.5)	.032 (11.1)	-.002 (.2)	.019 (3.6)	.068 (12.9)	.005 (1.1)	.1869	.1968
8. Add School Resources & Policies	.035 (8.8)	.031 (9.4)	-.025 (3.2)	.012 (2.0)	.066 (12.2)	.008 (1.7)	.1959	.2015
9. Add Teaching Strategies	.036 (9.3)	.035 (10.4)	-.004 (.5)	.030 (4.9)	.040 (6.3)	.017 (3.7)	.2202	.2105

Now let us examine what happens in row 5 when controls are added for the demographic background of the student and the school's student body—books in the home, number of siblings, gender, languages usually spoken at home and school means for books at home and proportion of students whose home language is different from the language of instruction. The estimated effects of exam systems on test scores remain strong: 5.6 percentage points for math and 3.7 points for science. The variables to be added to the regressions in models 6 through 9—TV watching, home math-science resources and activities, attitudes, school resources and policies and teaching strategies—are hypothesized to be influenced by the existence of external exams. Consequently, row 5 presents our best estimate of the total impact (including indirect effects) of having a provincial exam in the subject at the end of secondary school on IAEP test scores at age 13. Provincial exams have very large effects: 23.3 percent of a standard deviation (about four-fifths of a grade level equivalent) in mathematics and 18.3 percent of a standard deviation (about two-thirds of a grade level equivalent) in science. **In a standard deviation metric, the impact of provincial exams on math scores is larger than the 22.6 percent of a standard deviation decline in Math SAT test scores between 1969 and 1980 and 2.8 times the magnitude of the increase in Math SAT scores since 1980.**

Adding the demographic variables also causes the USA dummy to fall to zero. This implies that the 8.6 point gap between the math scores of Americans and English speaking Canadians can be fully explained by a combination of demographics and exam system effects. On the other hand, in mathematics Americans and Canadians attending English speaking schools lag way behind students from the same social class at French speaking schools in Canada. When one controls for the number of books in the home, the number of siblings, whether the language usually spoken at home is same as the school's language of instruction, and school means for books and language, students at Francophone schools score 10 points higher in math and 2.5 points higher in science.

Row 6 presents the results of adding 5 to 6 measures of family behavior and math-science resources to the regression. Four of the additional variables—time spent watching TV, school mean on TV watching, a reading for fun index and school mean on a home resources for math and sciences index—appear in both the math and the science regressions. The additional variable in the science regression is a watching science programs on TV index. The mathematics regression has two additional variables: a dummy for owning a calculator and a dummy that equals one if the student says yes to: "Do you ever use a computer for school work or homework?" Adding these variable substantially increases R squares. The coefficients on EXAM fall by 9 percent in the mathematics regression and by 19 percent in the science regression. It would appear that induced changes in home behavior (particularly TV watching) account for an important part of the impact of provincial exams on test scores at age 13.

Adding student and parental attitudes to the model has no effect on the coefficient on EXAM, suggesting that the three attitude questions available in this data set do not mediate the effects of provincial exams (contrast rows 6 and 7). Similarly adding teaching strategies to the model (contrast rows 8 and 9) fails to reduce the estimated effect of EXAM suggesting that teaching strategies also do not mediate the effects of provincial exams. However, adding school resources and policies to the model (contrast row 7 and 8) reduces the EXAM coefficient in the mathematics regression by 29 percent of its original (.056) level suggesting that school resources and policies do mediate some of the effects of provincial exams.

When a full set of controls for teacher, administrator and parent behavior are included in the model, EXAM continues to have major effects. The provincial exams appear to raise percent correct in both subjects by 3.5 to 3.6 points or 15 to 17 percent of a standard deviation. When we control for home characteristics, school resources and policies and teaching strategies, American 13 year olds are equal in math to Canadians at English speaking schools but 4 points below Canadians at French speaking schools. In science Americans are equivalent to Canadians at French speaking schools and 3 points ahead of Canadians at English speaking schools.[11]

Effects of School Demography on Math and Science Achievement: Table 4 presents the full set of results when all school characteristics and teaching strategy variables are included in the regression. The average number of books in the homes of the students of the school has large effects. Controlling on the actual number of books in the student's own home, doubling the school mean number of books at home increases math scores by 5.6 points and science scores by 3.5 points. Coming from a home where a different language is spoken than the language of instruction substantially lowers science scores but has little effect on math scores.

Table 4
The Determinants of Math and Science Achievement at Age 13

	Standard	Math Percent Correct		Science Percent Correct	
School Demography	Deviation	Coef	T Statistic	Coef	T Statistic
Log Avg Books in Home	.44	.080	(14.00)	.050	(10.33)
Speak Different Lang in Home	.21	.016	(1.45)	-.045	(4.97)
Sci-Math Materials in Home	.65	-.0078	(3.03)	-.0042	(1.91)
TV Hrs/Wk Avg in School	2.85	-.0085	(12.20)	-.0062	(10.85)
TV Hrs/wk of Student	8.65	-.0024	(12.52)	-.0017	(10.56)
Francophone School	.44	.040	(6.29)	.018	(3.73)
Has Curriculum-Based Exam	.50/.48	.036	(9.27)	.035	(10.40)
United States	.25	-.004	(.47)	.032	(4.95)
Non-Public School	.40	-.022	(5.16)	-.031	(8.03)
School Administrator Behavior					
Specialist Teachers in Subject	.50/.50	.020	(5.41)	.006	(1.93)
Teachers Studied Subj in Univ	.36/.36	.007	(1.46)	.003	(.83)
Specialized Science Laboratory	.92	.013	(6.92)	.005	(3.54)
Tracking in 8th Gr in Subject	.38/.27	.015	(3.31)	.009	(1.79)
w/Class Abil Group in Subj.	.41/.30	-.018	(4.76)	.001	(.34)
Log Size of 8th Grade	.93	-.0037	(1.81)	-.0052	(3.11)
Propor. Exper. Teachers	.22	.018	(1.82)	.008	(1.03)
Propor. New Teachers	.14	-.002	(.12)	.012	(.84)
Class Hours in Subject	.86/.79	.0002	(.13)	.0092	(5.10)
Teacher Behavior					
Avg. Homework Hrs/wk in Subject	.62/.44	.0074	(2.86)	.0044	(1.43)
Math Difficulty Index	.41	.000	(.06)	-.0024	(.75)
Quiz Index for Subject	.65/.56	-.0281	(10.33)	-.0280	(10.94)
Listen to Teacher Index	.77/.70	-.0365	(14.49)	-.0030	(1.40)
Math Group Problem Solving	.96	-.0228	(13.58)		
Math Work Alone on Problems	.70	.0239	(10.50)		
Science Do Experiments Ind.	.86			.0000	(.05)
Science Watch Experiments	.71			-.0116	(5.77)
Science Watch Films Index	.66			-.0111	(5.46)
Student Behavior and Attitudes					
Read for Pleasure Index	1.09	.019	(12.63)	.024	(19.25)
Have used Computer for Sch Wrk	.49	.027	(8.16)		
Have a Calculator	.31	.018	(3.59)		
Watch Science Programs on TV	.98			.0156	(10.98)
Subject Important to get Job	.63/.86	.0232	(8.60)	.0181	(10.98)
Parent want me to do well in Subj	.61/.95	.0110	(4.06)	.0106	(7.17)
Subj. Useful in Everyday Life	.87/.95	.0314	(16.39)	-.0103	(6.78)
Family Demographics					
Female	.50	-.026	(8.11)	-.058	(21.82)
Age	.28	.053	(9.77)	.034	(7.45)
Number of Siblings	1.36	-.0105	(9.14)	-.0110	(11.26)
Lang of Instruct & Home Diff--1	.28	-.015	(1.94)	-.016	(2.49)
Lang of Instruct & Home Diff--2	.22	.005	(.67)	-.031	(5.21)
Less than 10 books at Home	.19	-.100	(11.86)	-.107	(15.10)
Betw. 11 & 25 Books at Home	.32	-.063	(11.89)	-.060	(13.89)
Over 100 Books at Home	.50	.045	(12.93)	.037	(12.75)
Adjusted R Square	.2202	.2105			
Mean Square Error			.2179		.1780
Number of Observations			20,232		19,841
Mean Value of Dependent Variable			.474		.576

Students in the IAEP survey reported watching 14.7 hours of television a week. Heavy TV watching by the students at a school is apparently a sign of a school culture that does not encourage and support learning. If you attend a school where students average 20 rather than

15 hours of TV a week, your test scores are predicted to be 4.25 points lower and your science test scores are predicted to be 3.1 points lower. If you copy your friends and spend an extra 5 hours watching TV, your math score is predicted to go down a further 1.2 points and your science score is predicted to decline by 0.9 points.

In their influential 1990 book, John Chubb and Terry Moe argued that the constraints placed on public schools by bureaucracy and democratic government make them inherently less effective than non-public schools that must compete for students and that are, thus, required to survive a market test. Clearly, however, their theory does not apply to Canada. When background characteristics of the students are controlled, students at non-public schools know considerably less math and science at age 13 than public school students. Canadian nonpublic schools lag behind their public counterparts even though they receive considerable funding from provincial governments and local school districts.

As in the U.S., most nonpublic schools in Canada were started by religious denominations. Models were estimated which distinguished different types of nonpublic schools. When social class is controlled, public school students have significantly higher achievement in both math and science than students at religiously controlled schools. Students at secular nonpublic schools are slightly ahead in mathematics but significantly behind in science.

One is forced to conclude, therefore, that nonpublic schools are not inherently more effective at teaching math and science than public schools. My interpretation of these results is that religious schools probably place math and science education lower in their list of priorities than public schools. Their finances may also be more constrained. The religious denominations which control these schools and the parents who send their children to them are primarily seeking better discipline and ethical and moral climates that are consistent with their beliefs, not better math and science teaching. Fewer resources are devoted to math and science teaching and lower achievement results.

Effects of School Resources and Policies: The regressions yield some important insights into the school policies that are associated with high achievement in mathematics and science.[12] Schools that have invested in specialized science laboratories, that use specialist teachers to teach the subject and that have smaller numbers of students in each grade are more effective at teaching science and mathematics. Tracking also appears to help, though in science the effect is not large and only marginally significant. Within-class ability grouping of mathematics classes is associated with significantly lower math scores. Having more experienced teachers is associated with significantly higher scores in mathematics but not in science. When teaching strategies are included in the model, the share of math and science teachers who had studied the subject at university has no significant effect on test scores. Such teachers appear to use more effective teaching strategies, however, because, when teaching strategies are not controlled, the proportion of math teachers who studied the subject at university has a statistically significant positive effect on math scores (not shown). The proportion of teachers with fewer than 3 years of experience has no significant effects on scores.

The average number of classroom hours devoted to mathematics has no effect on math competence but assigning lots of homework does. For science, by contrast, devoting additional classroom hours to the subject has a big effect, while the effect of homework time is smaller and statistically insignificant.[13] This pattern of results suggests that doing problems at home is an effective strategy for learning mathematics, but less effective for science. For science, an hour in the classroom has twice the effect on learning that an hour of homework has.

Effects of Teaching Strategies: The conventional wisdom is that students spend too much time listening to teachers giving chalk and talk lessons or working alone at their desks. According to effective schools research, students do better on standardized tests when they attend schools that frequently test their own students. Science teachers are urged to use the discovery method—lots of experiments particularly experiments done by the student. Math teachers are urged to have their students solve mathematics problems in groups.

This last recommendation will probably sound problematic to economists because of the well know free rider problem. When groups work together to solve mathematics problems, solutions may be found more rapidly, but typically just one or two students are solving the problem. The others in the group may not understand how the problem got solved and may not even be paying close attention. It will often be perceived as uncool to ask for an explanation. Even if it is well explained by other members of the group, learning how someone else solved a problem is quite different from solving it yourself. Tests assess the individual's ability to solve problems, not their ability to solve problems as part of a group.[14]

The IAEP students were asked how frequently—“every day", "several times a week", "once a week", "less than once a week" or "never"—they experienced different types of teaching. Both math and science students were asked about tests and quizzes and about listening to the teacher give lessons. The math students were also asked "How often do you solve problems in small groups during mathematics class?" and "How often do you do mathematics exercises by yourself during mathematics classes?" Students who took the science test were asked how often they watched science films, how often they watched the teacher do experiments and how often they did experiments themselves. The variables employed in the analysis were an average of the school mean response (with one-third weight) and the student's own response (with two-thirds weight).[15]

These indicators of teaching strategy have strong relationships with test scores. Contrary to the effective school research, students in classes where quizzes were common score below students in classes that tested less frequently. Students in classes where the teacher spent a lot of time giving lessons do more poorly in mathematics but not in science. Students who report doing lots of experiments do not perform better than students reporting few such opportunities. At least, however, they are probably not watching science films or watching the teacher perform experiments which is associated with students knowing less science. Spending a good deal of class time in group problem solving activities seems to lower the student's ability to solve problems on their own in a testing situation. Doing exercises alone, on the other hand, is associated with significantly higher performance on the IAEP test. This last finding is consistent with the earlier finding that math homework has a significant effect on IAEP test scores.

Interesting as these relationships may be, one must be cautious about interpreting them as reflecting causality. Longitudinal data on changes in test scores as students progress through school was not generated by the IAEP. The proxy for teaching strategy describes only the methods employed in 8th grade, not earlier grades. Consequently, these results do not tell us much about the effectiveness of different modes of teaching. Teachers may have been varying their strategies based on the background and ability of their students. The findings of effective schools research have received a lot of publicity. Many troubled schools may have adopted the frequent quizzes strategy and this may be the reason why quizzes have a negative relation with test scores. Similar arguments can probably be made about some of the other indicators of teaching strategy.

Effects of Student and Parental Behavior and Attitudes: All of the indicators of voluntary participation in school like activities—reading for pleasure, watching science programs on TV, having a calculator and using a computer for schoolwork or homework-have the expected positive effect on mathematics and science test scores. Positive attitudes also had the anticipated effects. Students who report that their parents want them to do well in math or science do indeed do better. Students who believe that math and science are important for getting a good job also do better on the IAEP tests. Students who strongly agree that "mathematics is useful for solving everyday problems" also do better. Surprisingly, those who make the same claim about science do not do better on the IAEP science test.

VI. The Impact of Exam Systems on Parent, Student, Teacher and Administrator Behavior

The backwash effects of curriculum-based provincial exams on the behavior of 13 year olds, their parents, teachers and school administrators were examined by estimating models predicting these behaviors using schools as observations. The questionnaire completed by the principal provided the measures of school resources and policies. The student questionnaires provide our data on student, parent and teacher behavior. School means on each variable were calculated for the schools with at least 9 students in the school sample and these were the dependent variables analyzed.[16] The specification was the same for all dependent variables. Nine variables were used: 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, a dummy for religiously controlled school, a dummy for secular nonpublic school, a dummy for French speaking school, a dummy for USA and EXAM. For outcome variables specifically associated with either math or science, EXAM's definition was the same as the one used in the Table 4 regressions predicting test scores. When general school characteristics such as TV watching, total homework and class size were predicted, an average of the math and science EXAM variables was used.

The results are presented in Table 5. Each row represents a separate regression on data from 1366 to 1460 schools. The means and standard deviations across schools of each dependent variable are presented in columns 2 and 3. Coefficients on four of the variables—EXAM, USA, French Speaking school and log books in home-- appear in columns 4 through 8. Adjusted R squares appear in column 9.

Table 5
Effect of Curriculum-Based Exams on Parent, Teacher and School Administrator Behavior

	Hyp.	Mean	StdDev Schl	Curric Exam Coef	Curric Exam Tstat	U.S. Coef.	French Coef	LnBookHm Coef	Adj. R^2
Home Behavior									
TV-Sch. Avg.-Hrs/wk	-/-	14.7	2.85	-.64	(4.2)	.32	-2.1***	-3.5***	.272
Read for Fun Index	-/+	1.85	.28	.05	(2.8)	-.09**	.10***	.26***	.143
Use Computer for Sch Work	?/+	.40	.24	-.08	(6.3)	-.09***	-.11***	.150***	.180
Have Calculator	?/+	.88	.13	.05	(6.2)	.04***	.04***	.075***	.084
Watch Science programs on TV	-/+	.97	.38	.03	(1.5)	.04	.25***	.099***	.110
P. want me to do well-- Math	+/+	2.53	.22	.04	(3.4)	.05**	.02	.038**	.076
P. want me to do well-Science	+/+	1.67	.34	.06	(3.0)	.01	.10***	.172***	.063
Parent Talk about Math Class	+/?	.62	.17	.04	(3.8)	.08***	.03**	.030**	.034
P. Talk about Science Class	+/?	.47	.17	.06	(6.1)	.O6***	.00	.047***	.048
Teacher Behavior									
Total Homework—Hrs/wk	+/?	4.41	1.62	.76	(8.2)	1.52***	-.21	.02	.165
Math Homework—Hrs/wk	+/+	1.66	.64	.14	(3.6)	.16**	.008	.124**	.042
Science Homework—Hrs/wk	+/+	1.04	.47	.20	(7.0)	.19***	-.07*	.072*	.054
Emphasize Whole Number Oper	-/-	1.68	.49	-.13	(4.4)	-.12**	.09**	-.044	.029
Math Quiz Index	+/-	1.62	.52	.19	(7.6)	.53***	.70***	-.010	.368
Science Quiz Index	+/-	.89	.38	.12	(6.2)	.68***	.34***	-.056**	.331
Math Group Problem Solving	-/-	1.48	.62	-.09	(2.6)	.07	-.59***	-.124**	.131
Math Work Alone on Problems	+/+	3.22	.37	.03	(1.6)	-.05	-.13***	.052*	.060
Science Do Experiments Ind.	+/0	1.52	.63	.33	(9.6)	-.11*	.41***	-.050	.169
Science Watch Experiments	+/-	2.42	.47	.16	(6.0)	-.08	.24***	-.121***	.106
Science Watch Films Index	?/-	.94	.48	.02	(.8)	.40***	-.06	.006	.059
Math Listen to Teacher	?/-	3.28	.55	.00	(.1)	.18***	-.97***	-.110***	.587
Science Listen to Teacher	?/0	2.30	.48+	.04	(1.8)	.38***	-.60***	-.057*	.398
School Administrator Behavior									
Math Specialist Teachers	+/+	.48	.50	.27	(10.5)	.64***	.10**	.124***	.209
Science Specialist Teachers	+/+	.49	.50	.35	(13.7)	.57***	.00	.122***	.225
Took Math Courses in Univ	+/+	.66	.39	.15	(6.8)	.22***	-.05	.080***	.103
Took Science Courses in Univ	+/0	.69	.38	.20	(9.8)	.15***	-.17***	.046*	.173
Math Class Hours	+/0	3.97	.89	.42	(8.3)	.17*	.30***	-.253***	.114
Science Class Hours	+/+	3.00	.82	.24	(5.4)	.99***	-.O6	-.017	.174
Specialized Science Labs	+/+	1.95	.95	.63	(12.3)	.33***	.12	.065	.135
Tracking in 8th Gr. Math	+/+	.16	.37	.09	(4.7)	.63***	.07**	.042*	.174
Tracking in 8th Gr. Science	+/+	.08	.27	.03	(2.2)	.28***	.11***	.014	.077
w/Class Abil Group in Math	?/-	.23	.42	-.16	(6.7)	-.01	-.08**	-.007	.039
w/Class Abil Grp in Science	?/0	.10	.30	-.05	(2.9)	.03	.00	.003	.008
Propor. Exper. Teachers	+/+	.59	.24	-.03	(1.8)	-.00	.03	-.033*	.072
Propor. New Teachers	-/0	.16	.15	-.02	(1.6)	-.03*	.00	-.003	.049
Hours in School Year	+/0	949	89	13.2	(2.5)	34.0***	-10.6	4.90	.032
Class Size	-/+	24.8	6.1	1.9	(5.2)	2.1***	.1	.73	.073
Teacher Preparation Time	+/+	.31	.27	.06	(6.0)	.07***	-.02*	-.006	.083
Student Attitudes									
Math Important to get Job	+/+	2.56	.21	-.02	(1.4)	-.03	-.05**	.017	.055
Science Important to get Job	+/+	1.93	.33	-.07	(3.8)	-.13***	-.21***	.046*	.131

Source: Regressions predicting the characteristics of 1366 to 1460 Canadian and American secondary schools. Control variables included but not shown were religious school, independent school, share of students whose home language was different from the language of instruction and mean number of siblings. Mean school char. based on n gt 8.

Column 1 summarizes the hypotheses that were presented in the first section of the paper. To the left of the slash, /, is the expected sign (based on a priori reasoning and the literature) of the impact of EXAM on this measure of home or school behavior. A question mark appears here if no hypothesis was generated for this variable. The +, - and 0's appearing to the right of the slash mark summarize the analysis of IAEP data presented in Table 4. A + indicates that the variable had a significant positive effect on test scores at age 13. A - implies a significant negative effect. A zero, 0, indicates no significant relationship.

Effects on Home Behavior: The hypotheses about the behavior of parents are strongly supported. As predicted in **H**-3 and **H**-4, students in provinces with exams watch 40 minutes less television a week, and are 4 to 6 percentage points more likely to report that their parents want them to do well in the examination subject and are also more likely to report that their parents have talked to them about what they are learning in school.

Opponents of externally set curriculum-based examinations predict that they will cause students to cut back on learning activities that do not have a direct relationship to the exams. This hypothesis, **H**-5, was operationalized by testing whether exam systems are associated with <u>less</u> reading for pleasure and <u>less</u> watching of science programs like NOVA and Nature. Neither of these hypotheses is supported. Indeed point estimates of the effect of EXAM are positive (rather than the hypothesized negative) and the positive effect is statistically significant for reading for pleasure.

Effects on Teacher Behavior: It was hypothesized in **H**-6 that provincial exams would cause teachers to give more homework, to cover more difficult material, to schedule more quizzes and tests, to reduce the time that students spend doing group problem solving, increase the time that students work alone doing math problems and schedule more experiments in science class. All of these hypotheses are supported. Provincial exams are associated with students doing 45 additional minutes of homework per week and 10 to 12 additional minutes per week of homework in mathematics and in science. Emphasis on computation using whole numbers—a skill that should be learned by the end of 4th grade-declined significantly. More quizzes are assigned and more experiments are scheduled. Apparently, teachers subject to the pressure of a provincial exam 4 years in the future adopt strategies that are conventionally viewed as "best practice," not strategies designed to maximize scores on multiple choice tests.

But they also apparently see through the conventional wisdom and allocate less time to group problem solving activities. Also against conventional wisdom they may have also given students more in-class time to do problems on their own. This effect, however, is small and not significant.

Hypothesis (#7) that provincial exams would lead teachers to adopt strategies which cross section analysis indicates increase achievement is clearly rejected. Table 4 suggests quizzes and experiments lower IAEP test scores, but teachers in provinces with exams scheduled more quizzes and more experiments. This suggests that either (a) our cross section results do not reflect true causal relationships (because of the endogeneity problem already discussed), (b) teachers avoid strategies that maximize IAEP test scores because the provincial exams require different strategies or (c) that teachers are unaware of what strategies really maximize achievement and, instead, adopt practices which reflect a flawed conventional wisdom.

Effects on School Resources and Policies: Most of the hypotheses about how administrator behavior would be affected by provincial exams are supported. The most striking effect is the big increase in the use of specialist teachers and teachers with university courses in the subject. The percentage of specialist teachers is 27 points higher in mathematics and 35 points higher in science. In provinces with exam systems, the proportion of teachers who took courses in their subject at university is 15 percentage points higher for math teachers and 20 percentage points higher for science teachers. The specialized science labs index is also two-thirds of a standard deviation higher. Classroom instruction hours are 11 percent higher in mathematics, 8 percent higher in science and 1.4 percent higher overall. This suggests that the increase in class time devoted to math and science comes at the expense of something else. In addition, tracking appears to be more common particularly in mathematics and teachers get extra time to prepare their lessons.

The hypothesis that schools in exam provinces would try to hire more experienced teachers (in all subjects not just math and science) receives little support. They are slightly less likely to employ teachers with fewer than three years of tenure but they were also less likely to employ teachers with over 10 years of experience.

The other rejected hypothesis is the prediction that exam systems would stimulate reductions in class size. In fact, classrooms in provinces with exams have on average two additional pupils. Preliminary analyses of the determinants of test scores found a negative relationship between mean class size for the eighth grade and IAEP test scores. Could it be that principals "know" that large classes do not hurt student achievement on tests and that they obtain the resources necessary to hire more qualified teachers by increasing class size?

Hypothesis #9 that schools in exam system provinces would allocate resources in ways that maximize IAEP test scores also receives strong support. Of the ten administrator behavior variables that had significant positive effects on IAEP test scores, nine have significant positive relationships with EXAM.

I have specified no hypothesis for how within class ability grouping would be affected by external exams or how it would affect test scores. The cross section regressions uncovered a negative relationship with IAEP test scores and with EXAM systems. Here again the signs of these relationships are, as predicted by Hypothesis #9, the same.

Effects on Student Attitudes: Surprisingly, examination systems are not associated with a higher proportion of students thinking that mathematics or science are important for getting a job. Apparently, the availability of more reliable information on student performance in high school has not caused Canadian employers in Quebec (and presumably other provinces with examination systems) to ask applicants to provide information on high school grades. Job applications were obtained from seven large companies located in Montreal Quebec. All of them requested information about degrees and certificates of skills but none requested information on grades in secondary school. School transcripts were asked for in a few cases, but interviewers reported that this was to confirm graduation not to screen on grades in school.[17] Apparently, provincial exams increase rewards for studying by signaling performance to students, parents, colleges and universities not by signaling performance to Canadian employers.

How Does the USA Differ from Canada?

Column 6 of Table 5 presents estimates of the differences in student, teacher and administrator behavior between the USA and English speaking Canadian schools holding EXAM, school control and social class constant. American 13 year olds do less reading for fun, are more likely to have a calculator, but less likely to use a computer for school work. American parents are more likely to talk with their children about what they are learning in math and science classes. American 13 year olds in the IAEP study do 1.5 hours per week more homework than Canadian students. They take more quizzes, watch science films more and spend more time listening to teachers give lessons. Drill on simple arithmetic is less common.

American middle schools are much more likely to employ specialist mathematics and science teachers who have studied the subject in college. Science labs are better and an extra hour per week is typically devoted to teaching science. Tracking is much more common and the school year is nearly 3.6 percent longer. Class sizes are larger and teacher preparation time is greater as well. Relative to Canada, anyway, the adults in the system appear to be trying hard to do the things that both conventional wisdom and Table 4 regressions tell us raise test scores. Why then do American students score significantly below Canadian students on the IAEP? The answer is demography (fewer book in the home and extra siblings) and the absence of curriculum-based external exams at the end of high school (see table 3).

VII. Summary and Conclusions

Externally set curriculum-based examinations at the end of high school apparently have pervasive backwash effects on middle school students, their parents, teachers and school administrators. Holding the social class background of students constant, students from Canadian provinces with examination systems were substantially (23 percent of a standard deviation) better prepared in mathematics and 18 percent of a standard deviation better prepared in science than students from provinces lacking such exams. The effect of an exam system on mathematics achievement of 13 year olds is larger in a standard deviation metric than the decline in math SAT scores between 1969 and 1980 that has been such a focus of public concern. Other natural experiments yield similar findings. When adjustments are made for ethnicity, parent's education and family income of SAT test takers, New York State ranks higher on the SAT than any of the other 37 states where the SAT is taken by large numbers of students. The mathematics and science achievement of Swedish high school seniors declined in the years following the elimination of high/medium stakes curriculum-based exams.

The analysis also found that examination systems had pervasive effects on school administrators, teachers and parents. In the provinces with external exams, schools were more likely to:

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

At home students watch less TV, spend more time reading for fun, and are more likely to report their parents want them to do well in math and science. In addition, parents are more likely to talk to their child about what they are learning at school. None of the undesirable effects that opponents of external exams have predicted came about.

Lessons for Economic Analysis of Education Issues

Much of the economic research on elementary and secondary education has employed a production function paradigm in which test scores measuring academic achievement are the outputs, teachers are the labor input and students are goods in process. This research appears to suggest that schools are not efficiently run—money is spent in ways that do not appear to increase the presumed output, measured academic achievement.

This paper points in different directions. Students not teachers are the front line workforce of the education enterprise. School budgets pay only a fraction of the true opportunity costs of education. Schools are treated as multi-product organizations that adjust their product mix in response to effective demand. When schools gets additional budgetary resources, they distribute the money across a host of priorities many of which have little to do with academic achievement. Students are also consumers /investors who choose which goals (outputs) to focus on and how much effort to put into each goal. The behavior of each of the system's actors (teachers, administrators, school board, students and parents) depends on the incentives facing them. The incentives, in turn, depend upon the cost and reliability of the information (signals) that are generated about the various outputs of the system. Given their multiple constituencies, their multiple goals, the poor quality of the signals that are used to assess achievement of these goals, we need to rethink the conclusion that schools are inefficient. Maybe they are highly adaptive to a disfunctional incentive environment.

A wag recently described American secondary schools as ***“A place where young people come to watch adults work. "*** This captures in a nutshell the primary reason why American secondary school students lag behind their counterparts in Europe and East Asia. The economic theories of greatest relevance to this complex system are agency theory, game theory, signaling theory and consumer demand theory, not production theory. This paper only scratches the surface. Deeper plowing of these furrows will, I predict, yield a large crop of new insights.

Bibliography

Bishop, John. (1990) "What's Wrong with American High Schools? Can State Governments Fix It?" Center for Advanced Human Resource Studies Working Paper 90-17. The Economic Consequences of American Education, eds. Richard Aronson and Robert Thornton, New York: JAI Press, 1994.

Bishop, John. (1993) "Impacts of School Organization and Signalling on Incentives to Learn in France, the Netherlands, England, Scotland and the United States." Center for Advanced Human Resources Working Paper 93- to appear in Education, Training, and the Economy. ed. Christoph Buechtemann, New York: Russell Sage Foundation, to be published in Spring 1995.

Bradley, Ann. "Not Making the Grade: Teacher firing spurs debate over standards and expectations for students." Education Week, Sept. 13, 1993, p. 1, 19-21.

Chubb, John and Moe, Terry. Politics and Markets and America's Schools. Washington, D.C.: The Brookings Institution, 1990.

College Board. AP Yearbook 1988. New York: College Entrance Examination Board, 1988

Comber, L. C. and Keeves, John P. (1973) Science Education in Nineteen Countries. New York: John Wiley.

Cooper, Harris M. Homework. White Plains, New York: Longman, 1989.

Dynarski, Mark and Gleason, Philip. "Using Scholastic Aptitude Test Scores as Indicators of State Educational Performance." Economics of Education, Vol 12, No. 3, 203-211.

Elley, Warwick, How in the World do Students Read?, The Hague, The Netherlands: International Association for the Evaluation of Educational Achievement, 1992.

Graham, Amy and Husted, Thomas. "Understanding State Variation in SAT Scores." Economics of Education, Vol 12, No. 3, 197-202.

Husen, Torsten International Study of Achievement in Mathematics. New York: John Wiley & Sons, 1967.

International Assessment of Educational Progress, Learning Science. Princeton, NJ: Educ. Testing Service 1992a.

International Assessment of Educational Progress, Learning Mathematics. Princeton, NJ: ETS, 1992b.

International Assessment of Educational Progress, IAEP Technical Report. Volume 1, Princeton, NJ: ETS, 1992c.

Jencks, Christopher and Crouse, James. "Aptitude vs. Achievement: Should We Replace the SAT?" The Public Interest, 1982.

Kang, Suk. "A Formal Model of School Reward Systems." in Incentives, Learning and Employment, edited by John Bishop, Columbus Ohio: National Center for Research in Vocational Education, 1985.

Lapointe, Archie; Mead, Nancy and Phillips, A World of Differences. Educational Testing Service, Princeton New Jersey, 1989.

Lerner, Barbara. "Good News about American Education." Commentary, Vol. 91, No. 3, March 1990, 19-25.

Longitudinal Survey of American Youth. "Data File User's Manual" Dekalb: Public Opinion Laboratory, 1988.

McDonald, Judith. "The Canadian Educational System." The Economic Consequences of American Education, gds. Richard Aronson and Robert Thornton, New York: JAI Press, 1994.

McKnight, Curtis C. et al. The Underachieving Curriculum: Assessing US School Mathematics from an International Perspective. Stipes Publishing Co.: Champaign, IL, January 1987.

National Assessment of Educational Progress. The Science Report Card. Princeton, New Jersey: Educational Testing Service, 1988a.

National Assessment of Educational Progress. The Mathematics Report Card. Princeton, New Jersey: Educational Testing Service, 1988b.

National Center for Educational Statistics. The Digest of Education Statistics: 1992. Wash. D.C.: US Department of Education, 1992.

National Center for Educational Statistics. The Condition of Education: 1993. Vol. 1, Wash. D.C.: US Department of Education, 1993.

New York State United Teachers, "Blowing the Whistle on Education's 'Open Secret'." Albany, New York: New York State United Teachers, March 2, 1994.

Noah, Harold J. and Eckstein, Max A. "Tradeoffs in Examination Policies: An International Perspective." Paper presented at the Annual Meeting of the British Comparative and International Education Society, University of Bristol, September 15-17, 1988.

Postlethwaite, T. Neville and Wiley, David E. Science Achievement in Twenty-Three Countries. London: Pergammon Press, 1992.

Powell, Arthur; Farrar, Eleanor and Cohen, David. The Shopping Mall Hi School. New York, New York: Houghton Mifflin, 1985.

Raffe, D. "School Attainment and the Labor Market." Fourteen to Eighteen: The Changing Pattern of Schooling in Scotland, ed. D. Raffe, Aberdeen: Aberdeen University Press, 1984, 174-193.

Rehn, Gosta. (1980) "Sweden." Education and Youth Employment in Sweden and Denmark. Carnegie Council on Policy Studies in Higher Education.

Robitaille, David and Garden, Robert. The IEA Study of Mathematics II: Contexts and Outcomes of School Mathematics, New York: Pergamon Press. 1989.

Sizer, Theodore R. Horace's Compromise: The Dilemma of the American High School. Boston: Houghton Mifflin, 1984.

Slavin, Robert E. "When Does Cooperative Learning Increase Student Achievement?" Psychological Bulletin, 1985.

U.S. General Accounting Office, Educational Testing: The Canadian Experience with Standards, Examinations and Assessments. (written by Kathleen D. White) GAO/PEMD-93-11, Washington, D.C., April 1993, 1-74.

Endnotes

1. Curriculum-based exams are very different from the minimum competency exams adopted by many states. While minimum competency exams have apparently reduced the numbers of students with very low basic skills levels (Lerner 1991), the level of competence required to pass most of these exams is very low. They are typically first taken in 9th or 10th grade and most students pass on the first sitting. Transcripts indicate only whether the student eventually passes the test not achievement levels above the minimum. Thus, for the great majority of students, such exams have no further effect on incentives to study. Incentive effects are focused on the small minority who fail on the first try and must repeat the test. Unlike minimum competency exams, curriculum-based examinations are generally tied to taking a particular sequence of courses and are therefore effectively voluntary. They are much more demanding and how well the student does on the test (not just whether a minimum was exceeded) enters the student's record. They cover a broader range of subjects and the questions used are typically published after the test has been given. According to the National Assessment of Educational Progress 7-9 percent of 13 year olds score below the average 9 year old in reading and math and 15-17 percent score above the average 17 year old. When achievement differentials among students are as large as this, incentives for effort are stronger for most students if the full range of achievement is signaled rather than just whether the individual has passed some absolute standard. Under the single cutoff reward system, many students pass the standard without exertion and are, therefore, not stimulated to greater effort by the incentive. Some of the least well prepared students will judge the effort required to achieve the standard to be too great and the benefits too small to warrant the effort. They give up on the idea of meeting the standard. Only a few students will find the reward for exceeding the single absolute cutoff an incentive for greater effort (Kang 1985).

2 . Sweden not only ended its exit exams, it established new rules for selecting applicants for university. Swedish upper secondary students pursue specialized lines of study beginning in tenth grade. Universities had traditionally admitted recent graduates of academic lines of study in upper secondary school with three year durations. Reforms introduced in 1977 and earlier changed the rules governing competition for university places to favor those who had worked a few years after graduating from secondary school (Rehn 1980, p. 150). In addition, universities were not allowed to give preference to secondary school graduates who had pursued the more difficult longer duration academic lines of study in upper secondary school. If one-third of all applicants to university had pursued a two year vocational line of study, the universities and colleges were required to admit one third of their entering class from this group. This meant that a given student's chances of being admitted to popular majors in university were generally higher if the student pursued a less demanding shorter duration vocational line of study. As result in fall 1978, the toughest secondary school major, natural science, was also the easiest to get into. Ninety-three percent of those who selected it as a first choice were admitted. For those who wanted to enter a two year vocational line in secondary school, only 46 percent were admitted to their first choice program. These reforms also induced many secondary school graduates to postpone entry into higher education in order to accumulate enough work points to gain admission to preferred educational programs. Entrants to the study of medicine, for example, had an average age of 30 in 1977 and 1978 (Rehn 1980 p. 154). This change in university admissions policies certainly contributed to the decline in the number of Swedish upper secondary students taking rigorous courses in mathematics and science during the 1970s.

3 . The zero point on the scale for each year is the average performance for that year of the 14 year olds in England. England was selected for this standardization role because there was no change in the proportion of its age cohort who took the test. The standard deviation used as the metric measures individual variance for pooled data on 14 year olds from 11 (14 in 1983) industrialized countries. The vectors describe how an individual nation's scores and participation rates changed relative to England. The 1983 study reports separate means in for non-specialists (e.g. majors in economics, languages) and for students specializing in science. These two figures were averaged using the shares of the age cohort who pursue these two courses of study as weights (Postlethwaite and Wiley 1992 p. 6, 74).

4 . The decline in Sweden's relative performance in science appears to be an upper secondary school phenomenon. The relative performance of Swedish students at age 14 did not decline between 1970 and 1983/84. Of the eight countries that participated in both the First and Second International Studies of Science achievement of 14 year olds, four lost ground on Sweden and three gained ground. Japan and Hungary originally way ahead remained ahead, but their lead diminished. England and the United States were roughly equal to Sweden in 1971, but fell back in the succeeding decade. Finland and Italy were behind in 1971 but had essentially caught up by 1982. Dutch 14 year olds were way behind in 1971, but had surpassed Sweden by 1982. The lack of backwash effects at age 14 from the elimination of exit exams at age 18 is not difficult to explain. Admission to preferred specializations in upper secondary school has always been competitive. Which specialization is preferred has changed but the fact of competition has not. Certain lines of study are strictly rationed. For example, in 1978 only 22 percent of those giving nursing as their first choice specialization were admitted to the line. Admission to rationed lines is based on grades in lower secondary school (Rehn 1980).

5 . About 56 percent of 9th graders take the Mathematics Course 1 exam and, of these, 24 percent fail. Similar proportions of 10th and 11th graders take the global studies, biology and English exams. Failure rates were 20 percent in global studies, 18 percent in biology and 13 percent in English. The great bulk of those not taking Regents exams are in courses that are considerably less challenging than Regents level courses. The fact that nearly half of New York students are avoiding Regents courses because they perceive them to be too much work or too difficult suggests that the standard of the exam is about as high as is feasible considering current average achievement levels in the state.

6 . The Advanced Placement (AP) examinations are an exception to the generalization that the U.S. lacks a national system of curriculum-based examinations. Students who take these courses and pass the examinations may receive college credit for high school work. While it is growing rapidly, AP is still a very small program. In 1988 only 8,022 of the 22,902 U.S. high schools offered any AP courses. Only 52 AP exams were taken on average in each participating high school (The College Board 1988). Of the 11th and 12th graders in 1993, only 2.8 percent took an AP English exam, 2.3 percent took an AP history exam, 1.7 percent took the AP calculus exam, and 1.7 percent took an AP science exam (NCES 1993). The theory proposed herein argues that AP students learn more not just because they are a self selected group of highly able students but because the external examination aligns incentives in a way that induces both teachers and students to give higher priority to learning.

7. Dynarski and Gleason (1993) have also predicted state mean SAT scores while controlling school resource variables and characteristics of the state's population obtained from the Census. Graham and Husted's analysis is preferable for our purposes because it uses data on the background of the students who took the test and no effort was made to control school resource variables that might be influenced by the existence of Regents exams.

8 . In the provinces that lack provincial exams, some local school districts have district level exams for core subjects, but most have not. In any case, one would not expect local district subject exams to have as powerful incentive effects as provincial or national exams. Manitoba introduced its 12th grade examination in the winter of 1991 about the time the IAEP exam was being administered to 8th graders in the province. The new examination system was announced in June 1990 only 7 months earlier. The teaching of 12th graders was probably affected during the 1990/91 school year, but it is unlikely that any incentive effects had percolated back to the 8th grade in so short a time. The system introduced in Manitoba in 1991 rotated the subject assessed on a five or six year cycle. Starting in 1996, Manitoba will assess math and language arts every year and require the exams to count for 30 percent of the student's final grade. Saskatchewan is planning to introduce a provincial examination system and New Brunswick will soon expand their system to

include physics, chemistry and biology. Alberta's examination system was instituted in 1984, so it was 7 years old when the IAEP data was collected.

9 . In 1950 local shares of school funding were about 58 percent in both countries. Since then, the local share has fallen more rapidly in Canada than in the U.S. Funding levels vary less within Canadian provinces than within American states. The average within province coefficient of variation is .09 for Canada and .17 for the United States (McDonald 1993; National Center for Education Statistics 1992)). In some provinces negotiations over teacher salaries occur at the provincial level.

10 . The gap in science achievement appears to have narrowed over the 1980s.

11. The same sequence of models was estimated without the dummy variable for Francophone school. This change in specification does not appreciably change findings regarding the impact of EXAM systems.

12 . The school resource and school policy variables are based on a questionnaire filled out by the principal. A number of the variables constructed from this questionnaire appear to contain considerable measurement error. For example, student-teacher ratios calculated by dividing total enrollment by the size of the teaching staff range from 1.7 to 520 and have a standard deviation of 23. Estimates of the time during the school day that teachers have for preparing their lessons derived from class size and student-teacher ratios range from -15.8 to .93. Reported classroom hours of instruction in math and science also appear to contain measurement error. For the United States alone, the range is from 2.08 to 6.67 hrs/wk in mathematics and from 1.83 to 8.75 hrs/wk in science. One of the Canadian schools in the IAEP reportedly allocated only one hour a week to teaching science. The measurement error problems mean that coefficients on these variables are probably biased toward zero. It is also quite possible that school policies are in some cases jointly determined with average performance levels. Consequently one should view these coefficients as only suggestive of causal relationships, not structural estimates of a well specified causal model.

13. Harris Cooper's (1989) meta-analysis of randomized experimental studies found that students assigned homework scored about one-half of a standard deviation higher on post tests than students not receiving homework assignments. The impact of homework on the rate at which middle school students learn was also significant, though somewhat smaller. It is difficult to estimate the effect of time spent doing homework on learning in data sets where homework is not manipulated experimentally. Spending a lot of time doing homework might be due to a strong work ethic, strict parental supervision or a particularly demanding teacher. But it also might result from having difficulty doing the problems. Consequently, at the individual level, homework time will tend to be negatively correlated with unobservable ability, so coefficients will be biased. To minimize this problem, the specification uses school mean hours of homework rather than the time individuals spend on their homework to capture the effect of homework. Variations across schools in amount of time that is spent doing homework in particular subjects probably depends primarily on how demanding teachers choose to be and school policies regarding homework not unobservable student ability.

14. Experimental studies have shown that cooperative learning—organizing classes into small teams that work together and compete for nominal prizes—significantly improves student learning. In the cooperative learning programs designed by Robert Slavin students are grouped by the teacher into evenly matched teams of 4 or 5 members that are heterogeneous in ability. After the teacher presents new material, the team works together on work sheets to prepare each other for periodic quizzes. The team's score is an average of the scores of team members, and high team scores are recognized at the end of the week in a class newsletter or through group certificates of achievement. In some cases individual gain scores are the outcome measure that is averaged to generate the team score. Even though rewards are for group outcomes, it is individual performance that is assessed. On the quizzes, one student cannot do the work for the others. Class grades continue to depend on individual performance on quizzes, homework and tests.

This prevents the free rider problem from arising. Slavin's (1985) research implies that the key ingredients for successful cooperative learning are:

-- A cooperative incentive structure—award based on group performance—seems to be essential for students working in groups to learn better. This creates an incentive for students to help and encourage each other to learn.

-- A system of individual accountability in which everyone's maximum effort must be essential to the group's success and the effort and performance of each group member must be clearly visible to his or her group mates.

What seems to happen in successful cooperative learning is that the team develops norms supportive of hard study. Since the group is small and the interaction intense, the effort and success of each team member is known to other teammates. Such knowledge allows the group to reward each team member for his or her contribution to the team goal. When correctly structured, cooperative learning adds peer rewards for learning on top of the normal report card rewards for learning. The types of cooperative learning that experimental research has shown work well are very different from what usually goes under the rubric of group problem solving.

15. Other weighting schemes were tried including using the school means for these variables. This did not change the pattern of results discussed below.

16 . Some of the schools selected to participate in the IAEP had considerably fewer than 30 age eligible students. In developing the IAEP sampling frame, schools predicted to have only a few age eligible students were generally combined into larger super schools for purposes of drawing the sample. When one of these schools was selected, the target sample of 30-34 students was distributed among the schools forming the super school (IAEP 1992c). Principal questionnaires were completed in each school, but sometimes the number of student interviews was too small to provide reliable estimates of school means. When models are estimated including the very small schools, the estimated impacts of EXAM are slightly larger than the results shown in Table 5.

17. These practices are not a consequence of legal prohibitions on requesting and using such information. A government approved official Canadian Manpower form obtained from the University of Montreal's College Placement office requests such information.