SWEDISH ECONOMIC POLICY REVIEW 6 (1999) 349-398

Are national exit examinations important for educational efficiency?

John H. Bishop*

Summary

■ This paper analyses effects of national or provincial exit examinations on educational quality. On theoretical grounds, the paper argues that such examinations should increase high school student achievement, particularly in examination subjects, and that teachers and students and parents and school administrators should focus more on academic achievement when making school-quality decisions. On the negative side, exit examinations may lead to a tendency to concentrate on learning facts, rather than understanding contexts.

The hypotheses are tested using several datasets, including countries and provinces with and without external exit examinations. The conclusion from the empirical tests is that positive effects are likely to dominate. Students in countries with these exams tend to outperform students in other countries in science, math, reading, and geography, when national economic development levels are accounted for.

The paper also argues that the elimination of the Swedish exit examination system in the 1970s, in combination with changes in the way university applicants were selected, appears to have led to a decline in the number of upper secondary school students taking rigorous courses in mathematics and science.

The paper also analyzes the mechanisms behind the positive relation between external exit examinations and student achievements. It is found that exit examinations are associated with higher requirements for entry into the teaching profession, higher teacher salaries, higher shares of specialized teachers, and many teaching hours in examination subjects. But external exit examinations were not found to be associated with higher teacher-student ratios, more spending on education, or higher teacher satisfaction. ■

* *Associate professor, Consortium for Policy Research in Education and Center for Advanced Human Resource Studies New York State School of Industrial and Labor Relations, Cornell University*

SWEDISH ECONOMIC POLICY REVIEW 6 (1999) 349-398

Are national exit examinations important for educational efficiency?

John H. Bishop*

In many countries, national or provincial exit examinations certify and signal achievements of secondary school students to universities and employers. These examinations are thought to have significant effects on how teachers teach and how students study, so the character of these examinations has been a source of controversy in many countries. Efforts to reform secondary education almost always involve changes in examination systems. The English merged the old Certificate of Secondary Education (CSE) and the O level exams into the General Certificate of Secondary Education (GCSE). France broadened the list of Baccalaureate examinations to include many vocational specialities and has set a goal of more than 80% of the age cohort participating by 2000. The Brevet exam at the end of lower secondary school, which had been abolished in 1977, was re-introduced in 1986. "The reasons were that the results had been declining in the experience of many people ... (Kreeft, 1990, p. 6)." The Canadian provinces of Manitoba and New Brunswick are re-establishing curriculum-based exit examinations that had been discontinued in the early 1970s.

Curriculum-based external exit exams (CBEEEs) are also being considered in the US. The Competitiveness Policy Council, for example, advocates that

> external assessments be given to individual students at the secondary level and that the results should be a major but not exclusive factor qualifying for college and better jobs at better wages (1993, p. 30).

* *The Center for Advanced Human Resource Studies and the Consortium for Policy Research in Education (funded by the Office of Educational Research and Improvement, US Department of Education) supported the preparation of this paper.*

The American Federation of Teachers advocates a system in which:

> Students are periodically tested on whether they're reaching the standards, and if they are not, the system responds with appropriate assistance and intervention. Until they meet the standards, they won't be able to graduate from high school or enter college (AFT, 1995, p. 1-2).

These two quotes represent the views of many educational reformers in the countries that do not currently have a system of diploma examinations. These reformers argue that "curriculum-based external exit exam systems" (CBEEESs), based on world-class content standards, will improve teaching and learning of core subjects. Is this claim justified? This paper analyses data from four large-scale international studies of student achievement and attempts to answer this question.

1. Theory: why curriculum-based external exit exams change incentives

What is a CBEEES? Critics of moves to establish or re-establish exit examination systems point out that students already take lots of teacher-made tests. American students also take many nationally standardised tests. So the critics ask "Why should a CBEEES significantly improve incentives and learning?" The response of CBEEES advocates is that CBEEESs have uniquely powerful incentive effects because they have these six characteristics; they:

1. *Produce signals of student accomplishment that have real consequences for students.*
2. *Define achievement relative to an external standard, not relative to other students in the classroom or the school.* Fair comparisons of achievement across schools and across students at different schools are now possible. Costrell's (1994, 1997) formal analysis of the optimal setting of educational standards concluded that more centralised standard setting (state or national achievement exams) generally results in higher standards, higher achievement, and higher social welfare than decentralised standard setting (i.e., teacher grading or schools' graduation requirements).
3. *Are organised by discipline and keyed to the content of specific course sequences.* This focuses responsibility for preparing the student for particular exams on a small group of teachers.
4. *Signal multiple levels of achievement in the subject.* If only a pass-fail signal is generated by an exam, the standard will have to be set low enough to

allow almost everyone to pass and this will not stimulate the great bulk of students to greater effort (Kang, 1985; Costrell, 1994). By age 13, students differ dramatically in their achievement levels. On the *National Assessment of Educational Progress*, 7–9% of 13 year olds are four or more grade-level equivalents behind their age mates and 15–17% are four or more grade-level equivalents ahead. 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 signalled rather than just whether the individual has passed some absolute standard. When a test generates only a pass-fail signal, many students pass without exertion and are thus not stimulated to greater effort by the reward for passing. 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. Few students will find the reward, for exceeding a single absolute cut-off, an incentive for greater effort (Kang, 1985). Costrell agrees:

> The case for perfect information (making scores on external examinations available rather than just whether the individual passed or failed) would appear to be strong, if not airtight: for most plausible degrees of heterogeneity, egalitarianism, and pooling under decentralisation, perfect information not only raises GDP, but also social welfare (1994a or b, p. 970).

5. *Cover almost all secondary school students.* Exams for a set of elite schools, advanced courses, or college applicants will influence standards at the top of the vertical curriculum but will probably have limited effects on the rest of the students. The school system as a whole must be made to accept responsibility for how students do on the exams. A single exam taken by all is not essential. Many nations allow students to choose the subjects to be examined in and offer high- and intermediate-level exams in the same subject.
6. *Assess a major portion of what students studying a subject are expected to know or be able to do.* But it is not essential that the external exam assess every instructional objective. Teachers can be given responsibility for evaluating dimensions of performance that cannot be reliably assessed by external means.

1.1. Why and how are CBEEESs hypothesised to increase achievement?

National or provincial curriculum-based external exit examinations systems (CBEEESs) improve the signals of achievement available to colleges and employers, and this is likely to induce them to give academic achievement greater weight when they make admission and hiring decisions. Rewards for study and learning should grow and become more visible.

Effects on students. Rewards are necessary because learning is not a passive act; it requires the time and active engagement of the learner. Students have many other uses for their time and attention, so learning is costly for them. The intensity of their investment in learning depends on a comparison of benefits (intrinsic and extrinsic rewards for learning) to costs. A rise in the benefits of learning increases student effort and learning.

CBEEESs should also shift attention toward measures of absolute achievement and away from measures of relative achievement, such as rank in class and teacher grades. Advocates of CBEEESs hope that CBEEESs will reduce peer pressure against studying. Interviews I conducted during 1996 and 1997 with middle-school students in Ithaca, New York (a small city dominated by two universities) indicate that most boys internalise a norm against "sucking up" to the teacher. How does a student avoid being thought a *suck up*? He:

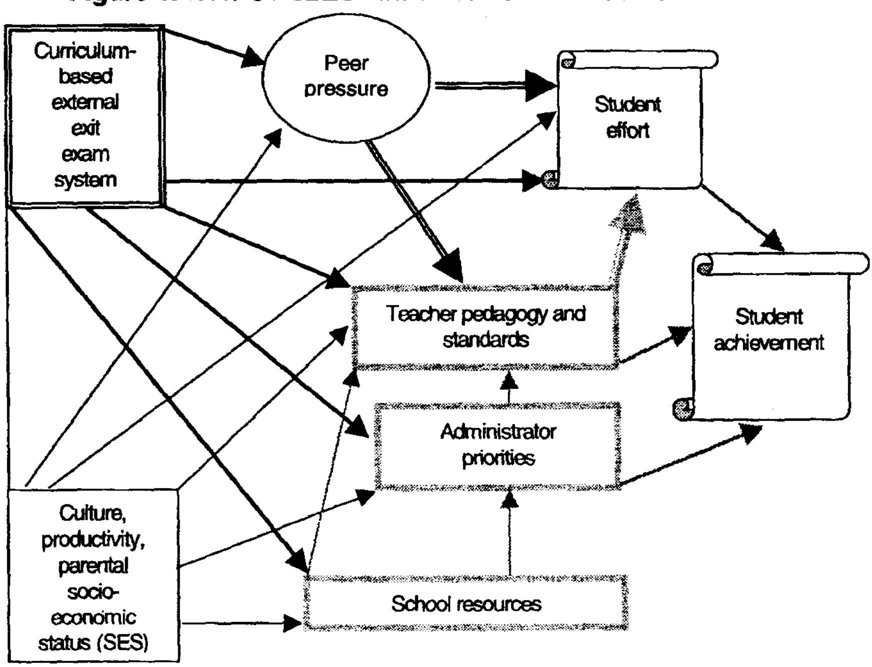
- Avoids giving the teacher eye contact
- Does not raise his hand in class too frequently; and
- Talks or passes notes to friends during class (this signals that you value friends more than rapport with your teacher).

Steinberg, Brown and Dornbush conclude similarly that "The adolescent peer culture in America demeans academic success and scorns students who try to do well in school (1996, p.19)." My conversations with Swedish students sometimes generate similar anecdotes.

Why are the studious called *suck ups*, *dorks*, and *nerds*? In part, it may be because grading exams on a curve means that study effort by one student tends to make it more difficult for others to get top grades. When exams are graded on a curve or college admissions are based on rank in class, the joint welfare of students is maximised if no one puts in extra effort. In the repeated game that results, side payments—friendship and respect—and punishments—ridicule, harassment, and ostracism—enforce the co-operative "don't study" solution. If, by contrast, students are evaluated relative to an outside standard, they no longer have a personal interest in getting teachers off track or persuading each other to refrain from studying. Peers should, in theory, become less tolerant of students who joke around in class or try to get the teacher off track.

Comparisons of the benefits and costs of focusing school resources and policies on academic achievement also influence parents, school administrators, and teachers. When a CBEEES is in place, exam results displace social class as the primary determinant of school reputations and this in turn should induce school staff to give enhanced learning higher priority. Teachers will upgrade curricula and assign more homework, and parents will demand better science labs and more rigorous teaching. School administrators will be pressured to increase the time devoted to examination subjects and hire more qualified teachers. The next section elaborates on the theory briefly described above.[1] Figure 1 illustrates the many paths by which CBEEESs are hypothesised to influence student achievement.

Figure 1. How CBEEESs influence student achievement



[1] Mathematical presentations of the theory are available from the author upon request.

2. Testable hypotheses:
Impacts of curriculum-based external exit exams

Students. The previously described theory predicts that:

Hypothesis	*Curriculum-based external exit examinations will result in ...*
1	Higher achievement. The effects should be strongest the year of the external examination. But they should reach down to 7th and 8th grade though maybe not down to early years of primary school.
2	Higher achievement, even when student characteristics, school resources, curriculum, teacher qualifications, and teaching techniques are held constant.

Parents. Curriculum-based external exams are also hypothesised to change incentives faced by parents and as a result, parents will put greater effort into trying to induce their children to study regularly.

Hypothesis	*Curriculum-based external exit examinations will ...*
3	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.

Opponents of external exams argue that focusing student attention on extrinsic rewards for learning will weaken student's intrinsic motivation to learn. George Madaus' list of possible negative effects includes "test scores come to be regarded by parents and students as the main, if not the sole, objective of education" and the result is

> undue attention to material that is covered in the examinations, thereby excluding from teaching and learning many worthwhile educational objectives and experiences (1991, p. 7).

If they are right, students in systems with external exams should be less likely to read for pleasure or to watch science programs such as *NOVA* and *Nature*.

Hypothesis	*Students will ...*
4	Spend less time watching science documentaries on TV and less time reading for fun.

School administrators. Local school administrators make hundreds of decisions that influence academic expectations and program quality (e.g., homework guidelines and whether to retain a popular but not very effective teacher). In many countries, schools are expected to achieve a host of often conflicting objectives: fostering self-esteem, providing counselling and supervising extra-curricular activities, musical training, health services, community entertainment (e.g., interscholastic sports). These other goals require additional staff and different kinds of staff, and so the goals may not be achieved by hiring teachers with strong backgrounds in calculus or chemistry.

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 grade-point averages (GPAs), complaining parents, a greater risk of being denied a diploma—are all negative.

When student learning is not assessed externally, the positive effects of choosing academic rigor are negligible and postponed. If college admission decisions are based on rank in class, GPA, and aptitude tests—and not on externally assessed achievement in secondary school courses—then 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 current graduates are more successful in local colleges. That, however, is even more uncertain and postponed. Publishing data on proportions of students meeting targets on standardised tests probably speeds the process by which real improvements in a school's performance influence its local reputation. But other indicators such as SAT test scores, proportions going to various types of colleges, and the socioeconomic background of the students tend to be more prominent. 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.

Few American employers pay attention to achievement in high school or school reputations when making hiring selections (Bishop, 1989, 1993; Hollenbeck and Smith, 1984). Consequently, students who study hard are not immediately rewarded with higher wage rates. Their greater competence is not fully recognised with higher wage rates until more than a decade after they graduate. Thus, higher standards benefit students as a group only after many years, so parents as a group have a reduced incentive to lobby for higher teacher salaries, higher standards, and higher school taxes.

External exams in secondary school subjects change the signalling environment. Hiring better teachers and improving the school's science laboratories now yields a visible payoff—more students passing the external exams and being admitted to top colleges. 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.

Hypothesis ***External exams will ...***

5 Cause priorities to shift in favour of achievement in examination subjects and away from inter-scholastic sports, band, and other activities intended to make school fun and entertain the public. Administrators and school boards will be induced to:

A • Improve the school's science laboratories (if science is an examination subject) and other facilities that contribute to learning in examination subjects

B • Offer additional courses in examination subjects and scale back offerings outside the core academic program

C • Increase the share of the school week devoted to examination subjects (when this is a local decision)

D • Lengthen the school day and school year (when this is a local decision)

E • Offer accelerated/enriched math and science courses

F • Use specialist teachers to teach examination subjects

G • Hire teachers with a thorough background in the field

H • Reduce class size in examination subjects

I • Give teachers additional preparation time

J • Pay higher salaries

K • Spend more per pupil.

Where students and parents choose their secondary school and state subsidies follow the student, the incentive effects of CBEEESs are magnified. In countries that have school choice and a CBEEES, newspapers typically publish league tables that report examination results by school. These results have major effects on enrolment applications the following year. Marginal instructional costs are typically below state aid per student, so schools at the top of the league table often expand (sometimes by bringing in temporary classrooms), forcing the schools with poor results to shrink and lay off staff.

Hypothesis *External exams will ...*

6 Induce larger shifts in the priority given academics when parents are able to choose which school their child attends and funding follows the student.

Teachers. In the US, 30% of the teachers say they "feel pressure to give higher grades than students' work deserves" and "feel pressure to reduce the difficulty and amount of work you assign" (Peter D. Hart Research Associates, 1994). Under a system of external exams, teachers and local school administrators lose the option of lowering standards to lower failure rates and raise self-esteem. Their response will be to strive to prepare their students for the external exam.

Hypothesis *External exams will ...*

7 Induce teachers to:

- A • Set higher standards
- B • Assign more homework
- C • Increase the number of experiments that students do in science class
- D • Have students solve mathematics problems alone rather than in groups
- E • Give more quizzes and tests
- F • Increase their use of other teaching strategies, which they believe improve exam performance
- G • Try less hard to entertain students
- H • Pay less attention to non-academic goals such as self-esteem, good discipline and low absenteeism.

Some educators argue that external exams can have negative effects on teaching. It is argued, for example, that "preparation for high stakes tests often emphasises rote memorisation and cramming of students and drill and practice teaching methods" and that

> some kinds of teaching to the test permits students to do well in examinations without recourse to higher levels of cognitive activity (Madaus, 1991, p. 7-8).

The assumption of opponents appears to be that the tests developed by individual teachers for use in their class are better than examinations developed by the committees of teachers that would have responsibility for developing state or national examinations. To the contrary, the tests that teachers presently develop for themselves are generally of very low quality. The Fleming and Chambers (1983) study of tests developed by high school teachers using Bloom's taxonomy of instructional objectives found that

> over all grades, 80 per cent of the items on teachers' tests were constructed to tap the lowest of the taxonomic categories, knowledge (of terms, facts or principles) (Thomas, 1991, p. 14).

Rowher and Thomas (1987) found that in colleges fully 99% of items on instructor-developed tests in American history required the integration of ideas, while only 18% of junior high school and 14% of senior high school test items required such integration. Secondary school teachers test low-level competencies because that is what they teach. Few students take state-mandated tests in history, so poor history teaching cannot be blamed on standardised tests. More evidence is needed on this issue, so tests are conducted on this hypothesis:

Hypothesis	*External exams will ...*
8	Cause teachers to focus on teaching facts and definitions, not the scientific process. Students will conduct fewer experiments in science class and computation will be stressed in mathematics.

CBEEE advocates argue to the contrary that well-designed external examinations that are graded by teachers will improve instruction. In May 1996, I interviewed several activists in the Alberta Teachers Union about the examination system in Alberta Canada. Even though the union and these teachers opposed the exams, they universally reported that serving on grading committees was "...a wonderful professional development activity (Bob, 1996)." Having to agree on what constituted excellent, good, poor, and failing responses to essay questions or open-ended math problems resulted in a sharing of perspectives and teaching tips that most found very helpful.

3. Do CBEEESs increase achievement? A look at the evidence

The hypothesis that CBEEESs improve achievement is tested by comparing nations and provinces that do and do not have such systems. Four different data sets are examined:

1. Science and mathematics achievement of 13 year olds in the 40-nation *Third International Math and Science Study*
2. The reading literacy of 14 year olds in the *International Association of the Evaluation of Educational Achievement's (IEA) Reading Study*
3. Science, math and geography scores of 13 year olds on the *International Assessment of Educational Progress* (IAEP) for 16 nations
4. Science and math scores of 13 year olds in nine Canadian provinces.

The theory predicts that CBEEESs influence societal decisions about education spending, administrator decisions about school priorities, teacher's decisions about standards and pedagogy, and student decisions about studying. Much of the ultimate impact of CBEEESs on student achievement derives from the changes they induce in spending, priorities, and pedagogy. In this section, the objective is to assess the *total* effect of CBEEESs on achievement (the sum of all the paths leading from the CBEEES to student achievement shown in Figure 1). Estimates of the total effects of CBEEESs are obtained from a reduced form model that controls for parental socio-economic status, national productivity levels, and national culture—and not the endogenous administrator, teacher and parent behaviours.

Section 4 presents models of the paths leading out of the CBEEES box in Figure 1. The relationships between CBEEFS and the resources devoted to K-12 schooling, administrative policies and priorities, and teacher pedagogy and standards are studied in a 42-nation cross-section and in comparisons of Canadian provinces with and without CEEEFS.

3.1. Third International Mathematics and Science Study

The *Third International Mathematics and Science Study* (TIMSS) provides 1994-95 data for 7th and 8th graders for 40 countries. The 1990-91 IEA reading study provides data on the literacy of 9th graders in 24 countries. To determine which TIMSS nations have CBEES, I reviewed comparative education studies, government documents, and education encyclopaedias and interviewed education ministry officials, embassy personnel, and Cornell graduate students from the country.[2] Twenty-two national school systems were classified as having CBEEESs for both subjects in all parts of the country: Austria, Bulgaria, Columbia, Czech Republic, Denmark, England, Hong Kong, Hungary, Ireland, Iran, Israel, Japan, Korea, Lithuania, the Netherlands, New Zealand, Russia, Scotland, Singapore, Slovak Republic, Slovenia and Thailand. France, Iceland, and Romania had CBEEESs in math but not in science. Australia, Canada, Germany, Switzerland, and the US had CBEEESs in some provinces/states but not in others. Norway has regular exit examinations in math, but examines in science only every few years. Latvia had an external examination system until very recently, so it was given a .5 on the CBEEES variable. The countries classified as not having a CBEEES in either subject were Belgium (Flemish and French systems), Cyprus, Greece, Philippines, Portugal, Spain, and Sweden. Following Madeus and Kelleghan (1991), the university entrance examinations in Greece, Portugal, Spain, Cyprus and the ACT and SAT in the US were not considered to be CBEEESs. University entrance exams should have much smaller incentive effects because students headed for jobs do not take them, and teachers can avoid responsibility for their students' exam results by arguing that not everyone is college material or that examiners have set an unreasonably high standard to limit enrolment in higher education.

Sweden was coded as a zero on the CBEEES variable because it eliminated its high-stakes, curriculum-based, external examinations at the end of secondary school in 1972 (Eckstein and Noah, 1993). Continuous

[2] Appendix A of Bishop (1998) provides a bibliography of documents and individuals consulted when making these classifications. The TIMSS report's information about examination systems does not distinguish between university admissions exams and curriculum-based exit exams, so its classifications are not useful for this exercise. In the TIMSS report, the Philippines, for example, is classified as having external exams, but its exams are university admissions exams similar to the SAT. South Africa was excluded because its education system was disrupted for many years by boycotts that were part of the campaign to end apartheid. Kuwait was excluded due to disruption of its education system by the Gulf War.

assessment by teachers became the basis for certifying achievement. But fearing grade inflation, the National Board of Education developed the multiple choice and short answer norm reference (*Centrala Prov)* subject examinations

> to help teachers grade students properly ... Teachers must use them and are not allowed to deviate more than .2 grade points from the standardised test class means in their final evaluation (Kreeft 1990, p. 15).

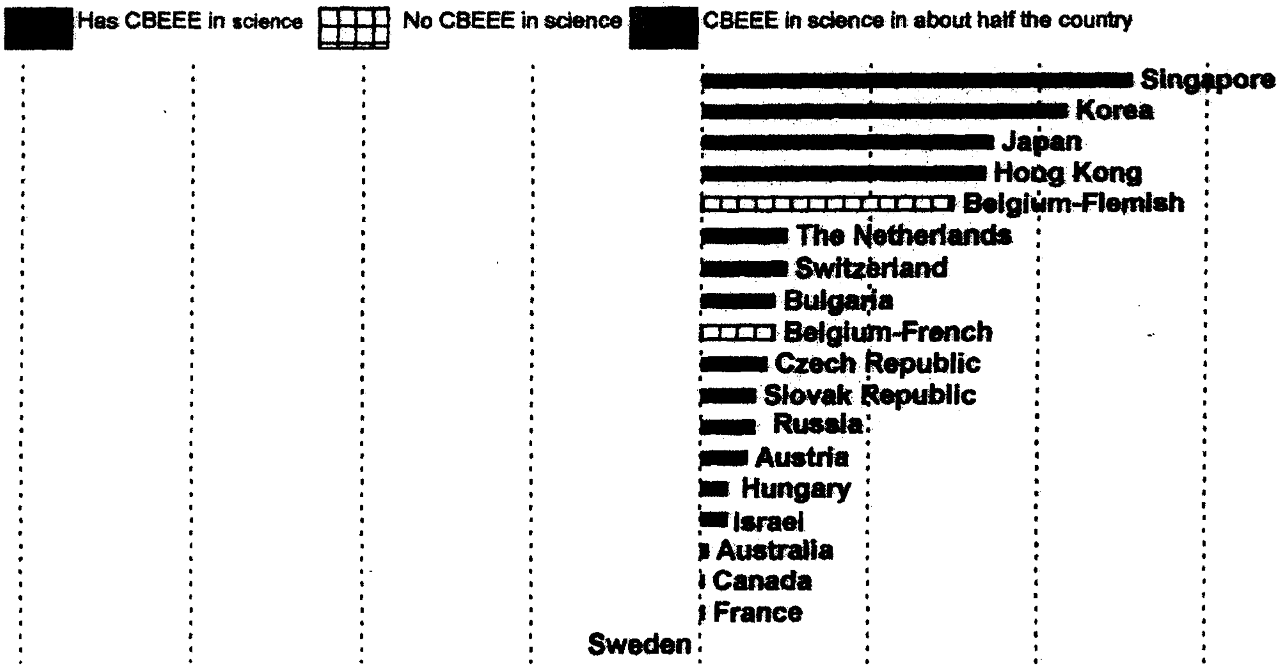
While teacher-awarded grades were supposed to follow a normal distribution centred on the class mean on the *Centrala Prov*, there is controversy about whether these mandates are being followed. In any case, the effect of students' performance on the *Centrala Prov* test on their grades is small. Swedish students I talked to did not perceive the *Centrala Prov* exams as carrying high stakes for themselves.

Figures 2 and 3 array the 40 TIMSS countries by the science and math achievement of their 13 year olds. Sweden ranks 14 in science and 19 in math. The gaps between the vertical grid lines represent one Swedish grade-level equivalent (GLE)—one-half the difference between 6th and 8th grade TIMSS test score means for Sweden. There are substantial achievement differentials between nations. In science, only Singapore is more than 1 GLE ahead of Sweden. Columbia, the Philippines, Lithuania, Romania, and Portugal are more than 2 GLEs behind. In math, Singapore, Korea, Japan, Hong Kong, and Flemish-speaking Belgium are more than 1 GLE ahead of Sweden. Columbia, the Philippines, and Iran are more than 2 GLEs behind. The countries represented by a solid black bar in the figures have a CBEEE in the subject. Countries represented by white squares do not. Note that the countries with a CBEEFS in the subject tend to have higher TIMSS scores.

Regression analysis. The mean 8th grade science and math test scores were regressed on average per capita gross domestic product for 1987–1991 deflated by a purchasing power parity price index, a dummy for East Asian nation and a dummy for CBEEES.[3] Results in Table 1 indicate that test scores are significantly higher in more developed nations, East Asian nations, and in nations with a CBEEES in the subject.

[3] TIMSS studied the two grades with the largest number of 13 year olds. The grade in the regression was the upper grade of the two studied. In Sweden, Norway, and Denmark, 7th grade was used because children start school at a later age. In England, Scotland, and New Zealand the 9th grade was used because children start grade 1 a year earlier than in most nations.

Figure 2. Math achievement at age 13.



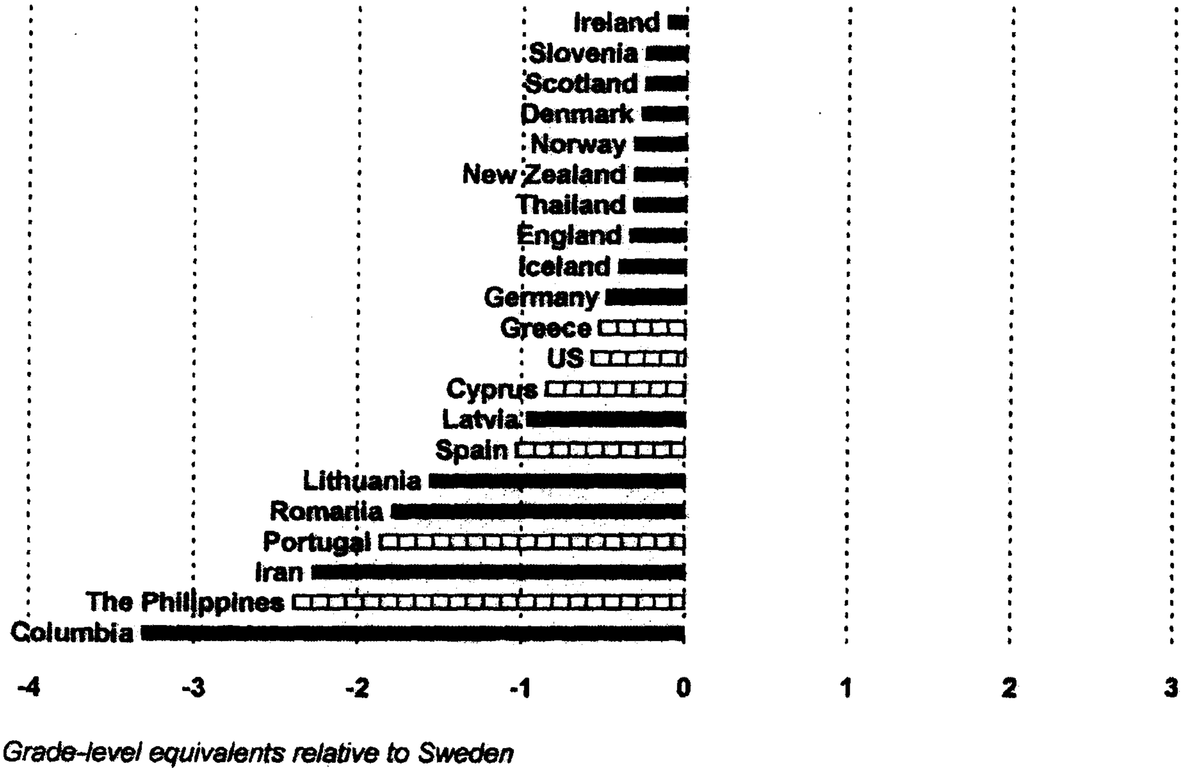
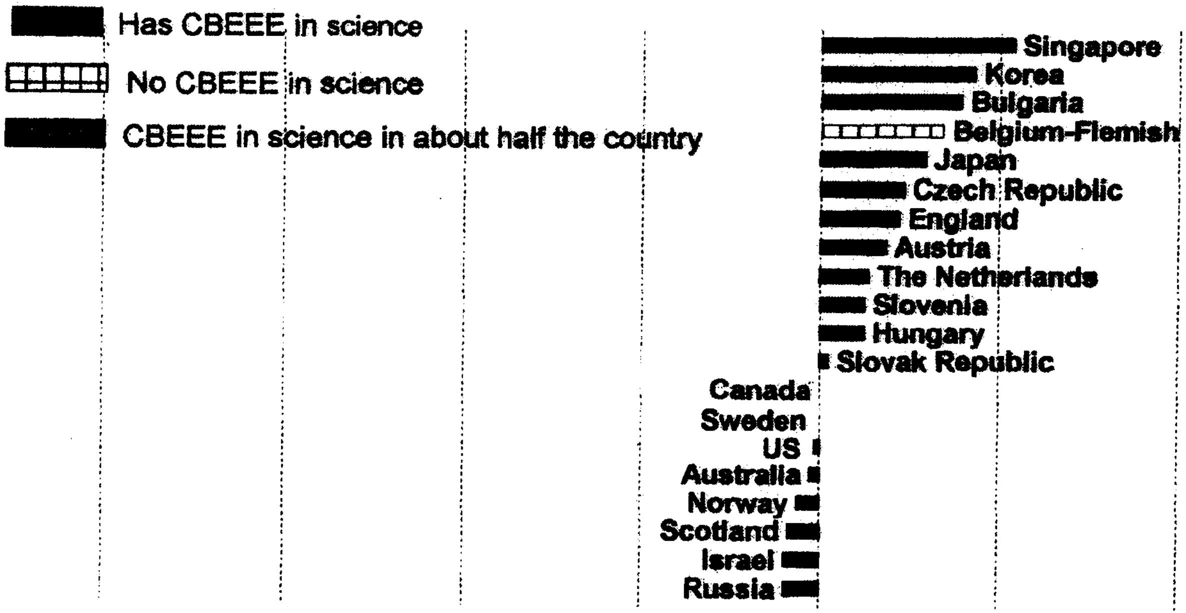

Ireland
Slovenia
Scotland
Denmark
Norway
New Zealand
Thailand
England
Iceland
Germany
Greece
US
Cyprus
Latvia
Spain
Lithuania
Romania
Portugal
Iran
The Philippines
Columbia
-4
-3
-2
-1
0
1
2
3
Grade-level equivalents relative to Sweden

Figure 3. Science achievement at age 13.



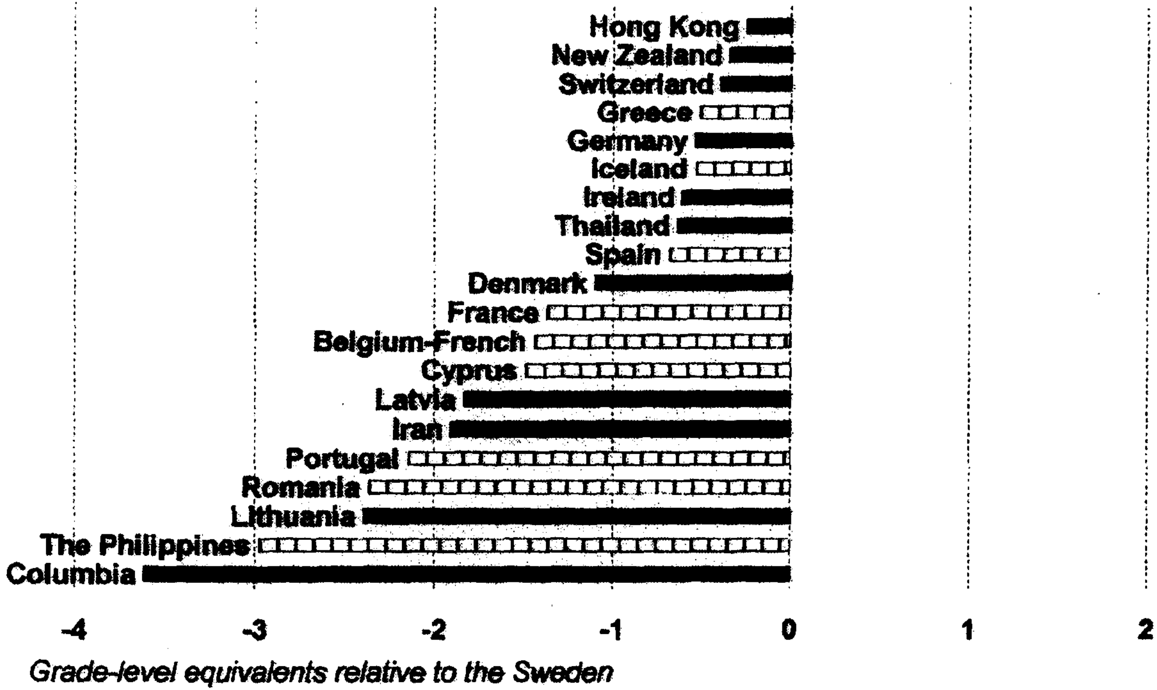

Hong Kong
New Zealand
Switzerland
Greece
Germany
Iceland
Ireland
Thailand
Spain
Denmark
France
Belgium-French
Cyprus
Latvia
Iran
Portugal
Romania
Lithuania
The Philippines
Columbia
-4
-3
-2
-1
0
1
2
Grade-level equivalents relative to the Sweden

Table 1. The effect of curriculum-based external exit ...

	Diploma exam	Private share	Private-share LT .11
TIMSS science-1994			
8th grade mean	42.2*** (3.20)		
Median 13 yr. olds	33.9** (2.50)		
Median 13 yr. olds	30.1** (2.30)		
Median 13 yr. olds	32.3** (2.34)	- 20.6 (.69)	
Median 13 yr. olds	27.4* (2.00)		-304* (1.85)
TIMSS mathematics-1994			
8th grade mean	37.3** (2.52)		
Median 13 yr. olds	25.6* (1.79)		
Median 13 yr. olds	21.1 (1.45)		
Median 13 yr. olds	27.7* (1.86)	18.1 (.60)	
Median 13 yr. olds	19.6 (1.37)		-366** (2.30)
IEA reading-1990			
Average, age adjusted	24.5*** (3.06)		
Average, age adjusted	23.6** (2.79)		
Average, age Adjusted	22.0** (2.67)	-24.7 (1.46)	
Average, age adjusted	17.2** (2.16)		-216** (2.33)

Notes: Grade-level equivalents are about 26 for science and 24 for math and reading. The TIMSS analysis is based on 39 nations. The analysis of IEA reading data is based on 24-25 nations. T values are in parentheses under the coefficients.

... examinations on science, math, and reading achievement.

Private share GT .11	Ed. spending[a)]	Per capita GDP 87-91	East Asia	Adj. R^2 RMSE
		30.4***	11.1	.296
		(2.78)	(.67)	36.0
		37.3***	18.7	.313
		(3.59)	(1.09)	36.9
	15.0*	31.5***	33.4	.381
	(1.94)	(3.01)	(1.89)	35.8
		39.3***	23.7	.302
		(3.61)	(1.27)	37.2
19.9		42.2***	33.5*	.342
(.54)		(3.95)	(1.77)	36.1
		39.8***	57.8***	.400
		(3.72)	(3.33)	38.0
		48.5***	66.1***	.460
		(4.67)	(3.94)	36.7
	6.6	48.2***	75.0***	.494
	(.81)	(4.44)	(4.14)	37.0
		46.8***	61.8***	.450
		(4.33)	(3.36)	37.1
71.3*		50.7***	74.8***	.520
(2.00)		(4.96)	(4.16)	34.6
		24.2***	-21.6*	.618
		(3.46)	(2.04)	16.6
	2.1	23.7***	-18.6	.604
	(.47)	(3.29)	(1.48)	16.9
		27.9***	-11.8	.628
		(3.73)	(.93)	16.4
-4.1		29.7***	-3.6	.634
(.22)		(4.28)	(.29)	15.1

a) per cent of GDP. *** indicates the coefficient is significant at the 1% level on a two-tail test; ** indicates the coefficient is significant at the 5% level on a two-tail test; * indicates the coefficient is significant at the 10% level on a two-tail test.

But the analysis of achievement at a particular grade level may be biased by differing policies regarding grade retention, school-entry age, and which grade was chosen for assessment. CBEEESs, for example, might be associated with high rates of grade retention.[1] So a preferable dependent variable is a measure of student achievement at some fixed age. Rows 3 and 4 of each panel present estimated models that predict the median test score for each nation's 13 year olds (Beaton et al, 1996a,b, Table 1.5). For countries not in this table, the 13-year-old median was estimated by age-adjusting the 7th and 8th grade means.[2] Switching to the age constant achievement somewhat reduces the estimated impact of the CBEEES, but the effects remain statistically significant. Using two-tail *t* tests, the CBEEES coefficient has a P = .08 in the mathematics model and a P = .01 in the science model.[3] The estimated impacts are substantively important: **1.3** US grade-level equivalents in science and **1.0** US grade-level equivalents in math.

One of the ways CBEEESs may improve achievement is by inducing greater social investments in education. Row 3 presents results of regressions that add the share of GDP spent on education to the standard model. Coefficients on this variable are positive for outcomes and significantly so for science. But the estimated impacts of spending are modest. A one percentage-point increase in the share of GDP devoted to educa-

[1] School attendance is not universal at age 13 in some less-developed countries participating in TIMSS. TIMSS publications do not report age-specific school enrollment rates, but they report an indicator that sets a lower bound on age-specific school enrollment rates—the proportion of the nation's 13 year olds who were in one of the two grades tested. Developing countries with rates below 80 were Columbia (45%), Iran (72%), Portugal (76%), Romania (76%), and Thailand (78%) (Beaton et al, 1996, Table A3).

[2] The Philippines, for example, had a math score mean of 399 in 8th grade and a mean of 386 in 7th grade. The mean age of 8th graders was 14; the mean age of 7th graders was 12.9. The math score for 13.5 year olds was estimated by interpolation between 7th and 8th grade means. Math 13.5 = 386 + (399-386)*((13.5-12.9)/(14-12.9)).

[3] Sweden's actual test scores (particularly in reading and science) lie substantially above the predictions for it generated from model 2. It might be argued that the *Centrala Prov* will have effects similar to a CBEEES and that Sweden should thus be recoded as a .5 or a 1 on the exam variable. When that experiment was tried, R squares and the size of the CBEEES coefficient grew. So reclassifying the *Centrala Prov* as a CBEEES actually strengthens the primary finding of the paper that CBEEESs raise academic standards.

tion increases the science achievement of 13 year olds by only one half a grade-level equivalent.

3.2. The IEA study of reading literacy

The bottom panel in Table 1 presents an identical analysis of IEA reading achievement data. To avoid problems of differing school-entry ages and grade-retention policies, the age standardised reading scores provided in Appendix E of Elley (1992) were used in the analysis. The IEA study defined and measured three types of reading literacy—narrative, expository, and document—and an average of the three scores is the dependent variable. The specification is the same as that used to study science and math achievement. Here, the exam variable is an average of the math and science CBEEES dummy variables used in the analysis of TIMSS data. The results are similar as well. Diploma exams and per capita GDP have significant positive effects on reading achievement. Countries with larger private school enrolment shares appear, *ceteris paribus*, to have lower reading achievement, though not significantly so.

3.3. Effect of the size of the private school sector

The hypothesis that a large private school sector instigates a competitive environment that makes all schools better is tested in the bottom two rows of each panel. Row 4 presents the results of adding the share of primary and middle school students who attend private schools to the model. Adding the private share leaves the coefficients on CBEEES unchanged and does not improve model fit. The private sector size variable has inconsistent and statistically insignificant effects on average achievement levels. In the models of science and reading achievement, the point estimate is negative. In the math equation, it is positive but tiny. Clearly there is no linear relationship between the size of the private school sector and student achievement.

I also tested for a non-linear relationship. This was accomplished by allowing the slope of the relationship between private sector share and achievement to shift at some arbitrary kink point (i.e., including a spline). Two separate slopes were estimated: one for the region from zero to .11, the current US private school enrolment share, and one for the range from .11 to 1.0. The kink point of .11 is above the median of the variable and slightly below the mean, which is .139. The coefficients on the lower range are all significantly negative. They imply that countries that lack any private schools, such as Sweden, will tend to have a more than one grade-level equivalent achievement advantage over countries such as the US with modest-size private sectors when other things—GDP, Asia, and exam systems—are held constant. The upper region coefficient from the mathematics regression is statistically significant and positive. This suggests that the large size of the private school sector in Belgium, Hong Kong, and the Netherlands may be one of the reasons why math achievement is high in these three countries.[4]

3.4. The impact of CBEEESs on inequality of achievement

Policy-makers are also interested in knowing if the CBEEES affects the variance of student achievement and the level. To address that question, models were estimated that predict the standard deviation of student achievement for the 39 nations that participated in TIMSS. The specification was, with just one exception, the same as that used to predict achievement levels. To deal with possible distorting effects of floors and ceilings on the TIMSS achievement scales, the achievement level was included as an independent variable along with per capita GDP and dummy variables for East Asia and for a CBEEES. The results are in the top panel of Table 2. CBEEESs neither increase nor decrease the variance of student achievement. Per capita GDP and the dummy for East Asian nations have no effect either. The achievement level is the only variable with a statistically significant relationship with the standard deviation of achievement.

[4] There are two possible reasons for this non-linear relationship—one causal, the other not. The causal explanation proposes that a growing private school sector will weaken support of public schools causing them to be under funded. If there are no alternatives to public schools, activist parents will "voice" their concern by running for PTA president or the school's board of governors. Their pressure, it has been hypothesized, keeps the schools first class. When private schools are an option for most parents, the activists "exit" and their positive influence on the quality of the public school may be lost. The only way to avoid this fate is to tie the fortunes of the two sectors together by requiring students in both sectors to take the same courses and the same exams and by tying the subsidy of private school student to the subsidy of public school students. This is what Belgium, Hong Kong, and the Netherlands have done and it has resulted in a very large private sector. The other explanation proposes that unsubsidized private sectors (such as the one that has captured 11% of the market in the US) spring up when the public schools do a poor job. If public schools are of uniformly high quality, private schools have no market niche to fill. If the nation chooses to fund private schools on the same footing as public schools, they end up with over half of the market and their pressure forces the public schools to become better.

Table 2. How is science and mathematics achievement different in nations with CBEEESs?

	CBEEE	Log GDP/pop 1987-91	East Asia	13 yr. olds achiev.	Adj. R^2 RMSE	No. obs.
Variance of achievement						
TIMSS science SD	1.5	-1.0	-.2	.077	.016	40
(US = 105.5)	(.37)	(.30)	(.04)	(1.67)	10.2	
TIMSS mathematics SD	2.8	-3.2	- 2.2	.17***	.529	40
(US = 90)	(.96)	(1.25)	(.55)	(5.20)	7.2	
International Assessment of Educational Progress—1991						
Science, % correct,	4.3	1.4	9.3**		.429	15
13 yr. olds (US GLE =6)	(1.71)	(.49)	(2.79)		4.07	
Math, % correct,	15.4***	4.6	14.1**		.648	15
13 yr. olds (US GLE = 8)	(3.93)	(1.10)	(3.40)		5.95	
Geography, % correct adjusted	1.8	-2.7	-3.6		.071	20
	(1.45)	(1.28)	(.92)		3.47	
Geography, % correct adjusted	2.6		-1.9		.039	20
	(1.66)		(.51)		3.53	

Notes: In TIMSS data, US grade-level equivalents are about 26 for science and 24 for math. T values are in parentheses under the coefficients. *** indicates the coefficient is significant at the 1% level on a two-tail test. ** indicates the coefficient is significant at the 5% level on a two-tail test.* indicates the coefficient is significant at the 10% level on a two-tail test.

3.5 Analysis of the 1991 International Assessment of Educational Progress

Science and mathematics. The 1991 *International Assessment of Educational Progress* (IAEP) is the third data set in which CBEEE effects can be tested. Fifteen nations are available for the analysis: England, France, Hungary, Ireland, Israel, Emilia Romagna/Northern Italy, Korea, Portugal, Scotland, Slovenia, Soviet Union, Spain, Switzerland, Taiwan, and the US. Canadian data is analysed separately. Data from Brazil, Jordan, and Mozambique were not used because of the low levels of industrialisation. In IAEP, schools were first sampled, then students within schools. Sampling frames generally excluded separate schools for special education students and often very small schools as well. Israel assessed only its Hebrew-speaking schools, The Soviet Union assessed Russian language schools in 14 of the nation's 15 republics. Switzerland assessed 15 of 26 cantons. A school's likelihood of selection was roughly in proportion to its estimated number of 13 year olds. In most countries, school non-response rates were extremely low. They were zero in Hungary, Slovenia, Korea, and Taiwan and 3% in Israel and the Soviet Union. The countries with high non-response rates were Switzerland (17%), Emilia Romagna (18%), Scotland (19%), the US (21%), and England (48%). When sampled schools declined to participate, an alternate was selected from the same stratum (IAEP, 1992a,b,c). Random samples of 30–34 (13 year olds) were selected from each school. Half were assigned to the mathematics assessment and half assigned to the science assessment. The average per cent correct (adjusted for guessing) for 13-year-old students was regressed on the same set of variables as in the analysis of the TIMSS data.

The results are in the bottom panel of Table 2. For mathematics the effect of curriculum-based external exams is highly significant and quite large. Because the US standard deviation was 26.8 percentage points in mathematics, the CBEEE effect on math was more than one-half of a US standard deviation or about 2 US grade-level equivalents. CBEEEs had a smaller non-significant effect on science achievement. East Asian students scored significantly higher than students in Europe and North America. Coefficients on per capita GDP were positive but not statistically significant.

Geography. Nine of the countries in the IAEP study assessed geography as well as mathematics and science. The countries participating in the geography assessment were Canada, Hungary, Ireland, Korea, Scotland, Slovenia, Soviet Union, Spain, and the US. Canada collected sufficient data to allow valid comparisons between provinces and between the Anglophone and Francophone school systems of the five provinces with dual education systems. Some of these provinces have CBEEEs and others do not, so including the Canadian provincial data in the study increases the power of our tests for the effects of exams (IAEP, 1992d).

Regressions were estimated that predict country/provincial means using the same specifications as above. The logarithm of per capita GDP 1987-91 had the wrong (a negative) sign, so preferred specification is one that does not include this variable. In the preferred model, curriculum based exams has an almost significant effect on geography achievement. The effect appears to be roughly 15% of a US standard deviation. The Canadian provinces without examination systems do significantly worse than the provinces that have examinations systems (F for Hyp = 3.9).

These results are consistent with the causal hypotheses previously presented. But causation is not proved, because other explanations can no doubt be proposed. Other sources of variation in curriculum-based exams must be analysed. Best of all would be studies that hold national culture constant. Bishop, Moriarty and Mane (1997) have found that when socio-economic background is held constant, students from New York state, the only US state with a CBEEES, out perform students in other states on the NAEP math assessment and on the SAT-I. The next section presents a comparison of math and science achievement of Canadian students, who live in provinces with curriculum-based diploma examinations, with comparable students in provinces without such examinations.

3.6. Comparing Canadian provinces

In 1990-91, the year the IAEP data was being collected, Alberta, British Columbia, Newfoundland, Quebec, and Francophone New Brunswick had curriculum-based provincial examinations in English, French, mathematics, biology, chemistry, and physics during the senior year of high school. These exams accounted for 50% of that year's final grade in Alberta, Newfoundland, and Quebec and 40% in British Columbia. The other provinces did not have curriculum-based provincial external exit examinations in 1990-91 (White, 1993). Ontario eliminated them in 1967, Manitoba in 1970, and Nova Scotia in 1972. Anglophone New Brunswick had provincial exams in language arts and mathematics but exam grades were not reported on transcripts or counted in final course grades. Canadian provincial exams are medium stakes, not high stakes tests. They influence grades but passing the examination is not essential for graduation. Employers appear uninterested in exam scores. Job application forms do not request that applicants report exam scores or grades.

The principals of schools sampled by IAEP completed questionnaires that describe school policies, school resources, and the qualifications of 8th grade mathematics and science teachers. Students were asked about books in the home, number of siblings, language spoken at home, hours of TV, hours doing homework, pleasure reading, watching science programs on TV, parental oversight of school work, and teaching methods of teachers.

Table 3. Effects of Canadian provincial diploma exams ...

	Diploma exam	Religious school board	Independent religious school	Independent non-sectarian school
Mathematics				
1	.062***			
	(8.31)			
2	.041***			
	(5.30)			
3	.035***	-.041***	.149***	.091***
	(4.70)	(5.77)	(6.66)	(3.68)
4	.051***	-.048***	.085***	.061***
	(7.58)	(5.93)	(4.45)	(2.83)
5	.049***	-.058***	.085***	.059***
	(7.41)	(7.23)	(4.51)	(2.81)
6	.039***	-.046***	.066***	.024***
	(5.25)	(5.26)	(3.44)	(1.13)
Science				
1	.035***			
	(6.09)			
2	.026***			
	(4.57)			
3	.025***	-.041***	.032*	.044**
	(4.40)	(5.77)	(1.95)	(2.36)
4	.026***	-.029***	-.017	.012
	(5.12)	(5.88)	(1.18)	(.73)
5	.027***	-.040***	-.020	.011
	(5.19)	(6.48)	(1.38)	(.68)
6	.009	-.029***	-.052***	-.015
	(1.52)	(4.09)	(3.40)	(.85)
Principal reports problems with:				
Absenteeism	.14***	.00	-.42***	-.35**
	(3.14)	(.01)	(3.31)	(2.44)
Discipline	.017	-.13**	-.34***	-.16
	(.40)	(2.55)	(2.78)	(1.15)

Notes: The school problem indexes range from 0 = "no problem" to 3 for "serious". The means (standard deviations) were .78 (SD=.72) for discipline and .82 (SD=.77) for absenteeism.

... and school governance on student achievement.

Elementary K-11, French	Books, size, sibs, different home language	Computers, calculators	School inputs, policies, teacher qualifications, homework	Adj. R^2
				.0476
X				.0914
X				.1482
X	X			.3291
X	X	X		.3606
X	X	X	X	.3717
				.0258
X				.1029
X				.1308
X	X			.3230
X	X	X		.3316
X	X	X	X	.3592
X	X			.1312
X	X			.0795

Column 8 variables are per cent of students who have a computer and per cent who have a calculator. The school policy variables include: hours of TV, homework, specialist teachers, teacher majored in subject taught, hours of science class, science labs, class size, preparation time, teacher experience, and tracking.

The effect of curriculum-based provincial exit exams taken by 12th graders on achievement of Canadian 13 year olds was examined by estimating models that predict achievement using schools as observations. The data set comprises 1362 Canadian schools. The dependent variable is the school-mean, per cent 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, students achieved an average of 47.2% in math and 57.3% in science. The standard deviation across Canadian 13 year olds is 24 points for the math score and 20 points for the science score.

Row 1 in the top two panels of Table 3 presents simple regressions that contain no controls for school characteristics. Row 1 tells us that students in provinces with exam systems scored 6.2 points higher in math and 3.5 points higher in science. Adding controls for school type (school includes elementary grades, school includes K-11th grade, Francophone school) in row 2 lowers the *exam* effect to 4.1 points for math and 2.6 points for science. Adding additional controls for three types of school, governorship in row 3 lowers the *exam* effect to 3.5 points for math and 2.5 points for science.

Now let us examine what happens in row 4 when controls are added for the demographic background of the school's student body—school means for books at home, number of siblings, and proportion of students whose home language is different from the language of instruction and size of the school. Here, schools in *exam* provinces are 5.1 points higher in math and 2.6 points higher in science. The variables added to the model in rows 5 and 6 are hypothesised to be influenced by the existence of external exams. So row 4 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. The effect of provincial exams is about one-half of a US grade-level equivalent.

Gains in mean achievement generated by exam systems do not come at the expense of greater inequality. Exam provinces have less variability of achievement across schools. The variance of school-mean science achievement is smaller in Alberta, British Columbia, and Quebec than in any other province. For math achievement, the variance was lowest in New Brunswick, British Columbia, Saskatchewan, Quebec, and Alberta. Regressions predicting the standard deviation of achievement among students at a school were run and within-school standard deviations for science achievement are smaller in exam provinces (Bishop, 1998).

Is the *exam* effect causal? One possible sceptical response to a causal interpretation of these findings is to point out that omitted variables and/or selection effects may be biasing the coefficient on the CBEEES indicator variable. Maybe the people of Alberta, British Columbia, Newfoundland, Quebec, and Francophone New Brunswick (the provinces with exam systems) place higher priority on education than the rest of the nation.[5] Maybe this trait also results in greater political support for examination systems. If so, we would expect that schools in the exam provinces should be better than schools in other provinces along other dimensions such as discipline and absenteeism, not just by academic criteria. But that's not the case. Exam systems induce students and schools to *redirect* resources and attention to learning/teaching exam subjects and away from the achievement of other goals such as low absenteeism, good discipline, and lots of computers. These competing hypotheses are evaluated in the bottom panel of Table 3. Contrary to the "provincial taste for education" hypothesis, principals did not report significantly fewer discipline problems and were significantly more likely to report absenteeism problems.

Adding endogenous student behaviour and school-policy variables. Row 5 of the top two panels has two additional control variables—calculator availability and computer use. Having a calculator at home and using computers for school work is associated with higher math achievement but not higher science achievement. Including these variables in the model has little effect on the CBEEES coefficient. The models presented in row 6 of each panel add a full set of endogenous school policy, school input, and student behaviour variables (see table notes for list). So they provide a test of whether CBEEESs have direct effects on achievement when endogenous school policies and inputs,

[5] This will seem like a strange idea to people who know Canada. Except for Alberta and British Columbia, these provinces have little in common. Historically they do not seem to have valued education more than the rest of Canada. Adult literacy is lower on average in these four provinces than in the rest of Canada (Jones, 1993). Newfoundland is quite poor and has significantly lower levels of adult literacy than the rest of Canada. The adult literacy study placed Quebec in the middle of the pack along with Ontario, Nova Scotia, and New Brunswick. Two of the top four provinces with respect to adult literacy have exam systems and two do not. Saskatchewan is number one when it comes to adult literacy, yet its 13 year olds lag substantially behind students from Alberta and British Columbia in mathematics and science.

homework assignments, and TV watching are all held constant. For science, the CBEEES coefficient is still positive, but it is very small and no longer statistically significant. Apparently, just about all of the effects of CBEEESs on science learning operate through school policies and indicators of student time allocation (hours doing homework and hours watching TV). But for math, there is a substantial and significant direct effect of CBEEESs on math achievement even when student time allocation is held constant.

3.7. Effects of Sweden's elimination of exit examinations

Still another way to assess the impact of examination systems is to study the effects of eliminating them. Spain, Portugal, Latvia, Sweden, and some Canadian provinces have ended examination systems, but only for Sweden is their trend data that spans the period of change. Consequently, Sweden's elimination of exit exams in the early 1970s is potentially very instructive.

But Sweden not only ended its exit exams—it also established new rules for selecting applicants for university. Upper secondary students pursue specialised lines of study beginning in 10th grade. Universities had traditionally admitted recent graduates of the academic lines of study in upper secondary school that had durations of three years. Reforms introduced in the 1970s changed the rules governing competition for university places to favour those who had worked a few years after graduating from secondary school (Rehn and Helveg Peterson, 1980, p. 150). This encouraged many secondary school graduates to postpone entry into higher education to accumulate enough work points to gain admission to preferred educational programs. Partly for this reason and partly because many nurses decided to return to school to become physicians, entrants to the study of medicine, for example, had an average age of 30 in 1977 and 1978 (Rehn and Helveg Peterson, 1980, p. 154).

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 a result, natural science, the toughest secondary school major, was very easy to get into during the autumn of 1978. Ninety-three per cent 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% were admitted to their first-choice program.

These changes in secondary school examinations and university admissions policies appear to have caused a decline in the number of Swedish upper secondary students taking rigorous courses in mathematics and science during the 1970s.

Mathematics at graduation. Eight countries participated in the *First International Mathematics Study* in 1964 and its replication in 1982 (see Figure 4). The proportion of Swedish 18 year olds taking college-prep mathematics in the final year of secondary school fell from 16% to 12%. This more selected group of students scored only slightly higher 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% and significantly improved their mean scores (Husen, 1967; Robitaille and Garden, 1989).

Science at graduation. The proportion of Swedish 18 year olds in academic lines of study that were given the IEA science exam fell from 45% to 28% 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 5).[6] Finland, by contrast, increased the proportion of the age cohort assessed in science from 21% to 41% 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 sci-

[6] The data necessary to measure changes in absolute levels of achievement between the *First* and the *Second International Science Studies* are not available. But comparisons of achievement relative to that of other countries are possible and are presented in Figure 5. The zero point on the scale for each year was chosen to be the average performance for that year of the English students preparing for A-level exams. England was selected for this standardization role because there was no change in the proportion of its 19-year-old age cohort, who were in the sample frame for the IEA science study. 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 a nation's scores and participation rates changed relative to England. The 1983 study reports separate means for non-specialists (e.g., majors in economics and 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).

ence exam and it experienced a substantial increase in relative test scores (Comber and Keeves, 1973, p. 168; Postlethwaite and Wiley, 1992, p. 74).

Figure 4. Mathematics at the end of high school

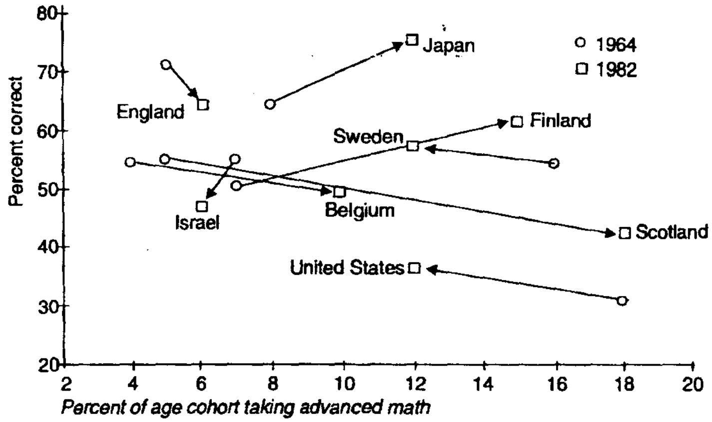
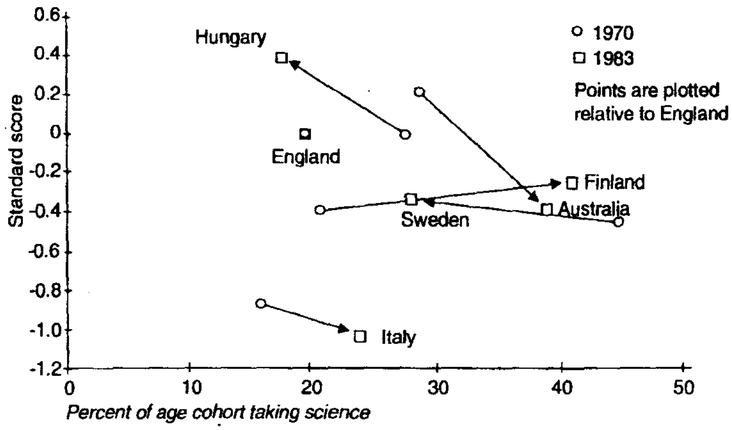


Figure 5. Science at the end of high school



4. Is parent, administrator, and teacher behaviour different in the presence CBEEESs? A look at the evidence

Are the teachers, school administrators, and parents in nations and provinces with CBEEESs behaving differently from their counterparts in nations and provinces that lack a CBEEES? This will be examined by estimating models that predict mean levels of these behaviours in a sample of 42 nations most of whom participated in TIMSS and in IAEP data on a sample of 1360 Canadian middle schools. The cross-section analysis of schooling inputs and indicators of student effort in 42 nations employs the same specification as that used in the cross-section analysis of TIMSS data on student achievement. It contains three variables: per capita GDP, a dummy variable for East Asia, and a dummy variable for CBEEES. Results are in Tables 4, 5, and 6.

The model that tested for the effects of CBEEESs in Canadian data was identical to the one presented in row 4 of Table 3. It contains 11 variables:

- Logarithm of the mean number of books in the home
- The mean number of siblings
- The proportion of the school's students whose home language was different from the language of instruction
- Logarithm of the number of students per grade in the school
- Dummies for schools run by a locally elected Catholic (or Protestant) school board
- Independent secular and non-secular schools
- Schools with primary grades
- Schools that include all grades (K-12) in one building
- French speaking schools
- A dummy for EXAM province.

Results are presented in Table 7.

Table 4. How is teacher and student behaviour different in nations with CBEEESs?

	CBEEE	Log GDP/ pop. 1987-91	East Asia	Adj. R^2 RMSE	No. obs.
Public investment in education					
% GDP spent on education	.24 (.81)	.55** (2.71)	-.86** (2.47)	.257 .76	41
Teacher qualifications index	1.53*** (3.80)	.70** (2.56)	-.18 (.39)	.304 1.04	42
% salary advantage of lower secondary teachers/all workers	30.6* (2.03)	8.6 (.26)	10.0 (.40)	.107 23.7	17
% salary advantage of upper secondary teachers/all workers	34.4 (1.64)	5.1 (.11)	12.7 (.36)	-.019 33.2	17
Teacher attitudes					
Teacher perception of relative status -1-100	- 6.1* (1.93)	-5.3** (2.19)	1.1 (.23)	.155 6.20	25
Society appreciates my work science teacher, % yes	- 5.7 (.60)	6.9 (.96)	37.5** (2.51)	.117 19.1	27
Would stay if teacher had opportunity to leave, % yes	- 10.7* (1.97)	- 5.0 (1.20)	1.5 (.18)	.063 11.0	27
Student time use					
Homework in other subjects (hours/week)	- 1.68** (2.14)	-2.03*** (3.81)	1.17 (1.31)	.279 1.76	37
TV and computer games (hours/week)	2.47** (2.06)	1.1 (1.33)	-.90 (.65)	.054 2.75	38
Playing or talking with friends (hours/week)	3.12** (2.64)	2.14*** (2.68)	-5.39*** (3.97)	.354 2.70	38
Playing sports (hours/week)	-.33 (.57)	.75* (1.92)	-3.65*** (5.55)	.509 1.31	38
Reading for pleasure (hours/week)	.05 (.83)	-.20*** (4.45)	.06 (.81)	.366 .15	38

Notes: *** indicates the coefficient is significant at the 1% level on a two-tail test. ** indicates the coefficient is significant at the 5% level on a two-tail test. * indicates the coefficient is significant at the 10% level on a two-tail test.

Table 5. How is science teaching different in nations with CBEEESs?

	CBEEE	Log GDP/ pop. 1987-91	East Asia	Adj. R^2 RMSE	No. obs.
Memorisation necessary to do well in science	-13.4** (2.31)	-4.3 (.98)	29.1*** (3.97)	.292 14.3	36
Hard work necessary to do well in science	-3.5 (1.30)	-.7 (.35)	6.3* (1.86)	.028 6.7	36
Natural talent necessary to do well in science	11.2* (2.02)	-7.1* (1.71)	11.4 (1.63)	.250 (13.7)	36
Hrs. teachers work outside school hours	.2 (.30)	.07 (.14)	11.8 (1.33)	-.017 1.73	36
Hours meeting/tutoring students after school-science	.42*** (3.00)	-.10 (.99)	.50*** (2.80)	.420 .35	36
Hours preparing or grading science tests	-.33 (1.30)	-.04 (.20)	.02 (.07)	-.034 6.3	36
Hours reading and grading science student work	.22 (.96)	.023 (1.37)	.64** (2.26)	.162 5.6	36
Frequency of student reasoning tasks in science	-.04 (.43)	-.27*** (4.08)	.03 (.31)	.293 .209	34
Hrs. science homework assigned/week by teacher	-.09 (.72)	-.18* (2.01)	-.06 (.42)	.050 .285	34
Hrs. science homework done/week, student report	-.88* (1.93)	-2.1*** (6.23)	.39 (.68)	.504 1.14	37
% gen. science students do experiments, pretty often+	26.8** (2.50)	23.5*** (3.17)	6.5 (.67)	.370 17.4	21
% gen. science students who like science	-.5 (.08)	-11.6** (2.74)	-.6 (.11)	.209 9.8	20

Notes: *** indicates the coefficient is significant at the 1% level on a two-tail test. ** indicates the coefficient is significant at the 5% level on a two-tail test. * indicates the coefficient is significant at the 10% level on a two-tail test.

Table 6. How is mathematics teaching different in nations with CBEEESs?

	CBEEE	Log GDP/ pop. 1987-91	East Asia	Adj R^2 RMSE	No. obs
Memorisation necessary to do well in math	-4.9 (.53)	-3.6 (.55)	16.0 (1.48)	-.016 21.7	37
Hard work necessary to do well in math	-2.0 (.68)	-.8 (.39)	3.7 (1.09)	-.045 6.9	37
Natural talent necessary to do well in math	8.3 (1.37)	-9.2 (2.17)	11.1 (1.60)	.212 14.0	37
Hrs./wk. math teachers work outside school hours	1.1 (1.25)	.15 (.24)	.21 (.21)	-.033 2.0	36
Hrs./wk. meeting/tutoring students after school, math	.34* (1.74)	-.13 (.91)	.39* (1.73)	.164 .45	36
Hrs./wk. preparing or grading math tests	-.23 (.85)	-.11 (.57)	-.23 (.72)	-.034 6.3	36
Hrs./wk. reading and grading math students' work	.31 (.98)	.22 (1.00)	.42 (1.16)	.018 7.3	36
Frequency of student reasoning tasks in math	-12.6* (1.83)	-.13.8*** (2.88)	2.6 (.33)	.161 15.7	34
Hrs./wk. of math homework assigned by teacher	-.07 (.35)	-.45*** (3.07)	.05 (.19)	.151 .484	38
Hrs./wk. math homework done, student report	-1.16** (2.13)	-1.84*** (4.81)	.81 (1.29)	.381 1.26	37
Teacher teaches entire class	-6.0 (1.10)	-10.4** (2.71)	10.5 (10.5)	.159 (1.66)	38
% quiz or test used pretty often+	-20.2** (2.51)	-2.2 (.39)	6.5 (.70)	.082 18.8	38
% math students who like math	-1.3 (.33)	-2.1 (.75)	-.4 (.08)	-.070 9.3	38

Notes: *** indicates the coefficient is significant at the 1% level on a two-tail test. ** indicates the coefficient is significant at the 5% level on a two-tail test. * indicates the coefficient is significant at the 10% level on a two-tail test.

Table 7. Effects of Canadian diploma exams on ...

School administrator behaviour	DepVar Mean	School SD
Specialist teachers do 8th grade math	.45	.50
Specialist teachers do 8th grade science	.46	.50
8th grade math teachers studied math in university	.64	.39
8th grade science teacher studied science in university	.69	.38
Math class hours/week	3.98	.88
Science class hours/week	2.93	.79
Good science labs	1.95	.95
Computers per student	.051	.043
Books per student	21	21
Average class size	24.8	6.1
Teacher prep time	.31	.27
Proportion teachers with 10+ yrs. experience	.60	.24
Proportion teachers, less than 3 yrs. experience	.16	.15
Teacher behaviour		
Total homework (hours/week)	4.41	1.62
Emphasises whole number computation	1.68	.49
Students do science experiments	1.52	.63
Teachers do science experiments	2.42	.47
Math quiz/test index	1.62	.52
Science quiz/test index	.89	.38
Students work on math problems alone	3.22	.41
Solve math problems in groups	1.46	.66
Home behaviour and attitudes		
TV hours/week	14.7	2.85
Read for fun index	1.85	.28
Watch *NOVA, Nature*	.97	.38
Parents talk about math class	.62	.17
Parents talk about science class	.47	.17
Parents want me to do well in math	3.54	.22
Parents are interested in science	2.18	.34
Believe science useful in everyday life	2.46	.31
Math useful in solving everyday problems	3.03	.31

Source: Analysis of IAEP data on 1338 Canadian schools.

Notes: Control variables not shown included: logarithm of the mean number of books in the home, the mean number of siblings, the proportion of the school's students whose home language was different from the language of in-

... administrator, teacher and parent behaviour.

Diploma exam			
Coefficient	**T statistic**	**Books in home**	**Adj. R^2**
.18	(6.9)	.074***	.280
.15	(5.6)	141***	.279
.19	(7.0)	.067**	.127
.19	(8.5)	.047	.199
.33	(5.9)	-.254***	.124
.16	(3.5)	-.006	132
.28	(5.6)	.037	.274
.001	(0.6)	.004	.195
1.58	(1.1)	3.13*	.157
-.15	(0.5)	.72*	.369
.001	(.01)	-.013	.222
-.025	(1.6)	-.037*	.090
.012	(1.2)	-.002	.078
.66	(7.1)	-.146	.149
-.09	(2.9)	-.011	.035
.28	(7.5)	.031	.145
.15	(5.4)	-.069*	.111
.10	(3.8)	-.074**	.391
.10	(4.9)	-.007	.206
.01	(.35)	.026	.060
-.04	(1.0)	-.072	.137
-.68	(4.2)	-2.69***	.255
.05	(2.8)	.264***	.115
.06	(2.7)	-.090***	0.91
.04	(3.4)	.016	.046
.06	(5.2)	.074***	.056
.05	(3.1)	.084***	.104
.06	(2.6)	.209***	.071
.06	(2.7)	-.097***	.095
.01	(0.5)	.114***	.084

struction, logarithm of the number of students per grade in the school, and dummies for schools run by a locally elected Catholic (or Protestant) school board, independent secular and non-secular schools, schools with primary grades, schools that include all grades in one building, and French speaking school.

The next section is organised around the hypotheses set forth in Section 2. I begin by examining hypothesis 5—the impact of CBEEESs and per capita GDP on indicators of school quality, such as spending, class size, teacher salaries, and qualifications (X in the theory in Bishop, 1996). I then turn to issues of teacher pedagogy—hypotheses 7 and 8—indicators of X and E in our theory. The section concludes with an analysis of how student and parental attitudes and behaviour are affected by per capita GDP and CBEEESs.

4.1. Education spending, teacher qualifications, and salaries: hypothesis 5-F to 5-K

The analysis of international cross-section data on spending and teacher qualifications presented in the top panel of Table 4 supports the hypothesis that rich countries will buy higher-quality schooling inputs (more X) than poor countries. More developed countries invest a larger share of GDP in K-12 education and also set higher minimum qualifications for entry into secondary school teaching jobs. Schools serving wealthier communities in Canada are more likely to have their teachers specialise in teaching one subject and are more likely to hire math teachers who studied the subject in university.

Contrary to our hypothesis, CBEEES nations do not spend significantly more on K-12 schooling (see row 1 of Table 4) and within Canada, class sizes are not smaller and time allowed for preparing lessons is not greater in CBEEES provinces (see the first panel of Table 7). But indicators of teacher quality such as "having studied the subject in university," 8th grade teachers being specialists in teaching their subject (rows 1-4 of Table 7), and the minimum qualifications for becoming a secondary school teacher are all significantly higher in CBEEFS provinces and nations. There is also evidence from a cross-section analysis of 17 advanced countries that relative pay levels are higher for secondary school teachers in CBEEES countries (see rows 3 and 4 of Table 4). But the higher pay has not resulted in teachers becoming more satisfied with their status in society. Teachers in CBEEES nations perceived themselves to have lower relative status and were significantly more likely to report wanting to leave the profession if an opportunity came along (see the second panel of Table 4). In addition, there is no tendency in the Canadian data for the teaching profession in CBEEES provinces to have a higher average experience level. This suggests the possibility that exit rates from the profession may be higher in CBEEES nations. Why are better-paid teachers less satisfied? Possibly because teachers who are part of a CBEEE system lose the ability to adjust their expectations for students to their own ability to teach the material. Students learn more, but teachers work harder and are under much greater stress because their 'success' or 'lack of success' as teachers is now more visible to others. This in turn helps explain why salaries are higher. Not only are more qualified teachers needed, they must receive extra compensation for the negative non-pecuniary characteristics (stress) of their job.

4.2 The quality of physical facilities: hypothesis 5-A

Our theory predicts that facilities that are not believed to directly improve learning in externally examined subjects will not be better in CBEEES provinces. I would put athletic facilities, computers, and libraries in this category. Facilities that are essential for high-quality instruction in examination subjects, such as science laboratories, will be better in the presence of a CBEEES system.[7] The analysis of Canadian data supports this hypothesis. Science labs are significantly better in CBEEES provinces. But library books and computers are no more prevalent in CBEEES provinces than in other provinces.

4.3. Instructional hours: hypothesis 5-B and 5-C

When local school administrators have discretion over the amount of instruction time that is allocated to examination subjects, our theory predicts a CBEEES will induce them to allocate a more time to examination subjects. It is not clear whether the extra time will be obtained by extending the school day or by reducing time spent in study halls and subjects that are not externally examined. The analysis of Canadian data supports this hypothesis. Scheduled hours of instruction were 8% higher in CBEEES provinces for mathematics and 5% higher for science. Total hours in the school year did not increase, so it appears the extra math and science instruction came at the expense of something else (see row 5 and 6 of Table 7).

4.4. Teacher pedagogy: hypotheses 7 and 8

Some educators worry that external examinations will lower the quality of instruction. The fear is that science teachers will be induced to focus on facts and definitions—not the scientific process. The dis-

[7] In addition, language labs are hypothesized to be better when listening and speaking skills are an important part of the external examination.

covery approach to teaching science—with students learning from experiments they conduct themselves–will be de-emphasised. In mathematics, computation will be emphasised to the exclusion of more advanced topics. Finally, students will start believing that doing well in the subject depends on "memorising the textbook and class notes." The results of testing these hypotheses are in Tables 5 and 6 and in the second panel of Table 7.

The analysis of TIMSS data found, to the contrary, that students were *less likely* to believe that memorisation was necessary to learn math and science in CBEEES nations. The difference was statistically significant for science. Teachers were *significantly more likely* to have students do experiments in class. The greater use of experiments to teach science was also found in the analysis of Canadian data. Teachers in CBEEES nations reported a significantly lower frequency of students doing reasoning tasks in math but not in science. In math, the emphasis placed on computation of whole numbers by 8th grade teachers was significantly lower in Canadian provinces with CBEEESs. There were no differences between provinces in the amount of time students spent working in groups to solve math problems or in the time spent doing math problems on their own. The overall conclusion from these comparisons is a resounding rejection of hypothesis 8. There is no evidence that the prospect of an external examination toward the end of secondary school causes the quality of instruction to decline in 8th grade. Along several dimensions there were no significant differences in pedagogy. Where CBEEESs had a statistically significant relationship with an indicator of pedagogy, the quality of instruction was, in every case but one, higher in the CBEEES nations/provinces.

According to the *Effective Schools* literature, frequent quizzes and tests is one of the traits of effective schools. This leads one to predict that teachers would give quizzes and tests more frequently in CBEEES nations and provinces. The Canadian data support this hypothesis, but the TIMSS data do not. Teachers in countries with CBEEESs spend about the same amount of time "preparing and grading tests" and "reading and grading student work" as teachers in countries without a CBEEES. Students report they are given fewer tests in CBEEES nations. The only significant difference in the allocation of teacher time was the finding that teachers in CBEEES nations spent considerably more time meeting with and tutoring students.

4.5. Student effort—homework versus TV and hanging out: hypothesis 7-B

The theory in Bishop (1996) could not sign the relationship between national wealth/productivity and student effort. But the empirical findings are unambiguous. Less homework is assigned in rich countries and a lot less is done. A doubling of GDP per capita reduces homework done by 4.2 hours per week or about one-third and increases time talking with friends, watching TV, or engaging in sports by 2.8 hours/week. Within Canada, students at schools that serve wealthier communities watch significantly less TV but tend to spend slightly less time (p=.22 on a 2 tail test) doing homework.

Our theory predicts that students will be expected to (and will) work harder when they face a CBEEES in the future. There is strong support for this hypothesis in the Canadian data. Students in CBEEES provinces spent significantly more time doing homework and significantly less time watching TV. But in the TIMSS analysis, teachers in nations with a CBEEES did not assign more homework than teachers in other nations. Students in CBEEES nations reported doing *less* homework. In fact, when the differences on all three homework variables are added together, students in CBEEES nations say they spend about 4 fewer hours a week doing homework, a reduction of about one-third relative to the international mean. They also reported spending an extra 5.6 hours a week watching TV, playing video games, and being with friends. There were no differences in time spent playing sports. Apparently, the image of the over-worked Japanese student preparing for examination hell does not characterise most CBEEES nations, at least not in 7th and 8th grade.

This finding suggests a need to revise the theory. One place where a revision may be needed is the assumption that teachers do not use direct measures of student effort such as homework assignments turned in as part of their grading formula. There might be two kinds of student effort–one invisible to teachers and the other visible to teachers and therefore part of the teacher's grade. When a CBEEE becomes the only signal of student achievement, the teachers might no longer be able to induce students to hand in long homework assignments. Under this assumption, a CBEEES might increase invisible student effort while simultaneously decreasing visible student effort. This did not happen in Canada because the external exam results were only a part of the final grade. TIMSS has data on how teachers evaluate students so it should be possible to investigate this hypothesis.

4.6. The hypothesised damaging effects of external rewards tied to external exams: hypothesis 4

Critics of externally set curriculum-based examinations predict that the exams will cause students to avoid learning activities that do not enhance exam scores. This hypothesis was operationalised by testing whether exam systems were associated with *less* reading for pleasure, *less* watching of science programs, such as *NOVA* and *Nature*, and fewer students having positive attitudes toward the subject. None of these hypotheses is supported. In the TIMSS data, students liking math and science and time spent reading for pleasure was unrelated to the existence of a CBEEES. In Canadian data, students in exam provinces spent significantly more time reading for pleasure, significantly more time watching science programs on TV, while watching less TV overall. Student attitudes were also more positive. Students in CBEEES provinces were significantly more likely to believe that "science is useful in everyday life." Regarding beliefs that "math is useful in solving everyday problems," there were no differences between CBEEES and non-CBEEES provinces. Finally, students in exam provinces were not more likely to say that extrinsic rewards (e.g., getting a better job) would result from doing well in math and science. Their view apparently reflects Canadian reality. A check of a small sample of job applications obtained from Canadian employers revealed that they typically do not ask for exam grades or GPAs.

4.7. Impacts on the behaviour of parents: hypothesis 3

Information on parental attitudes and involvement with their children's studies was available only in the IAEP data on Canada. Parents in CBEEES provinces were significantly more likely to talk to their children about their math and science classes, and their children were more likely to report that their parents "are interested in science" or "want me to do well in math." Parents in high SES schools were also more likely to talk to their children about science class and to be reported to be "interested in science" and to "want me to do well in math."

5. Summary

Theory predicts that improved signalling of academic achievement that results from curriculum-based external exit exams will raise achievement in examined subjects. Testing this hypothesis is difficult because it can be assessed only by comparing education systems, not by comparing individuals, classrooms, or schools. Consequently, the number of independent observations is inevitably small and the power of formal statistical tests is correspondingly reduced. The small number of independent observations also heightens worries about omitted variable bias and selection bias. Our approach to dealing with these inference problems is to find as many comparative data sets as possible that cover different sets of countries and provinces and different subjects and to test whether there is a consistent tendency for jurisdictions with CBEEESs to have higher achievement, as theory predicts.

Our review of the evidence suggests that theory and the claims of advocates of such examination systems are probably correct. Analyses of three very different international cross-section data sets found that students from countries with such systems outperform students from other countries at a comparable economic development level in four different subjects—science, mathematics, reading and geography. The proportion of Swedish students taking rigorous mathematics and science courses in upper secondary school fell after exit examinations were eliminated in the early 1970s. Additionally, students living in Canadian provinces (and US states) with such exams know more science and mathematics than students in other provinces/states. The variance of student achievement in countries and provinces with a CBEEES is no higher than in jurisdictions that lack such exams.

The paper then turned to an investigation of how the higher achievement comes about. CBEEESs are not associated with higher teacher-pupil ratios nor greater spending on K-12 education. But they are associated with higher minimum standards for entry into the teaching profession, higher teacher salaries, a greater likelihood of having teachers specialise in teaching one subject in middle school, a greater likelihood of hiring teachers who have majored in the subject they will teach, and additional hours of instruction in examination subjects. Teachers' satisfaction with their jobs appears to be lower, possibly because of the increased pressure for accountability that results from the existence of good signals of individual student achievement. Science labs are better in CBEEES jurisdictions. The number of computers and library books per student are unaffected by CBEEESs.

Fears that CBEEESs have caused the quality of instruction to deteriorate seem unfounded. Students in CBEEES jurisdictions are less likely to report that memorisation is the way to learn the subject and more likely to report that they did experiments in science class. Apparently, teachers subject to the subtle pressure of a provincial exam four years in the future adopt strategies that are conventionally viewed as "best practice," not strategies designed to maximise scores on multiple-choice tests. Quizzes and tests are more common, but in other respects, our indicators of pedagogy are no different in CBEEES jurisdictions. Students are also more likely to get tutoring assistance from teachers after school. They are not less likely to say that they like the subject, and they are more likely to agree with the statement that science is useful in every day life. Students also talked with their parents more about school work and reported that their parents had more positive attitudes about the subject.

Some of the evidence on how student effort varies with exam systems seems inconsistent. In the analysis of Canadian data, students in CBEEES provinces did more homework and watched less TV. But in the cross-section of nations analysis, students in CBEEES nations did considerably less homework and spent considerably more time watching TV and hanging out with friends. Additional theoretical and empirical work is needed to resolve this discrepancy.

Important as CBEEESs may be, they are not the only or even the most important determinant of achievement levels. General productivity levels and standards of living, and an East Asian culture appear to have even larger effects. CBEEESs are common in developing nations where achievement levels are often quite low (e.g., Columbia and Iran). Belgium and Sweden, by contrast, have high-quality education systems without having a CBEEES. But the results suggest that achievement would be even higher if the *Centrala Prov* in Sweden became an important part of students' final course grades (as occurs with Canadian diploma exams) and if school-by-school results on the exam were published in national newspapers. More research on the system level determinants of average achievement levels is clearly in order.

References

American Federation of Teachers (1995), Setting Strong Standards: AFT's Criteria for Judging the Quality and Usefulness of Student Achievement Standards (American Federation of Teachers, Washington, D.C.) 1-12.

Beaton, A. et al., (1996a), Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (CSTEEP, Boston College, Boston MA).

Beaton, A. et al., (1996b), Science Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study (CSTEEP, Boston College, Boston MA).

Bishop, J. (1989), The Productivity Consequences of What is Learned in High School, Journal of Curriculum Studies 22, 101-126.

Bishop, J. (1993), The Impact of Academic Competencies on Wages, Unemployment and Job Performance, in: B. Malkiel and C. Plosser (eds.), Carnegie/Rochester Conference Series on Public Policy 37,127-195.

Bishop, J. (1996), The Impact of Curriculum-Based External Examinations on School Priorities and Student Learning, International Journal of Education Research 23, 1-110.

Bishop, J. (1997), The Effect of National Standards and Curriculum-Based External Exams on Student Achievement, American Economic Review 87, 260-264.

Bishop J. (1998), Do Curriculum-Based External Exit Exam Systems Enhance Student Achievement?, University of Pennsylvania, Consortium for Policy Research in Education, CPRE Research Report RR-40 , 1-32.

Bishop, J., J. Moriarty and F. Mane (1997), Diplomas for Learning, not Seat Time: The Effects of New York State's Regents Examinations, Economics of Education Review, forthcoming.

Bob, Interview conducted in Calgary Alberta in May 1996.

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

Competitiveness Policy Council (1993), Reports of the Subcouncils, March (Competitiveness Policy Council, Washington, D.C.)

Costrell, R. (1994), A Simple Model of Educational Standards, The American Economic Review 84, 956-971.

Costrell, R. (1997), Can Centralized Educational Standard Raise Welfare, Journal of Public Economics, 65, 271-293.

Eckstein, M. and N. Harold (1993), Secondary School Examinations (Yale University Press, New Haven).

Elley, W. (1992), How in the World do Students Read? (International Association for the Evaluation of Educational Achievement, The Hague).

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

Hart (Peter D.) Research Associates (1995), Valuable Views: A Public Opinion Research Report on the Views of AFT Teachers on Professional Issues, American Federation of Teachers, Washington D.C, 1-24.

Hollenbeck, K., and B. Smith (1984), The Influence of Applicants' Education and Skills on Employability Assessments by Employers (National Center for Research in Vocational Education, Ohio State U., Columbus).

Husen, T. (1967), International Study of Achievement in Mathematics (John Wiley & Sons, New York).

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

International Assessment of Educational Progress (1992b), Learning Mathematics (Educ. Testing Service, Princeton, NJ).

International Assessment of Educational Progress (1992c), IAEP Technical Report. Volume 1 (Educ. Testing Service, Princeton, NJ).

International Assessment of Educational Progress (1992d), Learning About the World (Educ. Testing Service, Princeton, NJ).

Jones, S. (1993), Reading But Not Reading Well (National Literacy Secretariat, Ottawa Ontario) 1-22.

Kang, S. (1985), A Formal Model of School Reward Systems, in: J. Bishop (ed.), Incentives, Learning and Employability (National Center for Research in Vocational Education, Columbus Ohio).

Kreeft, H. P. J. (1990), Issues in Public Examinations, International Association for Educational Assessment.

Madaus, G. (1991), The Effects of Important Tests on Students: Implications for a National Examination or System of Examinations, American Educational Research Association Invitational Conference on Accountability as a State Reform Instrument, Washington, D.C., June, 1-19.

Madaus, G. and T. Kellaghan (1991), Examination Systems in the European Community: Implications for a National Examination System in the US, Contractor Report for the Office of Technology Assessment, US Congress, Washington, D.C, March 27, 1-100.

Postlethwaite, T. N. and D. E. Wiley (1992), Science Achievement in Twenty-Three Countries (Pergammon Press, London).

Rehn, G. and K. Helveg Peterson (1980), Sweden, Education and Youth Employment in Sweden and Denmark (Carnegie Council on Policy Studies in Higher Education, New York)

Robitaille, D. and R. Garden (1989), The IEA Study of Mathematics II: Contexts and Outcomes of School Mathematics (Pergamon Press, New York).

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

Steinberg, L. (1996), Beyond the Classroom (Simon and Schuster, New York) 1-223.

Thomas, J. W. (1991), Expectations and Effort: Course Demands, Students Study Practices and Academic Achievement, paper presented at a 1991 Conference on Student Motivation sponsored by the Office of Educational Research and Improvement.

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