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).

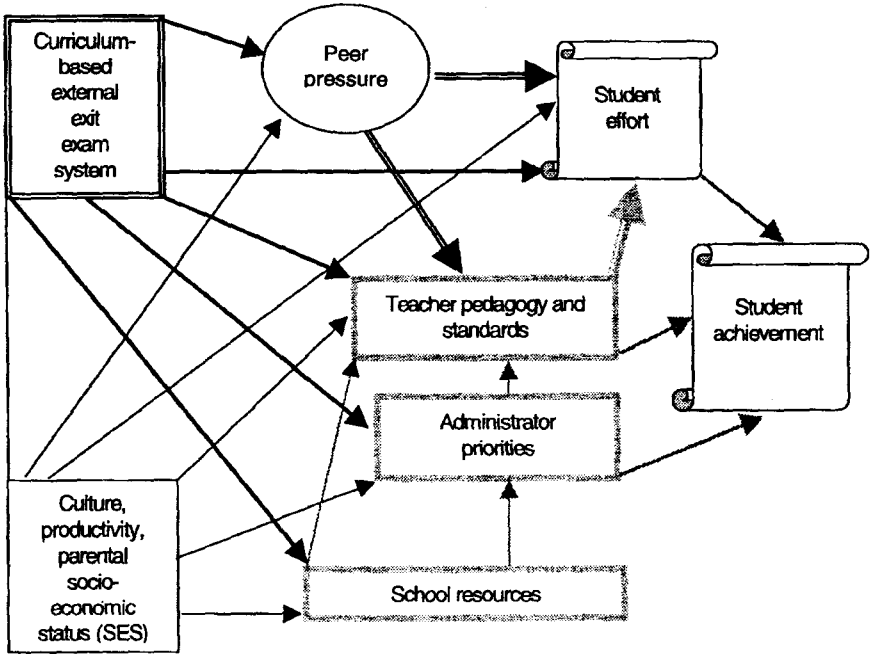
Steinberg, Brown and Dornbush conclude similarly that "The adolescent peer culture in America demeans academic success and scorns students who try to do well in school (1996, p.19)." My conversations with Swedish students sometimes generate similar anecdotes.

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.

CBEE 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 uni-

versally 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 CBEEES 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 CEEEES.

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.

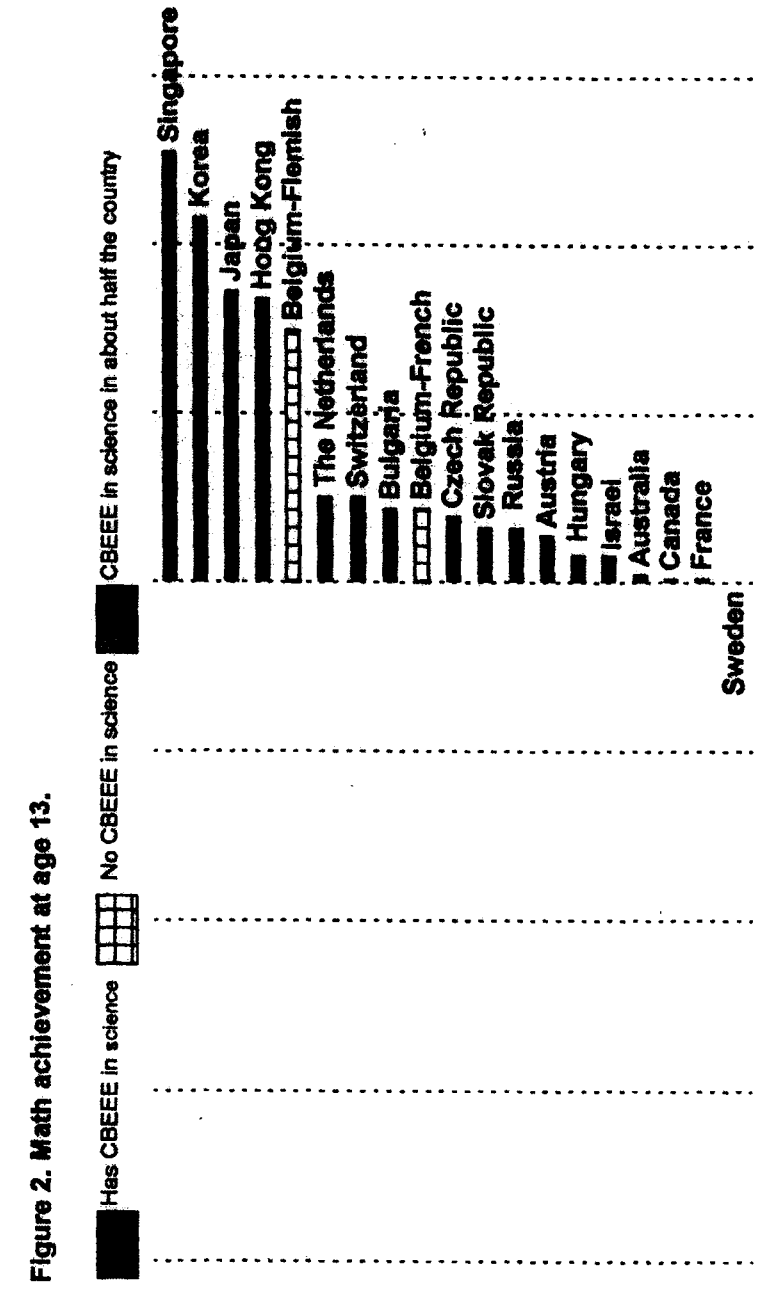


Figure 2. Math achievement at age 13.

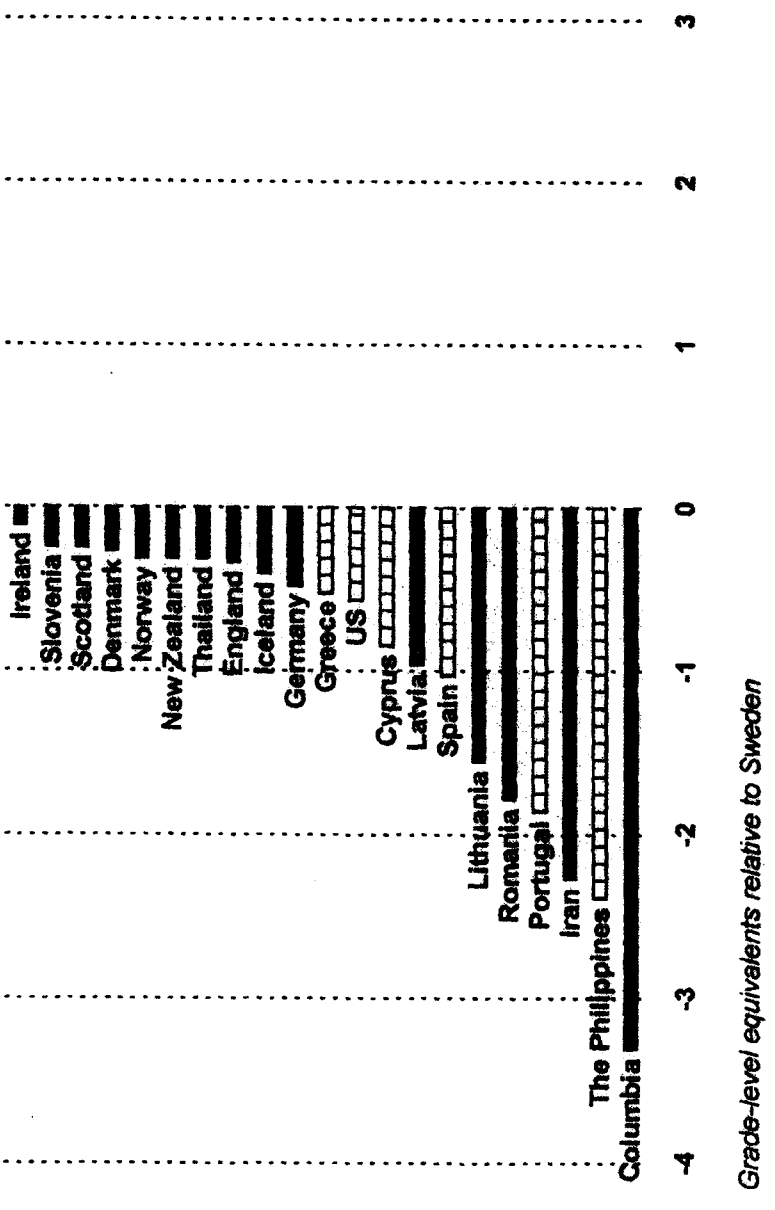

Ireland
Slovenia
Scotland
Denmark
Norway
New Zealand
Thailand
England
Iceland
Germany
Greece
US
Cyprus
Latvia
Spain
Lithuania
Romania
Portugal
Iran
The Philippines
Columbia
3
2
1
0
-1
-2
-3
-4
Grade-level equivalents relative to Sweden

Figure 3. Science achievement at age 13.

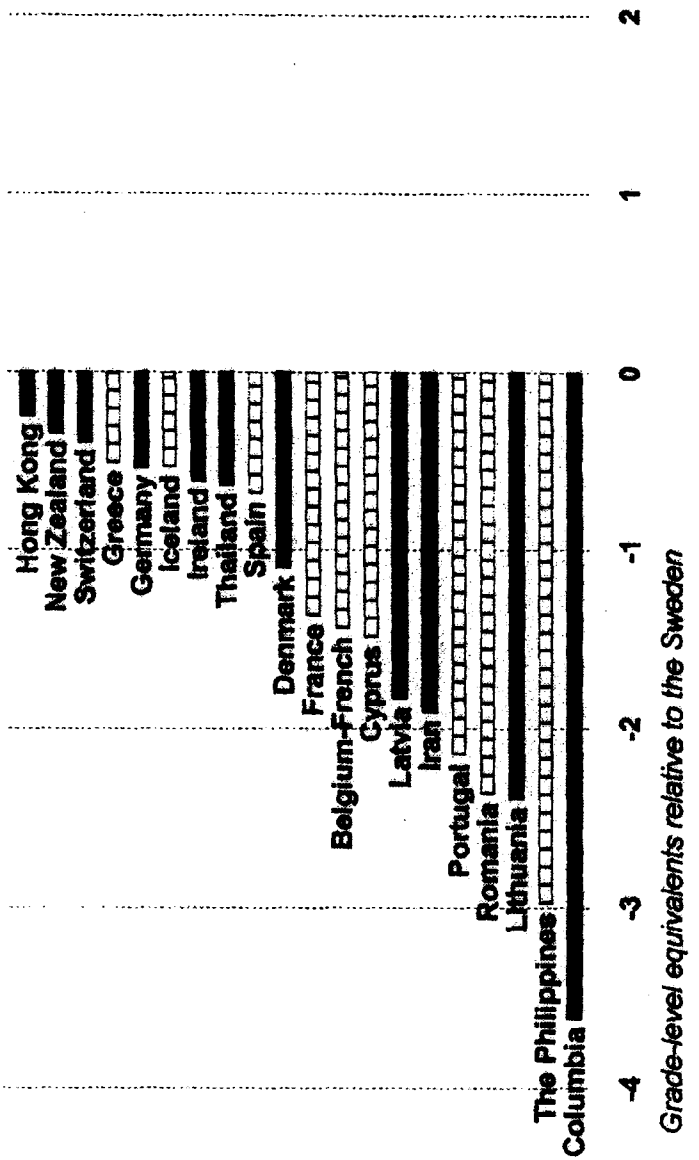

Hong Kong
New Zealand
Switzerland
Greece
Germany
Iceland
Ireland
Thailand
Spain
Denmark
France
Belgium-French
Cyprus
Latvia
Iran
Portugal
Romania
Lithuania
The Philippines
Columbia
2
1
0
-1
-2
-3
-4
Grade-level equivalents relative to the Sweden

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 ad-

vantage 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	1.8	-2.7	-3.6		.071	20
adjusted	(1.45)	(1.28)	(.92)		3.47	
Geography, % correct	2.6		-1.9		.039	20
adjusted	(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