

WHAT IS THE APPROPRIATE ROLE FOR STUDENT ACHIEVEMENT STANDARDS?

John H. Bishop*

Three presidents, the National Governors Association, and numerous blue-ribbon panels have called for the development of state content standards for core subjects and examinations that assess the achievement of these standards. 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 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” (American Federation of Teachers 1995, pp. 1–2).

American policymakers are trying to deal with low standards and weak incentives for hard study by making students, staff, and schools more accountable for learning. The education departments of the 50 states have responded by developing content standards for core academic subjects, administering tests assessing this content to all students, pub-

* Associate Professor of Human Resource Studies, New York State School of Industrial and Labor Relations, Cornell University. The preparation of this paper was made possible by support from the Center for Advanced Human Resource Studies and the Consortium for Policy Research in Education (funded by the Office of Educational Research and Improvement, U.S. Department of Education). The findings and opinions expressed in this report do not reflect the position or policies of the Office of Educational Research and Improvement or the U.S. Department of Education. This paper has not been formally reviewed or approved by the faculty of the ILR School. It is intended to make results of Center research, conferences, and projects available to others interested in human resource management in preliminary form to encourage discussion and suggestions.

lishing individual school results, and holding students and schools accountable for student achievement. While these efforts are generically referred to as standards-based reform, the mix of initiatives varies a great deal from state to state.

It is claimed that a curriculum-based external exit exam (CBEEE) system based on world-class content standards will improve the teaching and learning of core subjects. What evidence is there for this claim? What impacts have such systems had on school policies, teaching, and student learning? Outside the United States, CBEEE systems are the rule, not the exception. Within the United States, New York's Regents exams and North Carolina's end-of-course (EOC) exams are two examples of such systems. Do New York and North Carolina students outperform students with similar socioeconomic backgrounds from other states?

CURRICULUM-BASED EXTERNAL EXIT EXAMINATION SYSTEMS

While a number of states—for example, Maryland, Mississippi, Oklahoma, Arkansas, Tennessee, Virginia, Michigan—appear to be planning to implement CBEEE systems, only two states—New York and North Carolina—had established such systems by the beginning of the 1990s. State-sponsored end-of-course exam systems are provided in Appendix Table 1. The granddaddy of these is New York's Regents exam system. It has been in continuous operation since the 1860s. Panels of local teachers grade the exams using rubrics supplied by the state Board of Regents. Exam scores appear on transcripts and are the final exam mark that is averaged with the teacher's quarterly grades to calculate the final course grade. A college-bound student taking a full schedule of Regents courses would typically take Regents exams in mathematics and earth science at the end of ninth grade; mathematics, biology, and global studies at the end of tenth grade; mathematics, chemistry, American history, English, and foreign language at the end of eleventh grade; and physics at the end of twelfth grade. However, taking Regents courses and, therefore, Regents exams was voluntary until late in the 1990s. Prior to 1998, nearly half of the students chose to take "local" courses originally intended for noncollege-bound students, knowing that good grades could be obtained without much effort.

Between 1987 and 1991, North Carolina introduced end-of-course exams for Algebra 1 and 2, Geometry, Biology, Chemistry, Physics, American History, Social Studies, and English 1. Versions of these courses that are not assessed by a state test do not exist, so virtually all North Carolina high school students take at least six of these exams. Test scores appear on the student's transcript, and most teachers have been incorporating EOC exam scores in course grades. Starting in the year 2000, state law requires the EOC tests to have at least a 25 percent weight in the final

course grade. Clearly from this description one can see that even North Carolina's EOC exams and New York's Regents exams prior to 1999 carried only low to moderate stakes for students.

MINIMUM COMPETENCY EXAMINATIONS

Most states pursuing standards-based reform have established minimum competency exam (MCE) or other test-based school accountability systems that are quite different from curriculum-based external exit exam systems. Appendix Table 2 presents information on the end-of-grade (EOG) examination systems that a number of states have adopted, often to determine eligibility for honors diplomas or scholarships.

Eighteen states have MCE graduation requirements, and another 11 states are developing or phasing in MCEs. Minimum competency exams raise standards, but probably not for everyone.¹ The standards set by the teachers of honors classes and advanced college prep classes are not changed by an MCE. Students in these classes generally pass the MCE on the first try without special preparation. The students who are in the school's least-challenging courses experience the higher standards. Students pursuing a "do the minimum" strategy are told "you must work harder" if you are to get a diploma and go to college. School administrators want to avoid high failure rates, so they are likely to focus additional energy and resources on raising standards in the early grades and improving the instruction received by struggling students.

SCHOOL REPORT CARDS AND STAKES FOR TEACHERS AND ADMINISTRATORS

Formal systems for holding schools accountable are growing in popularity. In 1999, 37 states were publishing school report cards for all or almost all of their schools (Edwards 1999). Publicly identifying low-performing schools is intended to spur local school administrators and boards of education to undertake remedial action. Nineteen states had a formal mechanism for rewarding schools either for year-to-year gains in achievement test scores or for exceeding student achievement

¹ Minimum competency exams are in addition to—not a replacement for—teacher-imposed standards. In an MCE regime, teachers continue to control the standards and assign grades in their own courses. Students must still get passing grades from their teachers to graduate. The MCE regime imposes an additional graduation requirement and thus cannot lower standards (Costrell 1994). The Graduate Equivalency Diploma (GED), by contrast, offers students the opportunity to shop around for an easier (for them) way to a high school graduation certificate. As a result, the GED option lowers overall standards. This is reflected in the lower wages that GED recipients command (Cameron and Heckman 1991).

targets (Edwards 1999). Nineteen states had special assistance programs to help failing schools turn themselves around. If improvements were not forthcoming, 11 states had the power to close down, take over, or reconstitute failing schools.

Exactly how are these student and school accountability systems similar to or different from the curriculum-based external exit exam systems that are found abroad and in New York and North Carolina? We begin by noting the features they have in common. The following five criteria apply to CBEEEs and MCEs:

1. The exams produce signals of accomplishment that have real consequences for students and schools. While some stakes are essential, high stakes may not be necessary. Analyses of Canadian and U.S. data summarized below suggest that moderate stakes may be sufficient to produce substantial increases in learning.

2. The exams 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 possible. Costrell's (1994) analysis of the optimal setting of educational standards concluded that more centralized standard-setting (state or national achievement exams) results in higher standards, higher achievement, and higher social welfare than decentralized standard-setting (in other words, teacher grading or schools' graduation requirements).

3. The exams assess a major portion of what students are expected to know and be able to do. Studying to prepare for an exam (whether set by one's own teacher or by a state department of education) should result in the student's learning important material and developing valued skills. Some MCEs, CBEEEs, and teacher exams do a better job of achieving this goal than others. External exams, however, cannot assess every instructional objective. Teachers themselves must accept responsibility for evaluating dimensions of performance that cannot be reliably assessed by external means or that local leaders want to add to the learning objectives specified by the state department of education.

4. The exams cover all or almost all students. Exams for 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. With MCEs, in contrast, virtually all students are affected, and the school system as a whole must accept responsibility for how students do on the exams. A single exam taken by all is not essential. Many nations allow students to choose which subjects to be examined in and offer high- and intermediate-level exams in the same subject.

5. The exams are controlled by the education authority that establishes the curriculum for and funds K-12 education. Curriculum reform is facilitated because coordinated changes in instruction and

exams are feasible. Tests established and mandated by other organizations serve the interests of other masters. America's premier high-stakes exams—the SAT-I and the ACT—serve the needs of colleges to sort students by aptitude, *not* the needs of schools to reward students who have learned what high schools are trying to teach.

Curriculum-based external exit exam systems are distinguished from MCEs by the following additional features:

6. The system signals multiple levels of achievement in the subject.

If only a pass/fail signal is generated by an exam, and passing is necessary to graduate, the standard will almost inevitably be set low enough to allow almost everyone to pass after multiple tries. The great bulk of students will easily pass the test and will have no incentive to strive to do better. CBEEEs, in contrast, signal the student's achievement level in the subject being tested, so that all students, not just those at the bottom of the class, have an incentive to study hard in order to do well on the exam. Consequently, a CBEEE should be more likely to improve classroom culture than an MCE. 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 decentralization, perfect information not only raises GDP, but also social welfare" (1994, p. 970).

7. The system assesses more difficult material. Since CBEEEs are supposed to measure and signal the full range of achievement in a subject, they contain more difficult questions and problems. This induces teachers to spend more time on cognitively demanding skills and topics. MCEs, by contrast, are designed to identify which students have failed to surpass a rather low minimum standard, so they do not ask questions or set problems that students near that borderline are unlikely to be able to answer or solve.² This tends to result in excessive class time being devoted to practicing low-level skills.

8. The system is a collection of end-of-course exams. Since CBEEEs assess the content of a specific course, teachers of the course (or course sequence) being tested inevitably will feel responsible for how well their students do on the exam. Grades on EOC exams may be made part of the overall course grade, further integrating the external exam into the

² In 1996, only 4 of the 17 states with MCEs targeted their graduation exams at a tenth-grade proficiency level or higher. Failure rates for students taking the test for the first time varied a great deal: from highs of 46 percent in Texas, 34 percent in Virginia, 30 percent in Tennessee, and 27 percent in New Jersey to a low of 7 percent for Mississippi. However, since students can take the tests multiple times, eventual pass rates for the class of 1995 were much higher: 98 percent in Louisiana, Maryland, New York, North Carolina, and Ohio; 96 percent in Nevada and New Jersey; 91 percent in Texas; and 83 percent in Georgia (American Federation of Teachers 1996).

classroom culture. Alignment between instruction and assessment is maximized, and accountability is enhanced. Proponents argue that teachers will not only want to set higher standards, but will also find their students more attentive in class and more likely to complete demanding homework assignments. Teachers become coaches helping their team battle the state exam.

Those who are skeptical about the value of introducing CBEEEs point out that American students already take a lot of standardized tests. Why aren't the tests students already take (such as the ACT, the SAT-I, or commercially prepared norm-referenced achievement tests) sufficient? What's so special about the new CBEEEs that some states are introducing in their standards-based reforms?

Norm-referenced achievement tests such as the CAT, CTBS, ITBS, ITED, and Terra Nova are not curriculum-based external exit exams because they fail criteria one and eight. Students have no stake in doing well on these tests. They are not part of a course grade or important to the student in some other way, so many high school students fail to put much effort into answering all the questions correctly and completely.³ Where stakes are not attached to results, teachers and school administrators experience the consequences, rather than individual students. In most of the nation, tests that students have no reason to try hard on are the primary indicator of student achievement in school accountability systems. When this is the case, school ratings may reflect the school's success in getting students to try hard on state tests and not the quality of instruction throughout the school year. This reduces the validity of high school tests as measures of true student achievement and makes their use in school accountability systems problematic.

The SAT-I test is not a CBEEE because it does not fulfill criteria three, five, and eight. It fails to assess most of the material—history, science, economics, civics, literature, foreign languages, and the ability to write an essay—that high school students are expected to learn. The Scholastic Aptitude Test (SAT) was designed from the beginning to minimize

³ This observation is based on interviews with the directors of the testing and accountability divisions in Manitoba and New Brunswick. It is also based on the large increases in student performance that occurred in New Brunswick, Massachusetts, Michigan, and other states when no-stakes tests became moderate- or high-stakes tests (Hayward 2001). Experimental studies confirm the observation. In Candace Brooks-Cooper's master's thesis (1993), a test containing complex and cognitively demanding items from the NAEP history and literature tests and the adult literacy test was given to high school students recruited to stay after school by the promise of a \$10.00 payment for taking the test. Students were randomly assigned to rooms. Students in one room were promised a payment of \$1.00 for every correct answer greater than 65 percent correct. This group did significantly better than the other students, who were told different test-taking conditions, including the standard "try your best" condition. Similar results were obtained in other well-designed studies conducted by the National Center for Research on Evaluation, Standards and Student Testing (see Kiplinger and Linn 1993 and O'Neil et al. 1997).

backwash effects on teaching and student study habits. Indeed, when the machine-scored, multiple-choice SAT replaced the curriculum-based essay-style College Board Examinations, Harvard College's admissions director Richard Gummere was very candid about why the SAT had been adopted: "Learning in itself has ceased to be the main factor [in college admissions]. The aptitude of the pupil is now the leading consideration" (Gummere 1943, p. 5).

The subject-specific SAT-II achievement tests fail criteria one, four, and five. Stakes are very low—few colleges consider SAT-II results in admissions decisions—and few students take them. In 1982–83, only 6 percent of SAT-I test takers took a science SAT-II, and only 3 to 4 percent took one in history or a foreign language. Schools do not assume responsibility for preparing students for SAT-II tests.

The Advanced Placement (AP) examinations are the one exception to the generalization that the United States lacks a national CBEEE. The number of students taking AP examinations has been growing at a compound annual rate of 9 percent per year. In 1999, 686,000 students, about 11 percent of the nation's juniors and seniors, took at least one AP exam. Despite this success, however, 44 percent of high schools do not offer even one AP course, and many that do allow only a tiny minority of their students to take these courses (College Board 1999). Low participation means that AP exams fail criterion 5 and, consequently, are not a CBEEE system. They can, however, serve as a component of a larger system.

HOW ARE CBEEE SYSTEMS HYPOTHESIZED TO INCREASE ACHIEVEMENT?

Curriculum-based external exit exam systems fundamentally change the signaling of student achievement. In doing so, they transform the incentives faced by students, parents, teachers, and school administrators. CBEEE systems are, consequently, hypothesized to influence the resources made available to schools and the priorities of school administrators, teacher pedagogy, parental encouragement, and student effort.

Impact on Students

Curriculum-based external exit exam systems improve the signaling of academic achievement. As a result, colleges and employers are likely to give greater weight to academic achievement when they make admission and hiring decisions, so the rewards for learning should grow and become more visible. CBEEE systems also shift attention toward measures of absolute achievement and away from measures of relative achievement, such as rank in class and teacher grades. In doing so,

CBEEE systems ameliorate the problem of peer pressure against studying.

How serious a problem is peer pressure against studying? Steinberg, Brown, and Dornbusch's 1996 study of nine high schools in California and Wisconsin suggests that academic excellence is still not highly valued by peers in most schools:

The adolescent peer culture in America demeans academic success and scorns students who try to do well in school . . . less than 5 percent of all students are members of a high-achieving crowd that defines itself mainly on the basis of academic excellence. . . . Of all the crowds, the "brains" were the least happy with who they are—nearly half wished they were in a different crowd (pp. 16, 145–6).

Why do so many "brains" want to get out of their crowd? Don Merten's 1996 ethnography of Cronkite Junior High School provides a rich and perceptive description of why this is so. Documenting the thoughts and actions of the ostracized and the popular students, he describes the transformation of one student from outcast to socially acceptable classmate. His description of the student's journey from nerd to cool kid is a gripping illustration of the power of peer norms in middle school. In order to fit in, the student cast away the norms and values he had lived by in elementary school and had defended in seventh grade: empathy, helping others, being good. He adopted instead the more predatory anti-teacher persona promoted by the dominant/popular students in junior high school.

Unfortunately, the peer pressure against studying or excelling in school found in Cronkite Junior High School is not an aberration. In the Educational Excellence Alliance survey, 24 percent of students said, "My friends make fun of people who try to do real well in school." Fifty-six percent said, "My friends joke around and annoy the teacher."

The teachers and principals of many American middle schools have lost normative hegemony. In the eyes of most students, the "brains" exemplify the "I trust my teacher to help me learn" attitude that prevails in most elementary school classrooms. The dominant middle school crowd is saying that trusting teachers is baby stuff. It's "us" versus "them." Withdraw from alliances with teachers, they say, and get with the program of becoming popular with peers. Be like us, the popular crowds say. Spend your time socializing. Do not study too hard. Value classmates for their athletic prowess and their attractiveness, not their interest in history or their accomplishments in science.

Why are studious students treated as outcasts? In part, it is because exams are graded on a curve. When exams are graded on a curve or college admissions are based on rank in class, joint welfare is maximized if no one puts in extra effort. In the game that results, side payments

(friendship and respect) and punishments (ridicule, harassment, and ostracism) enforce the cooperative “don’t study” solution.

If, by contrast, students are evaluated relative to an outside standard, as they would be with CBEEs, they no longer have a personal interest in getting teachers off track or persuading each other to refrain from studying. There is less incentive for them to engage in peer pressure that demeans studiousness.

Impact on School Administrators

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 workload. 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 class rank, GPA, and aptitude tests—not externally assessed achievement in secondary school courses—upgraded standards will not improve the college admission prospects of next year’s graduates. Graduates will probably do better in difficult college courses and will be more likely to get a degree, but that benefit is uncertain and far in the future. Maybe over time, the school’s reputation and, with it, the college admission prospects of graduates will improve because the current graduates are more successful in local colleges. That, however, is even more uncertain and postponed. Publishing data on the proportions of students meeting targets on standardized tests probably speeds the process by which real improvements in a school’s performance influence its local reputation. However, 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 over which teachers and administrators have little control.

American employers historically have paid little attention to student achievement in high school or school reputations when selecting young workers (Bishop 1990, 1992 and Hollenbeck and Smith 1984). Those that do pay attention to achievement use indicators of relative performance such as GPA and class rank rather than results on an external exam as a hiring criterion. Consequently, higher standards do not benefit students as a group, so parents as a group have little incentive to lobby strongly for higher teacher salaries, higher standards, and higher school taxes.

External exams transform the signaling environment. Hiring better teachers and improving the school’s science laboratories now yield a

visible payoff—more students passing the external exams and being admitted to top colleges. This in turn is likely to lead to more spending on schools, more rigorous hiring standards for secondary school teachers, and a higher priority assigned to student learning in the allocation of school budgets.

Additionally, reform-minded administrators can use CBEEE results to inspire teachers to raise standards for all students. The superintendent of a suburban New York district that has been nationally recognized for raising student achievement levels observes: “[External validators like Regents exams] were the best and only way in which we could get teachers and staff to see themselves as others might see them and not just keep looking in the mirror and seeing themselves as they would like to see themselves” (author’s interview with a superintendent of an all-Regents high school, August 1997).

Impact on Teachers

Curriculum-based external exit exams often have profound effects on teacher-student relationships and on the nature of the student peer culture. Teachers who have taught in environments with and without CBEEEs, as I have, sense the difference. When a proposal was put forward in Ireland to drop the nation’s system of external assessments and have teachers assess students for certification purposes, the union representing Ireland’s secondary school teachers reacted as follows:

Major strengths of the Irish educational system have been:

- (i) the pastoral contribution of teachers in relation to their pupils, and
- (ii) the perception of the teacher by the pupil as an advocate in terms of nationally certified examinations rather than as a judge.

The introduction of school-based assessment by the pupil’s own teacher for certification purposes would undermine those two roles, to the detriment of all concerned. . . . The role of the teacher as judge rather than advocate may lead to legal accountability in terms of marks awarded for certification purposes. This would automatically result in a distancing between the teacher, the pupil, and the parent. It also opens the door to possible distortion of the results in response either to parental pressure or to pressure emanating from competition among local schools for pupils (Association of Secondary Teachers of Ireland 1990, p. 1).

Note how the Irish teachers feared that switching entirely to internal assessment would result in their being pressured to lower standards. For American teachers, such pressure is a daily reality. Thirty percent of American teachers say they “feel pressure to give higher grades than students’ work deserves,” and they “feel pressure to reduce the difficulty and amount of work you assign” (Peter D. Hart Research Associates 1995, p. 9). Under a system of external exams, teachers and local school

administrators lose the option of lowering standards to reduce failure rates and raise self-esteem. The only response open to them is to demand more of their students so as to maximize their chances of being successful on the external exams.

A further benefit of CBEEEs is the professional development that teachers receive when they are brought to centralized locations to grade the extended answer portions of examinations. In May 1996, I interviewed a number of teachers and union activists about the examination system in Alberta. Even though the union and these teachers opposed the exams, they universally shared the sentiment that serving on grading committees was "a wonderful professional development activity."⁴ 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.

On the other hand, many fear that external exams will have a negative effect on teaching. Opponents argue that "preparation for high-stakes tests often emphasizes rote memorization and cramming of students and drill and practice teaching methods" and that "some kinds of teaching to the test permit students to do well in examinations without recourse to higher levels of cognitive activity" (Madeus 1991, pp. 7–8).

CBEEE advocates counter by challenging the assumption implicit in the above argument that examinations developed by the committees of teachers working for state departments of education are or will be worse than the tests developed by individual teachers. In fact, the tests that teachers develop for themselves are generally of very low quality. As John Thomas discussed at a 1991 conference, Fleming and Chambers's 1983 study of tests developed by high school teachers found that "80 percent of the items on teachers' tests were constructed to tap the lowest of [Bloom's] taxonomic categories: knowledge (of terms, facts, or principles)" (Thomas, p. 14). Rowher and Thomas (1987) found that only 18 percent of history test items developed by junior high teachers and 14 percent of items developed by senior high teachers required the integration of ideas. College instructors, by contrast, required such integration in 99 percent of their test items. Secondary school teachers test low-level competencies because that is what they teach.

If care is taken in designing external exams, they can induce improvements in instructional practice. Sherman Tinkelman describes one such instance, based on his experience as New York state's assistant commissioner for examinations and scholarships:

⁴ Interview results are available from the author upon request.

For years our foreign language specialists went up and down the state beating the drums for curriculum reform in modern language teaching, for change in emphasis from formal grammar to conversation skills and reading skills. There was not very great impact until we introduced, after notice and with numerous sample exercises, oral comprehension and reading comprehension into our Regents examinations. Promptly thereafter, most schools adopted the new curricular objectives (1966, p. 12).

DO CBEEES INCREASE ACHIEVEMENT? A LOOK AT THE EVIDENCE

The hypothesis that curriculum-based external exit examination systems improve achievement can be tested by comparing nations, states, and provinces that do and do not have such systems. Here we examine five different data sets:

- science and mathematics achievement of eighth graders in 1995 and 1999 in the 50-nation Third International Math and Science Study (TIMSS);
- achievement of 14-year-olds in the Reading Literacy Study of the International Association for the Evaluation of Educational Achievement (IEA);
- science, mathematics, and reading literacy of 15-year-olds in the 2000 Program for International Student Assessment (PISA);
- science and mathematics scores of 13-year-olds in the International Assessment of Educational Progress (IAEP) for nine Canadian provinces; and
- SAT test results for New York state students compared with results for students in the rest of the United States.

The theory predicts that CBEEE systems influence administrators' decisions about school priorities, teachers' decisions about standards and pedagogy, and students' decisions about studying. Much of the ultimate impact of CBEEE systems on student achievement derives from the changes these systems induce in hiring decisions, school priorities, and teacher pedagogy. Bishop (1996) tested the effects of CBEEEs on most of these components using data on Canadian schools and students. In most of the analyses in the current paper, the units of observation are educational systems and the objective is to assess the total effect of CBEEE systems on student achievement. Total effects are estimated by a reduced-form model that controls for parental socioeconomic status, productivity, and national culture, not the endogenous administrator, teacher, and parent behaviors.

Third International Mathematics and Science Study (TIMSS)

TIMSS provides 1995 data for seventh and eighth graders for 40 countries. The TIMSS-Repeat study of eighth-grade achievement provides 1999 data for an additional 10 countries and a second measure of eighth-grade achievement for 25 countries. To determine which nations have curriculum-based external exit exams in secondary school, we reviewed comparative education studies, government documents, and education encyclopedias, and we interviewed education ministry officials, embassy personnel, and graduate students from those nations who were studying at Cornell University.⁵

The national school systems classified as having CBEEEs for both math and science in all parts of the country were Bulgaria, Czech Republic, Denmark, England, Finland, Hong Kong, Hungary, Indonesia, Iran, Ireland, Israel, Italy, Japan, Jordan, Korea, Lithuania, Malaysia, Morocco, the Netherlands, New Zealand, Poland, Russia, Scotland, Singapore, Slovak Republic, Slovenia, Taiwan, Thailand, Trinidad and Tobago, Tunisia, and Turkey. Three countries—France, Iceland, and Romania—had CBEEEs in mathematics but not in science. Four countries—Australia, Canada, Germany, and the United States—had CBEEEs in some states or provinces but not in others. Norway had regular exit examinations in mathematics, but exams only every few years in science. Latvia had an external examination system until very recently, so we gave it a 0.5 on the CBEEE variable. Sweden's unusual system of combining external assessment and teacher assessment was also assigned a 0.5. The countries classified as not having a CBEEE in either subject were Austria, Belgium (both Flemish- and French-speaking systems), Brazil, Chile, Colombia, Cyprus, Greece, Mexico, the Philippines, Portugal, Spain, Switzerland, and Venezuela.⁶

Countries with a CBEEE system in the subject tend to have higher TIMSS scores. Furthermore, achievement differentials across nations are

⁵ A bibliography of the documents and individuals consulted when making these classifications is available from the author upon request. 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. The Philippines, for example, is classified as having external exams by the TIMSS report, 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 anti-apartheid boycotts. Kuwait was excluded because of the disruption of its education system by the Iraqi invasion and the Gulf War.

⁶ Following Madeus and Kellaghan (1991), the university entrance examinations in Greece, Portugal, Spain, and Cyprus, and the ACT and SAT in the United States were not considered to be CBEEEs. University entrance exams have much smaller incentive effects because students who are headed into work do not take them, and teachers can avoid responsibility for their students' exam results by arguing that not everyone is college material or that examiners have set an unreasonably high standard to limit enrollment in higher education.

Table 1
 Academic Achievement in Nations with and without Curriculum-Based External Exit Examination Systems
 TIMSS, TIMSS-Repeat, and IEA Reading Study Data

	Curriculum- Based External Exit Exam	Log GDP per capita, 1999	East Asia	Adjusted R ² RMSE	Number of Observations
TIMSS Science (U.S. GLE=26)					
8 th grade, 1995	51.1*** (11.5)	35.2*** (8.6)	6.2 (14.8)	.487 32.1	40
8 th grade, 1995–99	36.9*** (12.1)	57.8*** (8.0)	17.4 (14.1)	.521 36.6	50
TIMSS Math (U.S. GLE=24)					
8 th grade, 1995	42.3*** (13.3)	35.2*** (8.6)	54.2*** (16.4)	.484 36.3	40
8 th grade, 1995–99	35.1*** (13.0)	65.3*** (8.3)	49.8*** (14.4)	.586 37.9	50
IEA Reading (U.S. GLE=24)					
Age-Adjusted Average, 14- year-olds, 1990	26.5*** (8.1)	29.1*** (9.0)	-17.6* (11.8)	.610 16.7	25

Note: Numbers in parentheses are *t*-statistics. TIMSS is the Third International Math and Science Study. GLE is grade-level equivalent. IEA is the International Association for the Evaluation of Educational Achievement. On a two-tail test, *** indicates $p < 0.01$, ** indicates $p < 0.05$, and * indicates $p < 0.10$.

Source: Author's calculations using TIMSS, TIMSS-Repeat, and IEA Reading data. When test-score data were available for both 1995 and 1999, the dependent variable was the average of the two estimates. Gross domestic product per capita data are from World Bank (2001).

very large. According to the 1995 scores in science, Singapore, Korea, Bulgaria, and Flemish Belgium are more than one U.S. grade-level equivalent (GLE) ahead of the United States.⁷ Colombia, the Philippines, Lithuania, Romania, and Portugal are more than three GLEs behind. In mathematics, Singapore, Korea, Japan, and Hong Kong are four or more GLEs ahead of the United States, while Colombia, the Philippines, and Iran are more than three GLEs behind.

We regressed the mean eighth-grade science and mathematics test scores on 1999 per capita gross domestic product deflated by a purchasing power parity price index, a dummy for East Asian nations, and a dummy for CBEEEs. The results of the analysis of TIMSS-95 scores are presented in the first and third rows of Table 1. The results of the analysis

⁷ A grade-level equivalent is defined here as the difference between seventh- and eighth-grade TIMSS test-score means for U.S. students. Overall, in the 1995 TIMSS, U.S. students ranked 15th in science and 31st in mathematics.

of merged TIMSS-95 and TIMSS-Repeat data are presented in the second and fourth rows. Both analyses indicate that test scores are significantly higher in more developed nations, East Asian nations, and nations with a CBEEE in the subject. Nations with CBEEEs are about 1.5 U.S. GLEs higher on the math and science tests in the combined TIMSS-95 and TIMSS-Repeat data. The differential is even larger when only the TIMSS-95 data are analyzed. Since exams are also likely to influence learning during upper secondary school, the total effect at the end of twelfth grade is likely to be larger still.

International Association for the Evaluation of Educational Achievement (IEA) Reading Literacy Study

The IEA conducted a study of the reading literacy of 14-year-olds in 1990–91. The bottom row of Table 1 presents an analysis of IEA reading achievement data identical to the TIMSS analysis. The IEA study defined and measured three different 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. The exam variable is an average of the math and science CBEEE dummy variables used in the analysis of the TIMSS data. The IEA results are similar to the TIMSS results. Fourteen-year-old students in nations with CBEEE systems are about one U.S. grade-level equivalent better at reading than students in nations without CBEEE systems.

Program for International Student Assessment (PISA)

PISA is a new system of international assessment focusing on the reading, mathematics, and science literacy of 15-year-olds. Each participating country selected a nationally representative sample of approximately 4,000 15-year-olds. The students completed a 20- to 30-minute background questionnaire and a 90-minute assessment consisting of a mix of multiple choice, short answer, and extended response questions. PISA is a distinctive assessment tool: “While other studies, such as TIMSS and NAEP, have a strong link to curriculum frameworks and seek to measure students’ mastery of specific knowledge, skills, and concepts, PISA is designed to measure ‘literacy’ more broadly. PISA’s content is drawn from broad content areas, such as space and shape for mathematics, in contrast to more specific curriculum-based content such as geometry or algebra” (U.S. Department of Education 2001, p. 5).

Principals of schools where students took PISA assessments also completed a background questionnaire about their schools. PISA assesses the cumulative educational experiences of all students at age 15 regardless of their grade level or the type of institution they are attending. “By

Table 2
 Academic Achievement in Nations with and without Curriculum-Based External Exit Examination Systems
 Program for International Student Assessment 2000 Data

	Curriculum- Based External Exit Exam	Log GDP per capita, 1999	East Asia	Adjusted R ² RMSE	Number of Observations
PISA 2000, 15-year-olds					
Science	30.6*** (9.9)	46.5*** (9.1)	43.2** (17.3)	.630 22.9	29
Math	38.3*** (12.7)	62.5*** (11.6)	40.5* (22.1)	.620 29.3	29
Combined Reading Literacy	31.8*** (7.7)	51.8*** (6.7)	10.3 (12.8)	.737 16.9	29
Retrieving Information	42.4*** (9.5)	64.1*** (8.2)	12.9 (15.8)	.747 20.7	29
Expected Years of Schooling, Ages 5–65					
Sum of Net Enrollment Rates	.08 (.61)	3.34*** (4.8)	.75 (.91)	.611 1.58	37
Sum of FTE Net Enrollment Rates	−0.8 (.46)	2.98*** (.36)	.47 (.69)	.698 1.20	37

Note: FTE is full-time equivalent. Numbers in parentheses are *t*-statistics. On a two-tail test, *** indicates $p < 0.01$, ** indicates $p < 0.05$, and * indicates $p < 0.10$.

Source: Expected years of schooling data are from Organization for Economic Cooperation and Development 2001 (p. 133); part-time enrollment counts as 0.5 year in full-time equivalent figures. PISA data are from U.S. Department of Education (2001).

assessing students near the end of compulsory schooling in key knowledge and skills, PISA provides information about how well prepared students will be for their future lives as they approach an important transition point for education and work" (U.S. Department of Education 2001, p. 3).

The first four rows of Table 2 present an analysis of PISA data on science, mathematics, and reading literacy. As in the TIMSS analysis, scores are significantly higher in more developed nations, East Asian nations, and nations with a CBEEE in the subject. While grade-level equivalents cannot be calculated for the PISA tests, estimated impacts of CBEEEs appear to be comparable to those in TIMSS and IEA reading studies. The effect of a CBEEE system is similar in magnitude to a doubling of a nation's productivity and income per capita.

These results are consistent with the causal hypotheses presented above. Causation is not proved, however, because other explanations can no doubt be proposed. Other sources of variation in curriculum-based exams need to be analyzed. Best of all would be studies that hold national

culture constant; our final two data sets allow us to do this: the IAEP data for nine Canadian provinces, and SAT comparisons for New York state versus the other states.

Before turning to these last data sets, however, we can use the OECD data to see whether there is evidence that curriculum-based external exit exams tend to push students out of school. Many believe that a tradeoff exists between the standards and quality of an educational system and the number of students who can or will stay in school into their late teens and twenties. In the policy debate within the United States, concern has been expressed that high- or medium-stakes student accountability will increase dropout rates and reduce college attendance rates. We tested this hypothesis by calculating how many years youth in each of the OECD nations spend in school (we summed the net enrollment rates of people aged 5 to 65) and then assessing what impact CBEEEs have on these estimates of expected years of schooling. The results are presented in the fifth and sixth rows of Table 2. CBEEEs had no effect on expected years of schooling. The only variable that had a significant effect on how long young people typically stay in school was the nation's income.

International Assessment of Educational Progress (IAEP) for Nine Canadian Provinces

When the Educational Testing Service canvassed countries about participating in the 1991 IAEP, Canada decided to collect sufficient data to allow reliable comparisons among provinces and between the Anglophone and Francophone school systems of the five provinces with dual systems.⁸ At the time, 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 percent of that year's final grade in Alberta, Newfoundland, and Quebec and 40 percent in British Columbia. The other provinces did not have curriculum-based provincial external exit examinations in 1990–91. Ontario eliminated them in 1967, Manitoba in 1970, and Nova Scotia in 1972. 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 uninter-

⁸ All French-speaking schools in New Brunswick, Saskatchewan, and Manitoba were invited to participate. Stratified random samples of 105 to 128 secondary schools were selected from the French-speaking school systems of Ontario and Quebec and the English-speaking school systems in all provinces, with the exception of Prince Edward Island.

Table 3
Effects of Curriculum-Based External Exit Exams in Canada

	Hy- poth- e- sign	Mean	School Standard Deviation	Curriculum- Based Exam		French speaking	Religious School Board	Log Books in Home	Adjusted R ²
	(1)	(2)	(3)	Coeff.	t-stat.	(6)	(7)	(8)	(9)
Achievement									
Mathematics	+	.470	.135	.051	(7.6)	.074***	-.048***	.145***	.329
Science	+	.541	.096	.026	(5.1)	.021***	-.036***	.116***	.323
Discipline Problems	0/+	.765	.720	-.017	(.4)	.19***	-.132**	-.282***	.080
Absenteeism Problems	0/+	.822	.766	.140	(3.1)	-.16**	.001	-.411***	.131
School Administrator Behavior									
Math Specialist Teachers	+	.45	.50	.18	(6.9)	.08**	-.195***	.074**	.280
Science Specialist Teachers	+	.46	.50	.15	(5.6)	-.03	-.103***	.141***	.279
Took Math Courses in University	+	.64	.39	.19	(7.0)	-.06*	-.120***	.067**	.127
Took Science Courses in University	+	.69	.38	.19	(8.5)	-.21***	-.172***	.047	.199
Math Class Hours	+	3.98	.88	.33	(5.9)	.31***	-.057	-.254***	.124
Science Class Hours	+	2.93	.79	.16	(3.5)	-.06	-.365***	-.006	.132
Computers per Student	?	.051	.043	.001	(.6)	-.006*	-.009***	.004	.195
Specialized Science Labs	+	1.95	.95	.28	(5.6)	.043	-.097	.037	.274
Teacher Behavior									
Total Homework Hours per Week	+	4.41	1.62	.65	(6.9)	-.48***	.621***	-.146	.149
Math Homework Hours per Week	+	1.66	.64	.21	(5.0)	-.08	.189***	.017	.051
Science Homework Hours per Week	+	1.04	.47	.16	(5.1)	-.11**	.149***	.089**	.054
Math Quiz Index	+	1.62	.52	.10	(3.8)	.64***	-.107***	-.074**	.391
Science Quiz Index	+	.89	.38	.10	(4.9)	.32***	-.102***	-.007	.206
Home Behavior and Attitudes									
Average Hours of TV per Week in School	-	14.7	2.85	-.68	(4.2)	-1.7***	.63***	-2.69***	.255
Read for Fun Index	?	1.85	.28	.05	(2.8)	.08***	.028	.264***	.115
Watch Science Programs on TV	?	.97	.38	.06	(2.3)	.21***	.068**	-.090***	.091

(continued on next page)

Table 3 (continued)
Effects of Curriculum-Based External Exit Exams in Canada

	Hy-	Mean	School Standard Deviation	Curriculum- Based Exam		French speaking	Religious School Board	Log Books in Home	Adjusted R ²
	poth- e- sized sign			Coeff.	t-stat.				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Parents Talk about Math Class	+	.62	.17	.04	(3.4)	.02	.044***	.016	.046
Parents Talk about Science Class	+	.47	.17	.06	(5.2)	-.01	.007	.074***	.056
Parents Want Me to Do Well in Math	+	3.54	.22	.05	(3.1)	-.01	.093***	.084***	.104
Parents' Interest in Science (0-4)	+	2.18	.34	.06	(2.6)	.12***	.109***	.209***	.071
Science Useful in Everyday Life	+	2.46	.31	.06	(2.7)	.18***	.141***	-.097***	.095

Note: On a two-tail test, *** indicates $p < 0.01$, ** indicates $p < 0.05$, and * indicates $p < 0.10$. Controls also included mean number of siblings, the proportion of the students who use a different language at home, the number of students in a grade, and dummy variables for independent schools, religious schools, K-11 schools, and schools including 4th grade.

Source: Author's regressions predicting the characteristics of 1,309 to 1,338 Canadian secondary schools.

ested in exam scores. Job application forms do not ask applicants to report exam scores or grades.

The principals of schools sampled by IAEP completed questionnaires describing school policies, school resources, and the qualifications of eighth-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 schoolwork, and teaching methods of teachers.

The effects of curriculum-based provincial exit exams taken by twelfth graders on the achievement and behavior of Canadian 13-year-olds, their parents, teachers, and school administrators were examined by estimating models predicting these behaviors using schools as observations. The data set comprises 1,338 Canadian schools. The model uses 11 explanatory 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 religious school board, independent secular schools, independent nonsecular schools, schools

with primary grades, schools that include all grades in one building, and French-speaking schools; and a dummy for province exam.

Altogether, regression analysis was performed for four achievement outcomes, 12 measures of school administrator behavior, eight teacher behaviors, and 11 student/parent attitudes and behaviors. Table 3 presents the results for each achievement measure and a representative subset of the other variables of interest.⁹ The first column presents the hypothesized sign of the relationship between CBEEE systems and that variable. The means and standard deviations across schools of each dependent variable are presented in columns two and three. The coefficient for the CBEEE dummy variable and its t-statistic are presented in columns four and five. The R^2 corrected for degrees of freedom is reported in the last column.

Provincial exit exams had large positive effects on student achievement: 19 percent of a U.S. standard deviation (about four-fifths of a U.S. grade-level equivalent) in mathematics and 13 percent of a standard deviation (about one-half of a grade-level equivalent) in science.

Exit exams also influenced the behavior of parents, students, teachers, and school administrators in Canadian provinces. Schools in exit-exam provinces scheduled significantly more hours of math and science instruction, assigned more homework, had better science labs, were significantly more likely to use specialist teachers for math and science, and were more likely to hire math and science teachers who had studied the subject in college. Eighth-grade teachers in exam provinces gave tests and quizzes more frequently. The following were not significantly affected by CBEEEs: hours in the school year, library books per student, class size, or teacher preparation time (results not shown).

Opponents of externally set curriculum-based examinations predict that they will cause students to avoid learning activities that do not enhance exam scores. This hypothesis was examined by seeing whether exam systems were associated with *less* reading for pleasure and *less* watching of science programs like NOVA and Nature. Neither of these relationships was found. Indeed, students in exam provinces spent significantly more time reading for pleasure and more time watching science programs on TV, while watching significantly less TV overall. Parents in these provinces were more likely to talk to their children about their math and science classes, and their children were more likely to report that their parents “are interested in science” or “want me to do well in math.”

Do CBEEEs skew teaching in undesirable ways? Apparently not. Students did more (not fewer) experiments in science class, and emphasis on computation using whole numbers—a skill that should be learned by

⁹ The remaining regression results are available from the author upon request.

the end of fifth grade—declined significantly (these results are not presented in the table). Apparently, teachers subjected to the subtle pressure of a provincial exam four years in the future adopt strategies that are conventionally viewed as “best practices,” not strategies designed to maximize scores on multiple-choice tests.

Students responded to the improved teaching by becoming more likely to report that science was “useful in everyday life.” The data provided no support for our hypothesis that CBEEEs would induce employers to pay greater attention to high school achievement. Students in exam provinces were not more likely to believe that math was important in getting a good job and were less likely to believe that science was important in job hunting (results not shown).

One possible skeptical response to these findings is to point out that the correlation between the exam and other outcomes may not be causal. Maybe the people of Alberta, British Columbia, Newfoundland, Quebec, and Francophone New Brunswick—the provinces with exam systems—place higher priority on education than do people in the rest of the nation. Maybe this trait also results in greater political support for examination systems. If so, we would expect schools in the exam provinces to be better than schools in other provinces along other dimensions, such as discipline and absenteeism, and not just by academic criteria.

Bishop (1996) predicts, to the contrary, that exam systems induce students and schools to *redirect* resources and attention to the learning and teaching of 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 third, fourth, and eleventh rows of Table 3. Contrary to the “provincial taste for education” hypothesis, principals in exam provinces had not purchased additional computers, did not report significantly fewer discipline problems, and were significantly more likely to report absenteeism problems.

Scholastic Aptitude Test (SAT) in New York State

In the early 1990s, New York state was the only state with a voluntary CBEEE system. In 1993, about 56 percent of ninth graders took the mathematics course 1 exam and, of these, 24 percent of those not taking Regents exams were typically in courses that were considerably less challenging than Regents-level courses. A system of minimum competency tests in specific subjects set a minimum standard for those not taking Regents courses but, as in other states, the passing standard was low.

New York’s students are more disadvantaged, more heavily minority, and more likely to be foreign-born than students in most other states. Among northern states, only Maryland, Delaware, and Illinois have a larger share of African-American pupils. Nationwide, only California has

Table 4
Determinants of Mean Total SAT-I Scores for States

	Basic Model	With Controls for Teacher-Pupil Ratio and Spending per Pupil
New York State Dummy	46** (2.7)	35* (2.0)
SAT Participation Rate	-68** (2.6)	-88*** (3.3)
Type of Test-Taking Population:		
Parents AA-BA+	370*** (6.4)	367*** (6.6)
Private School	60 (1.6)	69* (1.9)
Black	-135*** (3.2)	-113 (2.6)
Large School	-44* (1.8)	-36 (1.5)
Three or More Math Courses	85 (1.3)	45 (.7)
Three or More English Courses	-36 (.3)	-45 (.4)
R ²	.926	.933
RMSE	14.8	14.2

Note: Numbers in parentheses are *t*-statistics. On a two-tail test, *** indicates $p < 0.01$, ** indicates $p < 0.05$, and * indicates $p < 0.10$.

Source: Author's calculations.

a higher share of foreign-born population, and only California, Texas, Arizona, New Mexico, and Colorado have larger Hispanic population shares. In New York, literacy levels among adults are substantially below the national average (National Education Goals Panel 1993).

Consequently, when we compare student achievement levels, family background must be taken into account. Considering the high incidence of at-risk children, New York students do remarkably well. The proportion of students taking algebra, calculus, chemistry, and physics is generally above the national average. A larger proportion (9.4 percent) of New York's eleventh and twelfth graders are taking and passing AP exams in English, science, math, or history than any other state except Utah (National Education Goals Panel 1993).

Graham and Husted's (1993) analysis of SAT test scores in the 37 states with reasonably large test-taking populations found that New York state students did better than comparable students in other states. They did not, however, test the statistical significance of the New York state (NYS) effect and used an unusual log-log specification.

Table 4 presents the results of a linear regression predicting 1991

mean SAT-Math plus SAT-Verbal test scores for the 37 states for which data are available. With the exception of the dummy variable for New York state, all right-hand-side variables are proportions—generally the share of the test-taking population with the characteristic described. Clearly, New Yorkers do significantly better on the SAT than students of the same race and social background living in other states (row one).

When this model is estimated without the NYS dummy variable, New York has the largest positive residual in the sample. The next largest positive residual (Wisconsin's) is 87 percent of New York's residual. Illinois and Nevada have positive residuals that are about 58 percent of New York's value. Arizona, California, Colorado, Florida, New Mexico, Ohio, Rhode Island, Texas, and Washington have negative residuals greater than 10 points. Many of these states have large populations of Hispanics and recent immigrants, a trait that was not controlled for in the analysis. This makes New York's achievement all the more remarkable when one considers that Hispanics and immigrants are a large share of its schoolchildren.

For individuals, the summed SAT-Math plus SAT-Verbal has a standard deviation of approximately 200 points. Consequently, the differential between New York state's SAT mean and the prediction for New York (based on outcomes in the other 36 states) is about 20 percent of a standard deviation or about three-quarters of a grade-level equivalent.

Adding the teacher-pupil ratio and spending per pupil to the model reduces the NYS coefficient by 25 percent (column two). It remains significantly greater than zero, however. The teacher-pupil ratio has a significant positive effect on SAT scores. This suggests that heavy investment in K–12 schooling in New York state (possibly stimulated in part by the Regents exam system) may be one of the reasons why New York state students perform better than comparable students in other states.

The theory predicts that the existence of CBEEE systems will induce New York state to spend more on K–12 education and focus that spending on instruction. Indeed, New York's ratio of K–12 teacher salaries to college faculty salaries is significantly above average. New York teachers are also more likely to have master's degrees than are the teachers of any other state, except Connecticut and Indiana. New York ranks seventh in both teacher-pupil ratio and the ratio of per pupil spending to gross state product per capita (Bishop 1996).

Clearly, New York invests a great deal in its K–12 education system. If the cause of the high spending were a strong general commitment to education or legislative profligacy, we would expect spending to be high on both K–12 and higher education. This is not the case. New York is number one among the 50 states in the ratio of K–12 spending per pupil to higher education spending per college student.

The Regents exams have been low- to medium-stakes tests, not high-stakes tests. Exam grades counted for less than a quarter of the final grade in the course and influenced only the type of diploma received. Employers ignored exam results when making hiring decisions. Students were aware that they could avoid Regents courses and still go to college. Indeed some perceived an advantage to avoiding them:

My counselor wanted me to take Regents history, and I did for a while. But it was pretty hard, and the teacher moved fast. I switched to the other history, and I'm getting better grades. So my average will be better for college. Unless you are going to a college in the state, it doesn't really matter whether you get a Regents diploma (Ward 1994, p. 12).

Indeed, the small payoff to taking Regents exams may be one of the reasons why 40 to 50 percent of students elected to take watered-down local classes either to reduce their workload or to boost their GPA.

This has changed. In 1996, the Board of Regents announced that students entering ninth grade in 1996 had to take a new Regents English examination and pass it at the 55 percent level. The requirement to take and pass exams in five subjects applies to those entering ninth grade in 1999 or later. The English exam has become more challenging. The reading selections are longer and more difficult. The biggest change is that the exam is six hours rather than three, and students must write four long essays rather than two. One of the four essays asks for a response to two long literary passages that are presented to them for the first time. In January 2001, the prompt was:

Read the passages on the following pages [a memoir and an essay]. . . . Write a unified essay about the discovery of beauty. In your essay use ideas from *both* passages to establish a controlling idea about the discovery of beauty. Using evidence from *each* passage, develop your controlling idea and show how the author uses specific literary elements or techniques to convey that idea.

These prompts clearly call for deeper thinking about literature than the prompts used in past Regents exams. There is nothing rote or formulaic about teaching students how to handle essay questions like these. The pressures created by these exams are improving the teaching of literature and writing throughout the state. This is the true purpose of the Regents exam system.

CONCLUSIONS

Our review of the evidence suggests that the claims by advocates of standards-based reform that curriculum-based external exit examinations significantly increase student achievement are probably correct. Students

from countries with such systems outperform students from other countries at a comparable level of economic development on TIMSS-95, TIMSS-Repeat, PISA, and IEA reading studies. School enrollment rates are not reduced by CBEEE systems. Not only did students from Canadian provinces with such systems know more science and mathematics than students in other provinces, but they also watched less television and talked with their parents more about schoolwork.

Furthermore, schools in provinces with external exams were more likely to employ specialist teachers of mathematics and science; hire math and science teachers who had studied the subject in college; have high-quality science laboratories; schedule extra hours of math and science instruction; assign more homework in math, science, and other subjects; have students do or watch experiments in science class; and schedule frequent tests in math and science classes.

When student demography was held constant, New York state, the only U.S. state with a CBEEE system in the early 1990s, did significantly better than other states on the SAT-I test. The pressures created by these exams are improving the teaching of literature and writing throughout the state. This is the true purpose of curriculum-based external exit exam systems.

References

- American Federation of Teachers. 1995. *Setting Strong Standards: AFT's Criteria for Judging the Quality and Usefulness of Student Achievement Standards*. Washington, DC: American Federation of Teachers.
- . 1996. *Making Standards Matter: 1996*. Washington, DC: American Federation of Teachers.
- Association of Secondary Teachers of Ireland. 1990. Information Sheet opposing changes in Examination Systems.
- Beaton, Albert, et al. 1996. *Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.
- . 1996. *Science Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.
- Bishop, John. 1990. "Productivity Consequences of What Is Learned in High School." *Journal of Curriculum Studies* 22 (2): 101–26.
- . 1992. "The Impact of Academic Competencies on Wages, Unemployment and Job Performance." *Carnegie/Rochester Conference Series on Public Policy* 37 (December): 127–95.
- . 1996. "The Impact of Curriculum-Based External Examinations on School Priorities and Student Learning." *International Journal of Education Research* 23 (8): 653–752.
- Bishop, John, Joan Moriarty, and Ferran Mane. 1997. "Diplomas for Learning, Not Seat Time: The Effects of New York's Regents Examinations." Paper presented at the Regents Forum in Albany, New York (October).
- Brooks-Cooper, Candace. 1993. "The Effect of Financial Incentives on the Standardized Test Performance of High School Students." Cornell University, master's thesis (August).
- Cameron, Stephen V. and James J. Heckman. 1991. "The Nonequivalence of High School Equivalents." NBER Working Paper No. 3804 (August).
- Chubb, John and Terry Moe. 1990. *Politics, Markets, and America's Schools*. Washington, DC: Brookings Institution.

- College Board. 1999. "More Schools, Teachers and Students Accept the AP Challenge in 1998-99." August 31. New York: The College Board.
- Competitiveness Policy Council. 1993. *Reports of the Subcouncils*. March. Washington, DC: Competitiveness Policy Council.
- Costrell, Robert. 1994. "A Simple Model of Educational Standards." *American Economic Review* 84 (4): 956-71.
- Edwards, Virginia B. 1999. "Quality Counts '99: Rewarding Results, Punishing Failures." *Education Week* 18 (17): 87-93.
- Fleming, M. and B. Chambers. 1983. "Teacher-Made Tests: Windows on the Classroom." In *Testing in the Schools: New Directions for Testing and Measurement*. San Francisco: Jossey Bass.
- Graham, Amy and Thomas Husted. 1993. "Understanding State Variation in SAT Scores." *Economics of Education* 12 (3): 197-202.
- Gummere, Richard. 1943. "The Independent School and the Post War World." *Independent School Bulletin* 4 (April).
- Hayward, Ed. 2001. "Dramatic Improvement in MCAS scores." *Boston Herald* (October 16).
- Hollenbeck, Kevin and Bruce Smith. 1984. *Selecting Young Workers: The Influence of Applicants' Education and Skills on Employability Assessments by Employers*. Columbus, OH: National Center for Research in Vocational Education, Ohio State University.
- International Assessment of Educational Progress. 1992. *IAEP Technical Report*. Vol. 1. Princeton, NJ: Educational Testing Service.
- Kang, Suk. 1985. "A Formal Model of School Reward Systems," in *Incentives, Learning, and Employability*, edited by J. Bishop. Columbus, OH: National Center for Research in Vocational Education, Ohio State University.
- Kiplinger, Vonda and Robert Linn. 1993. "Raising the Stakes of Test Administration: The Impact on Student Performance on NAEP." Center for the Study of Evaluation Technical Report 360. National Center for Research on Evaluation, Standards, and Student Testing, University of California, Los Angeles.
- Lerner, Barbara. 1990. "Good News about American Education." *Commentary* 91 (3): 19-25.
- Madeus, George. 1991. "The Effects of Important Tests on Students: Implications for a National Examination or System of Examinations." Paper presented at the American Educational Research Association Invitational Conference on Accountability as a State Reform Instrument: Impact on Teaching, Learning, Minority Issues, and Incentives for Improvement, Washington, DC (June).
- Madeus, George and Thomas Kellaghan. 1991. "Examination Systems in the European Community: Implications for a National Examination System in the United States." Contractor Report for the Office of Technology Assessment, U.S. Congress, Washington, DC.
- Merten, Don. 1996. "Visibility and Vulnerability: Responses to Rejection by Nonaggressive Junior High School Boys." *Journal of Early Adolescence* 16 (1): 5-26.
- Mullis, Ina, et al. 1997. *Mathematics Achievement in the Primary School Years: IEA's Third International Mathematics and Science Study*. Chestnut Hill, MA: Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.
- National Education Goals Panel. 1993. *National Education Goals Report 1993*. Vol. 2. Washington, DC: Government Printing Office.
- . 1995. *Data for the National Education Goals Report: 1995*. Vol. 1. Washington, DC: Government Printing Office.
- O'Neil, Harold F., et al. 1997. "Final Report of Experimental Studies on Motivation and NAEP Test Performance." Center for the Study of Evaluation Technical Report 427. National Center for Research on Evaluation, Standards, and Student Testing, University of California, Los Angeles.
- Organization for Economic Cooperation and Development. Center for Educational Research and Innovation. 2001. *Education at a Glance 2001*. Paris: Organization for Economic Cooperation and Development.
- Peter D. Hart Research Associates. 1995. *Valuable Views: A Public Opinion Research Report on the Views of AFT Teachers on Professional Issues*. Washington, DC: American Federation of Teachers.

- Rohwer, William D. and John W. Thomas. 1987. "Domain Specific Knowledge, Cognitive Strategies, and Impediments to Educational Reform," in *Cognitive Strategy Research*, edited by M. Pressley. New York: Springer-Verlag.
- Steinberg, Laurence, Bradford Brown, and Sanford Dornbusch. 1996. *Beyond the Classroom*. New York: Simon and Schuster.
- Thomas, John W. 1991. "Expectations and Effort: Course Demands, Students' Study Practices, and Academic Achievement." Paper presented at the Office of Educational Research and Improvement Conference on Student Motivation.
- Tinkelman, Sherman N. 1966. "Regents Examinations in New York State after 100 Years." Albany, NY: The University of the State of New York and the State Education Department.
- U.S. Department of Education. National Center for Education Statistics. 1993. *Digest of Education Statistics: 1992*. Washington, DC.
- . 1996. *Pursuing Excellence: A Study of U.S. Eighth-Grade Mathematics and Science Teaching, Learning, Curriculum, and Achievement in International Context: Initial Findings From The Third International Mathematics and Science Study*. NCES 97-198. Washington, DC.
- . 2001. *Outcomes of Learning: Results From the 2000 Program for International Student Assessment of 15-Year-Olds in Reading, Mathematics, and Science Literacy*. NCES 2002-115, by M. Lemke, et al. Washington, DC.
- U.S. General Accounting Office. 1993. *Educational Testing: The Canadian Experience with Standards, Examinations, and Assessments*, by K. D. White. GAO/PEMD-93-11. Washington, DC.
- Ward, Deborah Hormell. 1994. "A Day in the Life." *New York Teacher* 25 (10): 10-12.
- World Bank. 2001. *The World Development Report 2000-2001: Attacking Poverty*. New York: Oxford University Press.

Appendix Table 1
Examples of End-of-Course (EOC) Examination Systems

State	Year Announced	Subjects (year first administered)	Score on Transcript	Part of Course Grade	Teachers Grade Exam	Honors Diploma Based on EOC Exam	Year Minimum Competency Exam (MCE) Begins	EOC Exam can substitute for MCE	Other Rewards for Student Achievement
NY	1865	English, Math, Biology, Chemistry, Physics, U.S. History, World History, Latin, Foreign Languages, Introduction to Occupations	Yes	Yes	Yes	Yes (40%)	1979	about 1992	In 1950s scholarships were based on Regents exams. Use in teacher assessment is a local option. Becomes primary high school graduation test after 2000–03.
NC	1984	Algebra I, Biology (1987); Algebra II, U.S. History (1988); Chemistry, Geometry (1989); English I, Physics, Social Studies (1990–91)	Yes	Most (25% after 2000)	Yes	2003	1980	No	State tests at earlier grades influence retention decisions.
CA	1983	Algebra I, Geometry (1987); U.S. History, Economics (1990); Biology, Chemistry (1991); Coordinated Science (1994); Writing (1996); Civics (1997); Literature, High School Math (1998); Physics, Spanish (1999)	Yes	No	No	Yes (1%)	2004	No	State tests at earlier grades influence retention decisions.
TX	1992	Biology (1995); Algebra I (1996); U.S. History, English (1999)	Yes	Most (required in future)	?	No	1987	2000	Scholarships based on course rigor and family income. State tests at earlier grades influence retention decisions.
TN	1992	Algebra I, Biology, English II (2001); Algebra II, Geometry, English I (2002); U.S. History, Chemistry, Physics (2003)	Yes	Yes	?	No	1985	2005	Becomes high school graduation test in 2005. Current honors diploma based on GPA.

Appendix Table 1 (continued)
Examples of End-of-Course (EOC) Examination Systems

State	Year Announced	Subjects (year first administered)	Score on Transcript	Part of Course Grade	Teachers Grade Exam	Honors Diploma Based on EOC Exam	Year Minimum Competency Exam (MCE) Begins	EOC Exam can substitute for MCE	Other Rewards for Student Achievement
MD	1995	English I, Civics, Algebra, Geometry, Biology (2001)	Yes	?	?	No	1983	2007	Becomes high school graduation test in 2007. Honors diploma based on rigorous courses and GPA since 1998.
MS	1994	Algebra, U.S. History (1997); Biology (1998)	?	?	?	No	1989	No	Merit Scholarship based on GPA and ACT scores.
VA	1996	English, Algebra I & II, Geometry, Earth Science, Biology, Chemistry, U.S. History, World History (1998)	Yes	Some	?	Yes	1981	2004	Becomes high school graduation test in 2004. State tests at earlier grades influence retention decisions.
OK	1999	English, U.S. History (2000); Math, Biology (2001)	Yes	No	No	No	None	No	State university and employers encouraged to use EOC exam in admission and hiring.
AR	1997	Math (1999); English (2002); Science, History (2004)	Yes	No	No	No	None	No	State tests at earlier grades influence retention decisions.

Source: Author's research using multiple reference materials.

Appendix Table 2
Examples of End-of-Grade (EOG) Examination Systems

State	Year Announced	Subjects (year first administered)	Score on Transcript	Part of Grade	Teachers Grade Exam	Honors Diploma Based on EOG Exam	Year High School Graduation Test Requirement Begins	Other Rewards for Student Achievement
OH	1987	12 th grade: Reading, Science, Math, Civics (1994–96)	Yes	No	No	Yes, in part	1994	\$500 scholarship based on EOG exam. Honors diploma based on rigorous courses, GPA, 12 th grade exams, or ACT.
CT	1991	10 th grade: English, Math, Science (1994)	Yes	No	No	Yes	None	
MI	1991	11 th grade: Math, Reading, Science, Writing (1997); Social Studies (1999)	Yes	No	No	Starts 1996, subject-by-subject	None	Beginning with 2000 graduates, \$2,500 scholarship awarded based on EOG exams.
PA	1991	11 th grade: Reading, Writing, Math (1999)	Some	No	No	Starts 2003	None	
OR	1991	10 th grade: English, Math (1996); Science (1999); Social Studies (2003)	Most; expect increase	Some	Teachers blind	Starts 2001 (proposed subject-by-subject)	None	Certification of Initial Mastery based on English and Math (2001), add Science (2002), Arts (2003), second language (2005), and Social Studies (2006).
IN	1993	10 th grade: English, Math (1997)	Most	No	No	No	2000	Graduation requirement also met by grade C or better in all Core 40 college prep courses or demonstrated 9 th grade achievement level. Honors diploma based on curriculum.
MA	1993	10 th grade: English, Math, Science (1998)	No, temporarily	No	No	Starts 2000	2003	As of March 2000, class of 2000 receive Certificate of Mastery based on EOG, AP, or SAT II scores.
IL	1997	11 th grade: Reading, Writing, Math, Science, Social Science (2001)	Yes	No	No	Starts 2002, subject-by-subject	None	
WI	1997	10 th grade: Reading, Writing, Math, Science, Social Science (2002)	Yes	No	No	No	2004	1997 high school graduation test legislation repealed in 1999; left to local districts.

Source: Author's research using multiple reference materials.