



Cornell University
ILR School

Cornell University ILR School
DigitalCommons@ILR

CAHRS Working Paper Series

Center for Advanced Human Resource Studies
(CAHRS)

4-1-2005

High School Exit Examinations: When Do Learning Effects Generalize?

John H. Bishop
Cornell University, jhb5@cornell.edu

Follow this and additional works at: <http://digitalcommons.ilr.cornell.edu/cahrswp>



Part of the [Educational Assessment, Evaluation, and Research Commons](#)

This Article is brought to you for free and open access by the Center for Advanced Human Resource Studies (CAHRS) at DigitalCommons@ILR. It has been accepted for inclusion in CAHRS Working Paper Series by an authorized administrator of DigitalCommons@ILR. For more information, please contact hlmdigital@cornell.edu.

High School Exit Examinations: When Do Learning Effects Generalize?

Abstract

This paper reviews international and domestic evidence on the effects of three types of high school exit exam systems: voluntary curriculum-based external exit exams, universal curriculum-based external exit exam systems and minimum competency tests that must be passed to receive a regular high school diploma. The nations and provinces that use Universal CBEEES (and typically teacher grades as well) to signal student achievement have significantly higher achievement levels and smaller differentials by family background than otherwise comparable jurisdictions that base high stakes decisions on voluntary college admissions tests and/or teacher grades. The introduction of Universal CBEEES in New York and North Carolina during the 1990s was associated with large increases in math achievement on NAEP tests. Research on MCTs and high school accountability tests is less conclusive because these systems are new and have only been implemented in one country. Cross-section studies using a comprehensive set of controls for family background have not found that students in MCT states score higher on audit tests like the NAEP that carry no stakes for the test taker. The analysis reported in table 1 tells us that the five states that introduced MCTs during the 1990s had significantly larger improvements on NAEP tests than states that made no change in their student accountability regime. The gains, however, are smaller than for the states introducing Universal CBEEES. New York and North Carolina. The most positive finding about MCTs is that students in MCT states earn significantly more during the first eight years after graduation than comparable students in other states suggesting that MCTs improve employer perceptions of the quality of the recent graduates of local high schools.

Keywords

high school, exit, examination, learn, international, curriculum-based, competency tests, test, Universal CBEEES, student, achievement, admission, admission

Disciplines

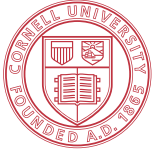
Education | Educational Assessment, Evaluation, and Research

Comments

Suggested Citation

Bishop, J. H. (2005). *High school exit examinations: When do learning effects generalize?* (CAHRS Working Paper #05-04). Ithaca, NY: Cornell University, School of Industrial and Labor Relations, Center for Advanced Human Resource Studies.

<http://digitalcommons.ilr.cornell.edu/cahrswp/4>



Cornell University
School of Industrial and Labor Relations
Center for Advanced Human Resource Studies

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

WORKING PAPER SERIES

High School Exit Examinations: When Do Learning Effects Generalize?

John H. Bishop

Working Paper 05-04



High School Exit Examinations: When Do Learning Effects Generalize?

John H. Bishop

Working Paper 05-04

<http://www.ilr.cornell.edu/cahrs>

This paper has not undergone formal review or approval of the faculty of the ILR School. It is intended to make results of Center research available to others interested in preliminary form to encourage discussion and suggestions.

Most (if not all) of the CAHRS Working Papers are available for reading at the Catherwood Library. For information on what's available link to the Cornell Library Catalog: <http://catalog.library.cornell.edu> if you wish.

Abstract

This paper reviews international and domestic evidence on the effects of three types of high school exit exam systems: voluntary curriculum-based external exit exams, universal curriculum-based external exit exam systems and minimum competency tests that must be passed to receive a regular high school diploma. The nations and provinces that use Universal CBEEES (and typically teacher grades as well) to signal student achievement have significantly higher achievement levels and smaller differentials by family background than otherwise comparable jurisdictions that base high stakes decisions on voluntary college admissions tests and/or teacher grades. The introduction of Universal CBEEES in New York and North Carolina during the 1990s was associated with large increases in math achievement on NAEP tests.

Research on MCTs and high school accountability tests is less conclusive because these systems are new and have only been implemented in one country. Cross-section studies using a comprehensive set of controls for family background have not found that students in MCT states score higher on audit tests like the NAEP that carry no stakes for the test taker. The analysis reported in table 1 tells us that the five states that introduced MCTs during the 1990s had significantly larger improvements on NAEP tests than states that made no change in their student accountability regime. The gains, however, are smaller than for the states introducing Universal CBEEES. New York and North Carolina. The most positive finding about MCTs is that students in MCT states earn significantly more during the first eight years after graduation than comparable students in other states suggesting that MCTs improve employer perceptions of the quality of the recent graduates of local high schools.

High School Exit Examinations: When Do Learning Effects Generalize?

High-stakes national exams, the SAT-1 and ACT, have been a rite of passage in America for more than a half a century. These college entrance exams are offered by two private organizations that are independent of the state education departments that set the curriculum and fund K-12 education.. Competitive pressures to keep unit costs low and attract customers—universities and students--from all fifty states prevent these tests from being comprehensive measures of learning during high school.¹ As Harvard's admissions director put it shortly after the college switched to the SAT-1, "*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)."

Most other nations have a very different approach to measuring academic achievement at the end of high school and signaling that information to universities and other interested parties. In Australia, Denmark, England, Scotland, Finland, France, Ireland, the Netherlands and many Canadian and German provinces, for example, high school exit examinations are developed by (or under the supervision of) the same Ministry of Education that establishes content standards for each subject, funds K-12 education and regulates the training and licensing of teachers. Taken over a period of two weeks or more, exams for each academic subject are about three hours long and require students to write essays, describe experiments and show how they solved multi-step problems. All students are typically required to take exams in a few core fields but students select the rest of their exam subjects. These universal curriculum-based external exit examination systems (Universal CBEEES) certify the learning of all students, not just those planning to go to university. The exams signal many different levels of achievement not just whether a student has exceeded a minimum standard. Doing poorly on these exams typically does not prevent one from graduating from secondary school; one completes high school with a record of **modest** accomplishment. Exam grades influence college admissions decisions just as the ACT and SAT-1 do in the U.S. They are requested on

job applications, appear on resumes and often influence hiring decisions. The questions used in the exams and the distribution of exam grades for each school are reported and discussed in local and national newspapers. The stakes are high for both students and schools. These examination systems are designed to simultaneously achieve four goals: induce teachers to set high standards, motivate students to learn what is being taught, recognize and reward them when they do and assist in the sorting of students across different post-secondary programs and employment options. The high stakes generated by the use of grades on these curriculum-based external exit exams in admissions and hiring decisions is what gives the examinations the leverage to achieve the other three goals of the system.

Students from these nations study harder than American students (Loveless, 2001) and did much better on TIMSS mathematics and science assessments at the end of upper-secondary school (Mullis et al, 1998; Takahira et al, 1998). The greater effort and achievement of the students in these nations may be due in part to the incentives created by their national exit exam systems. Blue ribbon panels studying American secondary education have frequently called for the introduction of high or moderate stakes curriculum-based achievement exams like those found abroad. In 1993, for example, the Competitiveness Policy Council proposed 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)."

Many states are pursuing a standard-based reform strategy that involves developing content standards for core academic subjects, administering tests assessing this content to **all** students, publishing individual school results and attempting in a variety of ways to hold schools and students accountable for student achievement. The most popular way of holding students accountable (twenty states in 2004-5) has been to require that they pass a **minimum competency test** (MCT) or a more difficult **standards-based exam** (SBE) to receive a regular high school diploma. This approach focuses pressure for higher standards on schools serving disadvantaged students and on students taking lower track courses.

A small group of states--New York, North Carolina, Virginia and a few other members of the Southern Regional Education Board---have developed or are developing systems of end-of-course exams that are intended to take over the accountability function currently served by the state's MCT. These systems are structured like the **Universal Curriculum-Based External Exit Exam Systems** found abroad. Compared to the European Universal CBEEES, however, student stakes tend to be lower because American colleges and employers seldom take exam grades into account when making admissions and hiring decisions.

The final type of high school exit exam system analyzed in this paper is **Voluntary Curriculum-Based External Exit Exam Systems** (Voluntary CBEEES). Advanced Placement and International Baccalaureate. Exams are growing rapidly and are of growing importance for admission to top colleges. For the first 125 years of their existence New York State Regents exams were voluntary. Despite their voluntary character, they were crucial in maintaining high standards in an ethnically diverse rapidly growing high school system.

The paper reviews empirical evidence on the effects of these three different approaches to assessing achievement and signaling it to students, parents, colleges, employers and the local community: It begins by describing the critical features of these approaches to signaling and accountability and how they contrast with the system of student stakes built around aptitude tests and teacher grades that prevailed during the 1960s and 1970s. Section 2 explains the theory of action behind the expectation that these new signaling/accountability systems will raise teacher standards and student effort and achievement above the levels that prevail when diplomas are based on seat time and high stakes college admission decisions are based largely on teacher grades and three-hour long multiple-choice format aptitude tests. The third section of the paper reviews the empirical literature on the effects of voluntary and universal curriculum-based external exit examinations and minimum competency examinations on learning, school attendance and labor market outcomes. The final section of the paper analyzes the impact of changes in state policies regarding signaling and accountability on gains in 8th grade NAEP mathematics achievement since 1990.

I. **Deconstructing Systems for Signaling Achievement and Holding Students Accountable**

Exactly how are domestic student accountability strategies similar to or different from the Universal CBEEES found abroad? We begin by noting the three features they all have in common. Minimum competency tests (MCTs), standards-based exams (SBEs), Universal CBEEES and Voluntary CBEEES all:

- 1. Produce signals of achievements that have real consequences for students and schools.** MCTs and SBEs are tests that must be passed to get a regular high school diploma. For CBEEES the nature and the magnitude of the rewards vary. In New York, North Carolina and Canada, CBEEE grades are averaged with teacher assessments to generate final grades for specific courses. In Europe and East Asia, exam results influence hiring decisions of employers and access to popular lines of study in university such as law and medicine. CBEEES sometimes make one eligible for a more prestigious diploma, a scholarship or confer rights to enroll in higher-level post-secondary institutions.
 - 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.²
 - 3. Are controlled by the education authority that establishes the curriculum for and funds K-12 education.** When a national or provincial ministry of education sponsors an external exam, it is more likely to be aligned with the national or provincial curriculum and to be used for school accountability; not just for student accountability. Curriculum reform is easier because changes in instruction and exams can be coordinated. The school system as a whole needs to accept responsibility for how all students do on the exams.
- The coverage of an exam system is the feature that distinguishes Voluntary CBEEES and college entrance exams such as the SAT-1 and ACT from MCTs, SBEs and Universal CBEEES.
- 4. MCTs, SBEs and Universal CBEEES are taken by almost every student.** Exams for a set of elite schools, advanced courses or college applicants may raise standards at the top

of the vertical curriculum, but are unlikely to have much effect on the rest of the students. Achievement gaps between high and low SES students are likely increase. Universality makes the school system responsible for how all students do on the exams. A single exam taken by all is not essential. Many nations (e.g. Netherlands, Germany, Ireland, Scotland, France and England) allow students to choose which subjects to offer for examination and offer high and intermediate level exams in the same subject.

Curriculum-based external exit exam systems are distinguished from MCTs and SBEs by the following additional features. CBEEES:

- 5. Assess a major portion of what students should 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 learning important material and developing valued skills. Some MCTs, SBEs, CBEEES and teacher exams do a better job of achieving this goal than others. External exams cannot assess every instructional objective, so teacher grades continue to have an important role.
- 6. Are collections of end-of-course exams?** End-of-course exams assess the content of specific courses (or sequences of courses). Assessment becomes better aligned with instruction, so teachers become more accountable. This also tends to align the interests of teachers, students and parents. Grades on end-of-course exams are often a part of the overall course grade further integrating the external exam into the classroom culture.
- 7. Signal 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 bulk of students who can easily pass the test are not induced to work harder (Becker and Rosen 1992; Costrell 1994; Betts and Costrell 2001). CBEEES signal achievement levels, so all students, not just those at the bottom of the class, have an incentive to study hard.³ Consequently, CBEEES should have a more pervasive effect on classroom culture than MCTs and SBEs.

- 8. Assess more difficult material.** Since CBEEES signal the full range of achievement in the subject, they contain more difficult questions and problems. This induces teachers to spend more time on cognitively demanding skills and topics. MCTs (and SBEs to a lesser extent) are designed to identify which students have failed to surpass a minimum standard, so they tend not to ask difficult questions.

America's college admissions tests—the SAT-1 and ACT—have some of these features: #1, #2, #7 and #8. However, they fail to satisfy four of the criteria defining Universal CBEEES. They are not controlled by the state education departments that fund and regulate K-12 education (#3), they are voluntary and so a sizable minority of students do not participate (#4), they cover only a small part of what students study during high school (#5) and they are not tied to specific courses and curricula (#6).

Now let's take a closer look at the three types of high school exit exams that have become increasingly important over the last 30 years.

Minimum Competency and Standards-Based Graduation Exams

Twenty-one states required students in the graduating class of 2005 to pass a series of tests before they were awarded a regular high school diploma (Quality Counts 2005. p. 91). According to a report of the Center on Education Policy (2004) seven states used minimum competency tests (MCTs) focusing on basic skills below the high school level and eleven states used "Standards-based Exams" (SBEs) that were aligned with state standards and generally targeted at the high school level. Half the states set no time limits for the tests. The other states gave students between 5 and 7 hours to complete the test battery. MCTs and SBEs raise standards, but not for everyone. The standards set by the teachers of honors classes and Advanced Placement classes are not changed by an MCT or a SBE. Students in these classes pass the MCT/SBE on the first try without special preparation. The higher standards are experienced by the students who are in the school's least challenging courses. School administrators want to avoid high failure rates, so they are expected to focus additional energy

and resources on raising standards in the early grades and improving the instruction received by struggling students. There is a danger, however, that teaching to such a test may narrow or 'dumb down' the curriculum for the majority of students who are not at risk of failing (Koretz et al 2001; Linn 2000, 2003).

Voluntary Curriculum-Based External Exit Examinations

Participation in the College Board's Advanced Placement program has been growing rapidly. In 2004 a total of 1,101,802 students (about 16 percent of high school seniors and smaller shares of 10th and 11th graders) took one or more Advanced Placement (AP) examinations (College Board 2005 p. 43).

The Regents of New York State have been sponsoring end-of-course exams ever since the 1870s. Panels of local teachers grade the exams using rubrics supplied by the state Board of Regents. Scores on the three hour exams appear on transcripts and are a final exam mark that is averaged with the teacher's quarterly grades to calculate the final course grade. Taking Regents exams was voluntary until late in the 1990s and nearly half of students took the easier 'local' courses intended for non-college bound students. Students taking a full schedule of Regents courses took about 10 Regents exams during their high school career.

Universal Curriculum-Based External Exit Examination Systems

In 1994 the New York City Board of Education decided that starting with those entering 9th grade in the fall of 1994, all students would have to take 3 Regents level math and 3 Regents level science courses before graduating. Two years later the State Board of Regents voted to require all students in the state to take Regents level courses in English, mathematics, American history, global history and a science and to pass the associated Regents exams. Ninth graders in 1996 were the first group to be affected. Ninth graders in 1999 had to pass all five exams. Supporting changes were also made in elementary and middle school curricula and school accountability tests.

North Carolina introduced end-of-course exams for Algebra 1 and 2, Geometry, Biology, Chemistry, Physics, Physical Science, American History, Social Science and English 1 between 1988 and 1991. Except for a four-year interlude in which some tests were made a local option, all students taking these courses were required to take the state tests. Test scores appear on the student's transcript and most teachers have been incorporating end-of-course exam scores in course grades. Starting in the year 2000, state law required the end-of-course exams to have at least a 25% weight in the final course grade (NC Assessment Brief, 1999-2002). A number of other states--Maryland, Mississippi, Oklahoma, Tennessee, Virginia etc.—are developing end-of-course examination systems.

Holding Secondary Schools Accountable

Formal systems for holding schools accountable are also growing in popularity. Forty-nine states publish school report cards and sixteen states have a formal mechanism for rewarding schools either for year-to-year gains in achievement test scores or for exceeding student achievement targets. Thirty-six states have special assistance programs to help failing schools turn themselves around. If improvements are not forthcoming, twenty-four states have the power to close down, take over or reconstitute failing schools (Quality Counts 2004, p.106-8). These tests typically carry low or no stakes for students but potentially moderate or high stakes for teachers and school administrators. The lack of real consequences is likely to result in many high school students not putting much effort into answering constructed response questions of tests that are not part of a course grade.⁴

II. Why are CBEEES and MCTs Hypothesized to Increase True Learning?

The purpose of the educational enterprise is learning. Engagement is essential to achieving this purpose. Students must come to school, pay attention, do homework, engage with the subject and construct their new knowledge in ways that allow them to retrieve it later. How are students induced to do all this hard work? Teachers try to make their subject interesting, but sixty-one percent of American students, nevertheless, say they “often feel bored” (OECD 2002 p. 330).

Additional motivators---diplomas, grades, exit exams, college admissions, etc.—are therefore essential. We assess each student’s learning, we honor it in ceremonies and signal (describe) it to parents, employers and colleges expecting them to reward the learning in their own way. The prospect of these external rewards strengthens incentives for students to attend school, participate in class and become engaged in learning. How these rewards manifest themselves also influences the priority that parents, school board members, teachers and administrators place on hiring better teachers, setting higher standards for students versus keeping school taxes low. Debates about MCTs and CBEEES are, at their root, debates about whether additional motivators are needed and how they should be structured.

The education leaders, politicians and policy analysts who support MCTs and/or CBEEES typically make the following arguments:

- Technical change and globalization have substantially increased the economic payoff to all types of academic and technical skills, so the current generation of students needs to achieve at higher levels than earlier generations.
- Many of the employers that offer good jobs have lost confidence in the high school diploma as a guarantor of literacy, numeracy and competence and, as a result, have become reluctant to hire recent high school graduates. States that force schools to set higher graduation standards by making the diploma contingent on passing a minimum competency test will raise achievement and help graduates get better jobs. This hypothesis is tested in the next section.
- The high stakes attached to the ACT and the old SAT-1-- tests that have little relationship to the high school curriculum--undermine incentives for students to develop high-level skills in history, science, foreign language, writing and English literature (Jencks and Crouse 1982; Board of Admissions and Relations with Schools 2002). CBEEES supporters argue that expanding AP and IB courses and introducing universal curriculum-based external exit exam systems will strengthen incentives to take rigorous courses and to study diligently (Mathews 2004, Ravitch 1995).

- Economists who have analyzed learning incentives agree that teacher grades are valuable motivators for students to try hard in class. The incentive effects of external assessments are different but complementary to the incentive effects of teacher grades (, Becker and Rosen 1992; Costrell 1994; Betts and Costrell 2001, Powell 1996). Both forms of evaluation should be used. (Coleman 1997, Board of Admissions and Relations with Schools 2002).
- Teacher assessments contribute a great deal to the valid assessment of student learning. Many important instructional goals can also be assessed externally and these assessments add a new and important perspective to the evaluation of learning. (Becker and Rosen 1992; Costrell 1994; Betts and Costrell 2001). Student evaluations that combine continuous and external assessment are more valid measures of learning than assessments based on only one methodology (Board of Admissions and Relations with Schools 2002).

The proposal is then that Universal CBEEES should supplement teacher grades, not replace them. When information from multiple sources is used, learning is measured more validly and the high stakes decisions that are based on the information should become better informed. Incentives to learn should strengthen. Opportunities for students and parents to game the system by seeking out easy graders or pressuring teachers to set lower learning standards should diminish (Competitiveness Policy Council 1993).

Figlio and Lucas (2001) have found that even though students learn substantially more when their teacher is a tough grader, parents do “not perceive tougher teachers to be better teachers (p. 20).” Difficult homework assignments intrude on parents’ time and often put the family under stress, so parents complain. This may be one of the reasons why 30 percent of American teachers feel pressured “to reduce the difficulty and amount of work you assign and “to give higher grades than students’ work deserves” (Hart 1995).” When the only signal of student achievement is teacher grades, parents seem to prefer high grades not high standards.

Teachers who work in systems with external exams are aware of their tendency to protect them from pressures to lower standards. When a proposal was tabled 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

(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 to either parental pressure or to pressure emanating from competition among local schools for pupils. (Association of Secondary Teachers of Ireland, Flyer, 1990, p. 1)

In the U.S. locally elected school boards and the administrators they hire make the thousands of decisions that determine academic expectations and program quality. Accountability advocates claim that when external assessment is absent, students and their parents benefit little in the near term from administrative decisions that opt for higher standards, more qualified teachers or a heavier student workload. The immediate consequences of such decisions are largely negative: higher local property taxes, more homework, having to repeat courses, lower GPA's, complaining parents and a higher risk of not graduating on time (Finn 1991, Ravitch 1995). Tests of some of these hypotheses are presented in section 3.

Opponents of external exams argue that focusing student attention on extrinsic rewards for learning will weaken student's intrinsic motivation to learn. George Madaus, for example, hypothesizes that "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 (1991b p. 7)." Madaus also points out that "preparation for high stakes tests often emphasizes rote memorization and cramming of students and drill and practice teaching methods" and that "some kinds of teaching to the test permits students to do well in examinations without recourse to higher levels of cognitive activity (1991 p. 7-8)." Some tests of these hypotheses are presented in the next section.

Advocates of external exams argue to the contrary that the end-of-course examinations developed by committees of experienced teachers are generally better than the teacher made final exams they replace.⁵ Proposed questions are carefully reviewed for ambiguity and bias and then pre-tested. The exams are published shortly after test day and receive another round of intense public scrutiny. States are trying to push teachers to give students better instruction in writing by adding externally set essay exams to their state testing programs. Well designed essay questions can also enliven class discussions and induce better teaching (see example in endnote).⁶

Steinberg, Brown and Dornbusch's (1996) recent study of nine high schools in California and Wisconsin concluded that:

The adolescent peer culture in [middle class] 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 (145-6).

James S. Coleman explains this phenomenon in the following way: "students who get especially high grades create negative externalities for other students, insofar as the teacher grades on the

curve....Often a norm arises in this case...; students impose a norm that restricts the amount of effort put into schoolwork (1990 p. 251).” External exams may be one way of changing student perceptions they are being graded on a curve. On external exams, everyone in the class can get an A. One is competing with anonymous students in other schools, not ones classmates. Peers should become less supportive of students who joke around in class or try to get the class off track and more supportive of those who cooperate with the teacher. Improved classroom culture should result in students learning more (Coleman et al 1997).

III. Do CBEEES and MCT/SBEs Increase True Learning?

Evidence from studies analyzing nationally representative data sets..

The web sites of most of the state education departments implementing high stakes testing systems report that growing numbers of students are reaching proficiency on the state’s MCT and school accountability tests. While flat or declining scores on a new high stakes test might reasonably lead one to conclude that true achievement has not improved, rising scores do not necessarily imply that true achievement has risen. Numerous authors (Linn 1990; Koretz et al 2001, Carnoy, Elmore and Siskin 2003) have pointed out that rising test scores might instead reflect teaching to the test—i.e. improved alignment of instruction with the topics and question formats found on the state’s high stakes test. Consequently, studies evaluating the effects of high-stakes testing on achievement must track their effects by studying scores on a zero-stakes audit test—e.g. TIMSS, PISA, National Assessment of Educational Progress (NAEP)—that represents a broader domain of knowledge than the content standards that informed the construction of the high stakes test.⁷ The issue is not whether the positive time trends on the state’s high stakes test are steeper than the trends on the audit test. That is almost guaranteed. The issue is “Do the audit tests respond to the introduction of a CBEEES or MCT?” and if so “How large is the response?” To assist the reader in judging whether estimated effects are substantively important, I have translated all impact estimates into a common metric: U.S. grade level equivalents.

Review of studies employing nationally representative data sets: The hypothesis that universal curriculum-based external exit examination systems improve achievement has been tested by comparing nations and provinces that do and do not have such systems. In most studies of impacts, national mean test scores (for an age group or a grade) were regressed on per capita gross domestic product deflated by a purchasing power parity price index, a dummy for East Asian nation and a dummy for Universal CBEEES. Analyzing 1994-95 Third International Math and Science Study (TIMSS) data, Bishop (1996, 1997) found that 13 year old students from countries with medium and high stakes Universal CBEEE systems outperformed students from other countries at a comparable level of economic development by 1.3 U.S. grade level equivalents (GLE) in science and by 1.0 GLE in mathematics. Analysis of data from the 1990-01 International Association for the Evaluation of Educational Achievement's study of the reading literacy of 14 year olds in 24 countries found that students in countries with Universal CBEEES were about 1.0 GLE ahead of students in nations that lacked a Universal CBEEES (Bishop 1999). Analysis of data from both waves of TIMSS data collection also implies that Universal CBEEES have highly significant effects (of about 1.5 GLEs) on the math and science achievement in 8th grade (Bishop 2003). Analyses of year 2000 data on 15 year olds from the Program for International Student Assessment (PISA) also yields large statistically significant estimated effects of CBEEES on reading, mathematics and science literacy of native-born students (Bishop 2003). Achievement gaps between high and low SES students are also significantly lower in nations that have a Universal CBEEES (Bishop and Mane 2004)

Four other studies (Ludger Wößmann (2000, 2003a, 2003b; Fuchs and Wößmann 2004) have conducted hierarchical analyses of the entire TIMSS and PISA micro data sets and included a comprehensive set of controls for family background, teacher characteristics, school resources and policies at the individual and school level. Wößmann (2000) found that 8th graders in Universal CBEEES nations were about 1.1 international grade level equivalents ahead in mathematics and about 0.8 international grade level equivalents ahead in science. He also found

that learning gains between 7th and 8th grade were significantly larger in Universal CBEEES nations.

Another five studies compare students living in different provinces/states in Germany, Canada and the United States. German provinces with centralized secondary school exit examinations have significantly higher scores on the TIMSS assessments (Jurges, H., Schneider, K. and Buchel, F., 2003). Students attending school in Canadian provinces with Universal CBEEES were a statistically significant one-half of a U.S grade level equivalent ahead in math and science of comparable students living in provinces without Universal CBEEES (Bishop 1997, 1999a). In 1990 New York State's Regents exam system was the only example of a voluntary curriculum-based external exit exam system in the United States. Graham and Husted's (1993) analysis of 1991 SAT test scores in the 37 states with reasonably large test taking populations found that New York State students did much better than students of the same race and social background in other states. Bishop, Moriarty and Mane (2000) confirmed Graham and Husted's SAT findings and also found that 1992 NAEP math scores of New York 8th graders were significantly higher than in other demographically similar states. Analyzing NELS-88 data Bishop, Mane, Moriarty and Bishop (2001) found that New York students learned about a half a GLE more between 8th grade and 12th grade than comparable students in other states. Controlling for ethnicity, social background and other standard's based reform policies, 8th graders in New York and North Carolina in 1996-98 were about one-half of a GLE ahead of comparable students in other states in reading, math and science. In these cross section analyses state minimum competency tests had small (less than 10 percent of a GLE) non-significant effects on achievement (Bishop, Mane, Moriarty and Bishop 2001, Jacob 2001).

What was the primary mechanism by which Universal CBEEES increase student achievement? The impacts of Universal CBEEES on school policies and instructional practices have been studied in the TIMSS data and in the Canadian International Assessment of Educational Progress data. Universal CBEEES were not associated with higher teacher-pupil ratios nor greater spending on K-12 education. They were, however, associated with higher

minimum standards for entry into the teaching profession, higher teacher salaries, a greater likelihood of having teachers specialize in teaching one subject in middle school and a greater likelihood of hiring teachers who have majored in the subject they will teach. Teacher satisfaction with their job was significantly lower, possibly because of the increased pressure for accountability that results from the existence of good signals of individual student achievement. Schools in Universal CBEEES jurisdictions devote significantly more hours to math and science instruction and build and equip better science labs. For homework time the Canadian and TIMSS studies got contradictory results (Bishop 1996, 1997, 1999b).

What about the quality of instruction and student attitudes toward the subject? Students in Universal CBEEES nations and Canadian provinces were significantly less likely to say that memorization is the way to learn the subject and significantly more likely to do experiments in science class. Quizzes and tests were significantly more common in Canadian CBEEES provinces, but in other respects these provinces were not significantly different on a variety of indicators of pedagogy. They were just as likely to enjoy the subject and they were significantly more likely to believe that science is useful in every day life and more likely to talk with their parents about school work. Students in the TIMSS study were significantly more likely to get tutoring assistance from teachers after school. Madaus's prediction that students would avoid opportunities to learn material that is not likely to be on the exam was not supported. Students in Canadian provinces with CBEEES spent significantly more time reading for fun and watching science documentaries (Bishop 1996). The study using TIMSS data found no relationship between CBEEES and reading for fun. (Bishop 1999b).

Do CBEEES Improve the Functioning of Decentralized Education Systems?

Advocates of external measurement of student achievement with important stakes attached argue that it will improve the functioning of decentralized education systems. Parents will be better able to judge which schools are doing a good job. The information will influence choice of school and strengthen competitive pressures for excellence. Ministries of Education no

longer need to try to improve education quality by rigidly specifying inputs—teacher qualifications, salaries, budget allocations and textbooks. Instead, teachers and school administrators can be given authority to use their local knowledge about teacher talents and budget circumstances to maximize school quality (Finn 1991, Ravitch 1995). Publishing data on achievement, it is hypothesized, creates accountability pressures that induce teachers and administrators to place greater emphasis on improving academic achievement. Tests of these hypotheses have been supportive. Bishop's (1999b) analysis of IAEP data found that controlling on student background, math achievement of students in private schools was higher only in the provinces that required externally set diploma exams at the end of secondary school. Analyzing TIMSS and PISA data, Ludger Wößmann (2002, 2003b,) found that school autonomy over salaries and teacher influence over course content, textbooks and budget allocations had positive effects on student achievement in nations with external exams. In nations without external exams, by contrast, high levels of school and teacher autonomy were associated with lower student achievement. This is a promising line of research. Since changes in school governance and autonomy are commonly proposed as a way to make schools more efficient, it is critical that we understand how the effects of school choice and autonomy are influenced by the measurement and signaling of student achievement.

Does Better Signaling of Achievement Influence School Attendance and Labor Market

Success?

What effects do high stakes curriculum-based external exit exam systems have on high school enrollment rates and college attendance? Many believe 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. Bishop and Mane (2004) recently evaluated the effects of Universal CBEEES on school enrollment rates of 15 to 19 year olds and of 20-24 year olds, upper-secondary graduation rates and years spent in school using Organization of Economic Cooperation and Development data. Universal CBEEES had no significant effect on any of these

indicators. The statistically significant predictors were per capita GDP and the share of upper-secondary students in pre-vocational and career-technical educational programs.

Well-controlled cross-section studies of aggregate state level data have concluded that enrollment rates and graduation rates are negatively related to the total number Carnegie units required to graduate. These studies found no tendency for aggregate completion rates to be significantly lower in states with MCT or SBEs (Bishop and Mane 2000; Lillard and DeCicca 2001, Jacobs 2001, Dee 2003). When, however, the analysis is conducted separately for schools in low income neighborhoods, for disadvantaged students or for students with below average grades in 8th grade, a number of studies have found to have found that high school enrollment and completion rates are lower in MCT/SBE states. (Reardon 1996, Bishop, Mane, Moriarty and Bishop 2001, Bishop and Mane 2004). The longitudinal NELS-88 data set allows a more refined look at the distributional effects of Voluntary CBEEEs and MCT/SBEs on high school completion. Students with low or average GPAs in 8th grade were significantly more likely to get their diploma late or to get a GED when they were from New York or a state with an MCT/SBE. The proportion of 8th graders who eventually got either a regular diploma or a GED was no different in New York but significantly lower for low GPA students from other MCT/SBE states (Bishop, Mane, Moriarty and Bishop 2001). As in Europe, fast paced instruction and high standards for getting an academic diploma results in some students taking longer to get the diploma and other students switching over to less demanding programs of study.

Critics of high stakes testing argue that teaching to exit exams diminish the time spent on more important skills that would help students in college and in jobs. If this were the case, we would expect students in states with graduation tests to be less likely to go to and stay in college and less likely to get good jobs. When this was tested, however, eighth graders in states with high school exit exams were found to be **more** likely to go to college and equally likely to graduate from college (Bishop, Mane, Moriarty and Bishop 2001; Bishop and Mane 2005).

Economic theory predicts that raising graduation standards will improve the average quality of high school graduates and raise their mean wage and earnings (Betts and Costrell 2001). Analysis of HSB and NELS-88 data support this prediction and contradicts claims to the contrary. Controlling on high school completion, college attendance and local labor market characteristics, students from states with MCTs earned significantly more--**9 percent more** in the calendar year following graduation-- than students from states without a MCT. The MCTs also helped recent graduates get jobs that offered better opportunities for training and advancement (Bishop, Mane, Moriarty and Bishop 2001). As a result, eight years after graduating from high school, those growing up in MCT states earned between \$1100 and \$2000 per year more than those who had attended high school in states without graduation exams (Bishop and Mane 2005). Diplomas that reflect both teacher judgments and external exams appear to be worth more in the labor market than diplomas awarded for seat time only or GED certificates based solely on test scores.

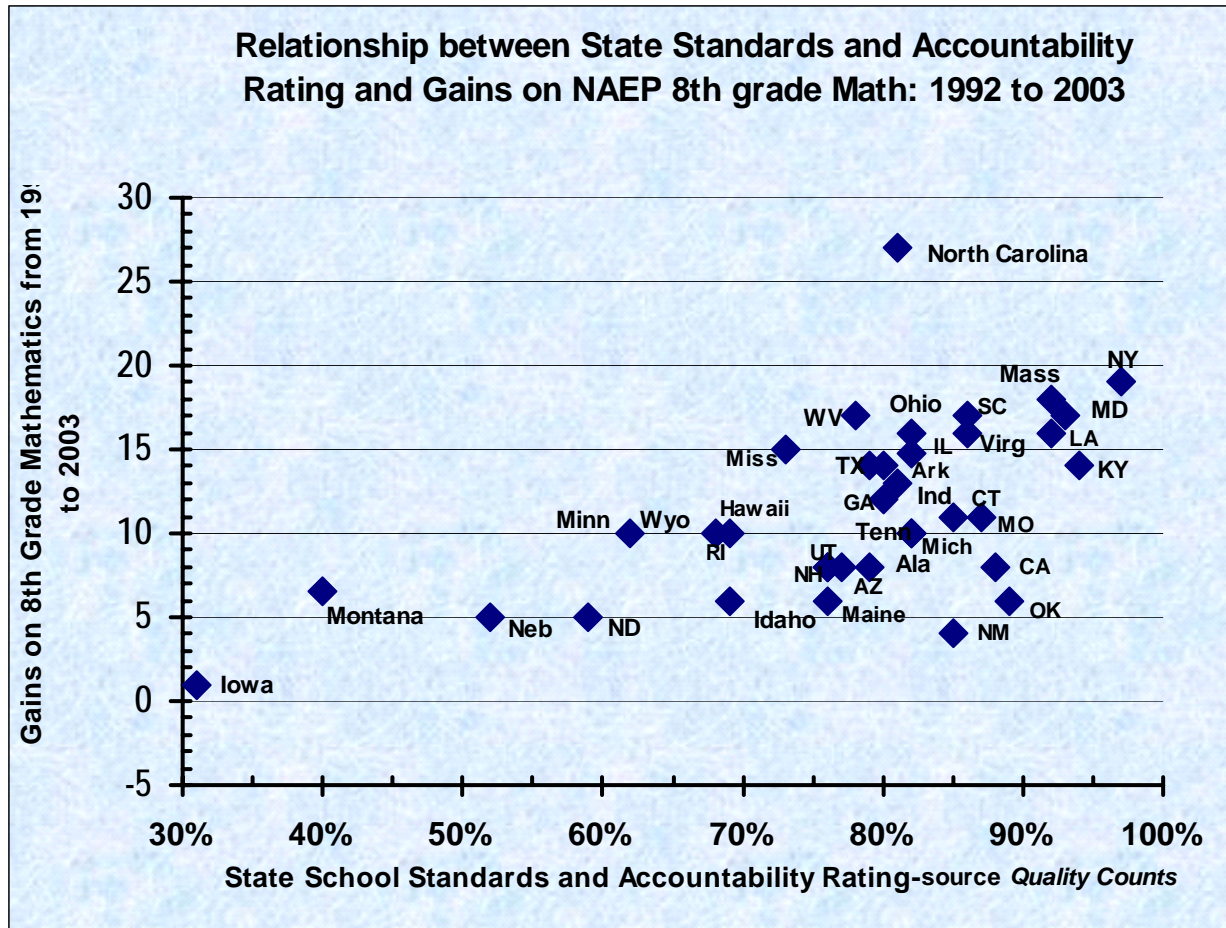
IV. Effects of Introducing High School Exit Exams on Achievement Gains since 1990

Another way to assess the effects of exit exams is to compare achievement gains in states that have recently introduced exit exams to gains in states that have not. The first such study was by Norman Fredrickson (1994). He found that states introducing “high stakes” testing systems (minimum competency tests for graduation for the most part) achieved larger gains on National Assessment of Educational Progress (NAEP) mathematics questions between 1978 and 1986.

Subsequent studies have all examined data from the 1990s a period during which many states were introducing standards-based reform strategies holding schools accountable for improving student achievement. The indexes of high-stakes testing used in these studies largely reflect the growth of school accountability testing systems not high school exit examinations. Martin Carnoy and Susanna Loeb (2003) found that 4th and 8th grade math achievement gains from 1996 to 2000 were significantly larger in states with strong test-based

accountability. Effects were particularly strong for Blacks and Hispanics and remained large when adjustments were made for changes in exclusion rates. Hanushek and Raymond (2003a, 2005) also report that states introducing test-based accountability tended to have larger test score gains from 4th to 8th grade. They also conclude special education placement rates did not rise more rapidly in states introducing test-based accountability. Barak Rosenshine (2003) excluded states with big increases in exclusion rates and then compared four-year NAEP test score gains of the remaining high-stakes states to the gains in states with no stakes. He concluded “that students in the clear high-stakes states were, indeed, learning mathematics and reading that was beyond the specific content of the statewide tests (p. 3).” Henry Braun (2004) study of gains between 1992 and 2000 concluded: “For each grade, when we examine the relative gains of states over the period, we find that the comparisons strongly favor the high-stakes testing states. Moreover, the results cannot be accounted for by differences between the two groups of states with respect to changes in the percent of students excluded from NAEP over the same period (p. 2).”

Figure 1 plots the relationship between gains on 8th grade NAEP math tests from 1992 to 2003 and Education Week’s January 2003 overall rating of the quality of each state’s standards and school accountability system. The picture tells the same story as the studies discussed above. School accountability systems were not well developed in 1992, so the positive relationship visible in Figure 1 suggests that the introduction and continuing development of school accountability systems during the 1990s may have had positive effects on NAEP math scores. Some of these states, however, were also introducing new student accountability systems during this period. During this period Hawaii ended its MCT, five states introduced new minimum competency tests and two introduced a universal CBEEES. Previous studies of impacts have either not distinguished between student and school stakes or analyzed one without controlling for the other. What effects did the new exit exams have and how do the impacts of student accountability compare with the impacts of school accountability?



Plans for new high school graduation tests and school accountability systems are announced many years in advance of actual implementation. These announcements start a change process that affects elementary as well as secondary school teachers. Consequently, it will typically be a “half a generation” (CEP, 2004 p. 26) before students exhibit the full effects of a new MCT, SBE or Universal CBEEES.⁸ This implies that statistical power is maximized by measuring change over a long period--one that runs at least from the announcement date to many years after the new graduation requirements are imposed.

Multivariate regressions were run to assess the effects of introducing school accountability and various types of student accountability on test score gains from 1992 to 2003 and from 1990 to 2003. Increases in the exclusion of students from testing tend to generate spurious increases in mean test scores that may bias efforts to evaluate high stakes testing. To avoid such a bias and to adjust for allowing testing accommodations in 2003, changes in the

exclusion of students from NAEP testing were included in the regressions. Results are presented in Table 1. The coefficients on the exclusion variables imply that an increase in exclusion rates removes from the NAEP sample students who tend to score about 68 points (5.6 GLEs) below the statewide average. This is a high but not implausible estimate of the size of the bias that results. When this variable is dropped (compare row 2 to row 3 or row 6 to row 7), coefficients on the school and student accountability variables hardly change at all. Carnoy and Loeb (2003) and Braun (2004) came to similar conclusions about the lack of an effect of changing exclusion rates on estimates of the effect of high stakes testing for school accountability.

Table 1--Effects of Introducing Student and School Accountability Systems on NAEP Mathematics Achievement in 8th Grade

Achievement Growth 1992→2003	Q.C. Index of Standards & Accountability	Universal Curriculum-Based Exit Exam	New H.S. Minimum Competency Grad Test [†]	Test Score in 1992 or 1990	Change in exclusions from 1990-92 to 2003	Adj. R Square # obs.
8 th Grade	.232*** (.055)				.722** (.285)	.377 41
8 th Grade	.187*** (.052)	7.44*** (2.50)	2.22+ (1.43)			.418 41
8 th Grade	.189*** (.057)	7.10*** (2.32)	2.19* (1.32)		.677** (.256)	.499 41
8 th Grade	.081* (.047)	7.23*** (2.84)	4.06*** (1.16)	-.238*** (.055)	.679*** (.209)	.666 41
Growth 1990→2003						
8 th Grade	.236*** (.059)				.876*** (.319)	.406 37
8 th Grade	.177*** (.050)	10.54*** (2.59)	3.08** (1.64)			.530 37
8 th Grade	.183*** (.048)	9.60*** (2.46)	2.44* (1.41)		.641** (.274)	.586 37
8 th Grade	.090+ (.055)	9.01*** (2.27)	3.00** (1.31)	-.199** (.075)	.689*** (.252)	.652 37
Independent Variables						
Mean	80.63	.045	.091	267	2.90	
Std. Dev.	11.11	.211	.362	8.95	2.38	

Author's analysis of state NAEP data. The dependent variable is the 2003 test score with accommodations allowed minus the 1992 or 1990 test score without accommodations. Data on 1992 was not available for Alaska, Illinois, Kansas, Montana, Nevada, South Dakota, Vermont and Washington, so they are not included in the estimations. New York and North Carolina introduced universal curriculum-based external exit exams during the 1990s. The Quality Counts Index of Standards and Accountability is the mean of the 2003 overall rating and an average of 1996 and 1997 Standards and Accountability ratings.

The Quality Counts school accountability index is a significant predictor of the growth of mathematics achievement. In models with no controls for student accountability, a two standard deviation (22 point) increase in the 1996-2003 School Accountability Index led to a 5 point gain in math achievement. The estimated effects of a two standard deviation increase in the index drops to 4.2 points when student accountability variables are included and drops to 1.8 points when 1992 test scores are also included in the model.

The five states-- Indiana, Massachusetts, Minnesota, Ohio and Virginia—that shifted from a no-student stakes environment to a MCT or SBE between 1992 and 2003 had significantly bigger gains in 8th grade mathematics than other states with similar QC school accountability ratings and changes in exclusion rates. On the NAEP mathematics test 12 points is roughly a grade level equivalent, so the predicted effect of introducing a MCT or SBE is between 18 and 34 percent of a grade level equivalent (GLE).

During the period from 1992 to 2003, the two states that added a Universal CBEEE system to an existing MCT system, North Carolina and New York, improved their math achievement by three-fifths of a GLE more than other states with similar Quality Counts school accountability ratings and changes in rates of exclusion. For the thirteen-year period from 1990 to 2003, the estimated effect of the Universal CBEEES is about three-quarters of a GLE.⁹ The predicted effect of a state shifting from no student accountability to a Universal CBEEES is the sum of the MCT/SBE and the Universal CBEEES coefficients or between .8 and 1.1 grade level equivalents.

V. Summary and Discussion

This paper has reviewed international and domestic evidence on the effects of three types of high school exit exam systems: voluntary curriculum-based external exit exams, universal curriculum-based external exit exam systems and minimum competency tests that must be passed to receive a regular high school diploma.

Voluntary CBEEES like AP, IB and California's Golden State Exams are particularly difficult to evaluate because they are taken by self-selected groups of students. Participation in New York's Regents exams was over 50 percent in the early 1990s, so the high achievement of New York students compared to similarly disadvantaged student populations in the early 1990s is suggestive evidence that Voluntary CBEEES have positive effects. However, there might be other reasons for the outstanding achievement of New York students, so more research on Voluntary CBEEES is needed before conclusions can be drawn.

Universal curriculum-based external exit exam systems are found all over the world, so many studies of their impacts have been conducted. The nations and provinces that use Universal CBEEES (and often teacher grades as well) to signal student achievement have significantly higher achievement levels and smaller differentials by family background than otherwise comparable jurisdictions that base high stakes decisions on voluntary college admissions tests and/or teacher grades. The introduction of Universal CBEEES in New York and North Carolina during the 1990s was associated with large increases in math achievement on NAEP tests.

Research on MCTs and high school accountability tests is less conclusive because these systems are new and have only been implemented in one country. Cross-section studies using a comprehensive set of controls for family background have not found that students in MCT states score higher on audit tests like the NAEP that carry no stakes for the test taker. The analysis reported in table 1 tells us that the five states that introduced MCTs during the 1990s had significantly larger improvements on NAEP tests than states that made no change in their student accountability regime. The gains, however, are smaller than for the states introducing Universal CBEEES, New York and North Carolina. The most positive finding about MCTs is that students in MCT states earn significantly more during the first eight years after graduation than comparable students in other states. This suggests that MCTs may be improving employer perceptions of the quality of the recent graduates of local high schools.

MCTs delay a significant number of graduations and reduce the number of standard high school diplomas awarded. They also increase the number GEDs and Certificates of Attendance awarded. Rates of school attendance and college completion remain essentially the same. European Universal CBEEES, by contrast, do not to depress school attendance rates or upper secondary completion rates.

Probably the most important finding of the paper is the remarkable ability of European style Universal CBEEES to substantially increase academic achievement without decreasing school enrollment and graduation rates. Minimum competency graduation requirements, by contrast, clearly have much smaller (possibly no) effects on achievement and also reduce the number of students getting a regular high school diploma.

Why are European Universal CBEEES so much more successful? First, they signal the full range of student achievement to universities and to employers, so all students get increased rewards-- better jobs and access to preferred university programs—if they study harder. An MCT, by contrast, focuses all of its high stakes incentive effects on a few low achieving students who were already at high risk of dropping out. Most students pass the MCT on the first try. Once they pass, the stimulus to studying and paying attention in class generated by the MCT goes away. Only in the minority of very troubled schools where the majority of students are at risk of failing the MCE is student culture likely to be changed by the high stakes test.

Second, doing poorly on a European Universal CBEEE means you graduate with a record of **modest accomplishment**. It does not prevent you from graduating altogether. Employers and universities take that record into account when they make their decisions. Students with poor exam grades are able to enter less prestigious forms of postsecondary education.

Thirdly, end-of-course exams pressure individual teachers to improve their teaching. Their colleagues will know how their students do on the exam. Since the stakes for the students are high, parents and school administrators are likely to encourage them to set high standards. MCTs, by contrast, typically cover material studied in many different courses taught by different

teachers. Sometimes they are administered in the fall. Under these circumstances, individual teachers are not considered responsible for how students taking their class this term do on the MCT test. **When everyone is responsible for student performance, no one is responsible.**

Fourthly, the component exams of these Universal CBEEES are more challenging and higher in quality than the MCT and SBE exams that dominate student accountability in the U.S. The challenge and quality of an exam depend on the level and complexity of the tasks students are required to perform, not the percent correct pass-fail cut score. The primary goal of any high or moderate stakes exam should be improving teaching and learning. Teachers should be proud to be preparing their students to take it. The long-term political viability of standards-based reform depends on our ability to improve quality and credibility of the exams used to measure student achievement.

Bibliography

- Amerein, A. L. and Berliner, D. C. (2002) "The Impact of high-stakes tests on student academic performance." Educational Policy Studies Laboratory, Arizona State University, <http://asu.edu/educ/epsl/EPRU/documents/EPSSL-0211-125-EPRU.pdf>.
- American Federation of Teachers. (1995) Setting Strong Standards: AFT's criteria for judging the quality and usefulness of student achievement standards. Washington, D.C.: American Federation of Teachers, 1-12.
- Association of Secondary Teachers of Ireland, Information Sheet opposing changes in Examination Systems, 1990.
- Becker, William and Sherwin Rosen. (1992) "The Learning Effect of Assessment and Evaluation in High School." Economics of Education Review, Vol. 11, # 2, pp.107-118.
- Betts, Julian and Robert Costrell. (2001) "Incentives and Equity under Standards-Based Reform" Brookings Papers on Education Policy 2001, edited by Diane Ravitch, (Washington, DC: The Brookings Institution), 9-74.
- Bishop, John H. (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 H. (1997) "The Effect of National Standards and Curriculum-Based External Exams on Student Achievement." American Economic Review, May 1997.
- Bishop John H.. (1999) "Are National Exit Examinations Important For Educational Efficiency?" Swedish Economic Policy Review, Vol. 6, #2, Fall 1999, 349-401.
- Bishop, John H. "Nerd Harrassment, Incentives, School Priorities and Learning," Earning and Learning, ed. by Susan Mayer and Paul Peterson, (Washington, DC: Brookings Institution Press, 1999b).
- Bishop John H., (2003) "What is the Appropriate Role of Student Achievement Standards," Education in the 21st Century: Meeting the Challenge of a Changing World, ed. Yolanda Kodrzycki, Boston, MA: Federal Reserve Bank of Boston.
- Bishop, John H., Joan Moriarty and Ferran Mane, (2000) "Diplomas for Learning: not Seat Time," Economics of Education Review. Vol 19, No. 3, 2000.
- Bishop, John H. and Ferran Mane. (2001) "The Impacts of Minimum Competency Exams Graduation Requirements on High School Graduation, College Attendance and Early Labor Market Success." Labour Economics, Vol. 8, n^o2, Spring 2001, 203-222.
- Bishop, John H., Ferran Mane, Michael Bishop and Joan Moriarty. (2001) "The Role of End-of-Course Exams and Minimum Competency tests in Standards-Based Reforms." Brookings Papers on Education Policy 2001, edited by Diane Ravitch, (Washington, DC: The Brookings Institution), 267-346.

- Bishop, John H. and Ferran Mane (2004) "Educational Reform and Disadvantaged Students: are they better off or worse off?" Presented at CES-IFO conference in Munich Germany, September 2004.
- Board of Admissions and Relations with Schools (2002) The Use of Admissions Tests by the University of California. University of California office of the President. <http://www.ucop.edu/news/sat/boars.pdf>
- Braun, Henry. (2004) "Reconsidering the impact of high stakes testing." *Educational Policy Analysis Archives*. Vol. 12, no. 1.
- Carnoy, Martin and Susanna Loeb. (2003) "Does External Accountability Affect Student Outcomes?" Educational Evaluation and Policy Analysis Vol. 24, # 4, 2003.
- Carnoy, Martin, Richard Elmore and Leslie Santee Siskin. (2003) *The New Accountability*. (New York: Routledge Farmer)
- Center on Education Policy, (2004) State High School Exit Exams: A Maturing Reform. (Washington, DC: CEP, August 2004), 1-259.
- Coleman, James S. (1990) *Foundations of Social Theory*. (Cambridge, Mass: Harvard University Press).
- Coleman, James, Barbara Schneider, Stephen Plank, Katheryn Schiller, Roger Shouse and Huayin Wang. (1997) Redesigning American Education. (Boulder Colo: Westview Press)
- College Board. (2005) Advanced Placement Report to the Nation 2005. http://www.collegeboard.com/prod_downloads/about/news_info/ap/2005/ap-report-nation.pdf
- Competitiveness Policy Council, (1993) Reports of the Subcouncils, March, Washington, D.C.: Competitiveness Policy Council.
- Costrell, Robert. (1994) "A Simple Model of Educational Standards," *American Economic Review*, vol. 84, no. 4, (1994) pp. 956-971.
- Dee, Thomas. (2003) "The 'First Wave' of Accountability." No Child Left Behind? The Politics and Practice of School Accountability. Eds. Martin West and Paul Peterson, Washington, DC: Brookings Institution Press, 2003. 292-323.
- Education Week staff. (2004, 2005) "Quality Counts." Education Week, January 8, 2004 and January 2005.
- Farber, Henry and Robert Gibbons. (1996), "Learning and Wage Dynamics." *Quarterly Journal of Economics*, pp. 1007-47.
- Figlio, David and Maurice Lucas. (2001) "Do High Grading Standards Affect Student Performance." Univ. of Florida, 1-38.
- Finn, Chester E. (1991) We Must Take Charge: Our Schools and our Future. (New York, The Free Press)

- Fleming, M. and Chambers, B. (1983) Teacher-made Tests: Windows on the Classroom. San Francisco: Jossey Bass.
- Fredericksen, Norman. "The Influence of Minimum Competency Tests on Teaching and Learning" (Princeton, New Jersey: Educational Testing Service, 1994).
- Fuchs, T. and Wößmann, L. (2004) "What Accounts for International Differences in Student Performance? A Re-examination using PISA Data." CESifo Working Paper 1235 (Munich, CESifo).
- Graham, Amy and Husted, Thomas. "Understanding State Variation in SAT Scores." *Economics of Education Review*, Vol 12, No. 3, 197-202.
- Gummere, Richard, (1943) "The Independent School and the Post War World." *Independent School Bulletin*. Vol 4, April, (quoted in Arthur Powell, Standards, 1995 Chap 6).
- Hanushek, Eric A. and Margaret Raymond. "Improving Educational Quality: How Best to Evaluate Our Schools?" Education in the 21st Century: Meeting the Challenge of a Changing World, ed. Yolanda Kodrzycki, 2003a, Boston, MA: Federal Reserve Bank of Boston.
- Hanushek, Eric A. and Margaret E. Raymond (2003b) "Shopping for Evidence Against School Accountability" Education Next, 3(3), Summer 2003 unabridged version of "High Stakes Research"
- Hanushek, Eric A. and Margaret E. Raymond (2005) Does School Accountability Lead to Improved Student Performance? "*Journal of Policy Analysis and Management* Spring 2005 (forthcoming)
- Hart (Peter D.) Research Associates, (1995) "Valuable Views: A public opinion research report on the views of AFT teachers on professional issues." Washington D.C.: American Federation of Teachers, 1995, 1-24.
- Jacobs, Brian A. (2001) "Getting Tough? The impact of high school graduation exams." *Educational Evaluation and Policy Analysis*. Vol. 23 no 2, Summer, 99-122.
- Jencks, Christopher & Crouse, James. "Aptitude vs Achievement: Should We Replace the SAT?" The Public Interest, 1982, pp. 21-35.
- Koretz, Daniel, Daniel McCaffrey and Laura Hamilton. (2001) "Toward a Framework for Validating Gains under High-Stakes Conditions." National Center for Research on Evaluation, Standards and Student Testing. CSE technical Report 551, December 2001, 1-45.
- Lillard, Dean and Phillip DeCicca. (2001) "Higher Standards, More Dropouts? Evidence within and across Time," *Economics of Education Review*, Vol 20 # 5, 459-73
- Linn, Robert. (2000) "Assessments and accountability." *Educational Researcher*, 29(2), 4-16. also at CSE Technical Report 490.
- Linn, R. L. (2003). "Accountability: Responsibility and Responsible Expectations." *Educational Researcher*. 32(7), 3-13.

- Loveless, Tom. (2001) *How Well are American Students Learning*, Brown Center Report on American Education 2001, Washington, DC, The Brookings Institution, p. 18-28, http://www.brookings.edu/dybdocroot/GS/brown/bc_report/BC_Report_hp.htm
- Madaus, George. (1991) "The Effects of Important Tests on Students: Implications for a National Examination or System of Examinations." American Educational Research Association Invitational Conference on Accountability as a State Reform Instrument, Washington, D.C., June 1991, 1-19.
- Mullis, I. V. S., Martin, M.O., Beaton, A. E., Gonzalez, E. J., Kelly, D. L. & Smith, T. A. (1998) Mathematics and Science Achievement in the Final Years of Secondary School: IEA's Third International Mathematics and Science Report. Available at: <http://isc.bc.edu/timss1995i/MathScienceC.html>
- Organization of Economic Co-operation and Development. (2002) Education at a Glance. Paris, France: Organization for Economic Co-operation and Development.
- Powell, Arthur. (1996) "Motivating Students to Learn: An American Dilemma" in Rewards and Reform. Susan Fuhrman and Jennifer O'day, Jossey Bass, 1-59.
- Ravitch, Diane. (1995) National Standards in American Education: A Citizens Guide (Washington DC: The Brookings Institution).
- Reardon, S. (1996) *Eighth grade minimum competency testing and early high school drop out patterns*. Paper presented at AERA meetings in New York.
- Rosenshine, B. (2003) "High Stakes Testing: Another Analysis." *Educational Policy Analysis Archives*. Vol. 11, no 4, from <http://epaa.asu.edu/epaa/v11n24/>
- Rohwer, W.D. and Thomas, J.W. (1987) "Domain specific knowledge, cognitive strategies, and impediments to educational reform." in M. Pressley (Ed.), Cognitive Strategy Research. New York: Springer-Verlag.
- Steinberg, Laurence, Brown, Bradford and Dornbusch, Sanford. (1996) Beyond the Classroom. New York: Simon and Schuster, 1-223.
- Takahira, Sayuri; Patrick Gonzales, Mary Frase and Laura Hersh Salganik. (1998) Pursuing Excellence: A Study of U.S. Twelfth-Grade Mathematics and Science Achievement in International Context Third International Math and Science Study, National Center for Educational Statistics, p. 92.
- Thomas, John W. (1991) "Expectations and Effort: Course Demands, Students Study Practices and Academic Achievement." paper presented at a 1991 Conference on Student Motivation sponsored by the Office of Educational Research and Improvement.
- Tinkelman, Sherman N. (1966) "Regents Examinations in New York State after 100 Years." Albany, N.Y: The University of the State of New York, The State Education Department, 1-15.
- Wößmann, Ludger (2000) "Schooling Resources, Educational Institutions, and Student Performance: The International Evidence," Kiel Working Paper No. 983, (May 2000) Kiel

Institute of World Economics, Germany, <<http://www.uni-kiel.de/ifw/pub/kap/2000/kap983.htm>> 1-88.

Wößmann, Ludger. (2003a) "Central Exit Exams and Student Achievement: International Evidence." No Child Left Behind? The Politics and Practice of School Accountability. Eds. Martin West and Paul Peterson, Washington, DC: Brookings Institution Press, 2003. 292-323.

Wößmann, Ludger. (2003b) "Central Exams as the "Currency" of School Systems: International Evidence on the Complementarity of School Autonomy and Central Exams." DICE Report - Journal for Institutional Comparisons 1 (4): 46-56, 2003.

Endnotes

¹ Most of the skills and knowledge—literature, economics, civics, foreign languages and writing—that high school students are expected to learn is not assessed by these tests. The ACT's science and history subtests are very short and not linked to specific curricula. The SAT-I does not assess history and science. Recently, the College Board added a 25 minute essay and 35 minutes of multiple-choice questions on grammar and usage to the SAT-1. SAT-2 exams assess individual subjects but fewer than sixty colleges require applicants to submit three or more SAT-2 scores. Few students take them. In 1982-83, for example, 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.

² Costrell's (1994a) analysis of the optimal setting of educational standards concluded that more centralized standard setting (state or national achievement exams) with a local option to set even higher standards results in higher standards, higher achievement and higher social welfare than decentralized standard setting (ie. teacher grading or schools setting their own graduation requirements).

³ Costrell's analysis of optimal standard setting concluded: "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)."

⁴ This observation is based on interviews with the directors of the testing and accountability divisions in Manitoba and New Brunswick Canada and the large increases in student performance that occurred in New Brunswick, Massachusetts, Michigan and other states when no-stakes tests become moderate or high-stakes tests. Experimental studies confirm the observation. In Candace Brooks-Cooper master's thesis, 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 a test. Students were randomly assigned to rooms and one group was promised a payment of \$1.00 for every correct answer greater than 65 percent correct. This group did significantly better than the students in the other test taking conditions, one of which was the standard try your best condition. Candace Brooks-Cooper, 1998.

⁵ Fleming and Chambers (1983) study of tests developed by high school teachers found that "over all grades, 80% of the items on teachers' tests were constructed to tap the lowest of [Bloom's] taxonomic categories, knowledge (of terms, facts or principles)"(Thomas 1991, 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.

⁶ Judge for yourself. New York's English Regents exam asks students to write four essays over a six hour period. One of the prompts always has the following "critical lens" format.

*Write a critical essay in which you discuss **two** works of literature you have read from the perspective of the statement that is provided to you in the 'Critical Lens.' In your essay, provide a valid interpretation of the statement as you have interpreted it, agree or disagree with the statement as you have interpreted it and support your opinion using specific references to appropriate literary elements from the two works. Guidelines....*

** Use the criteria suggested by the critical lens to analyze the works you have chosen*

** Avoid plot summary. Instead use specific references to appropriate literary elements (for example: theme, characterization, setting, point of view) to develop your analysis....*

** Follow the conventions of standard written English.*

In June 1999 the 'critical lens' was: "In literature, evil often triumphs but never conquers." In June 2000 it was: "It is not what an author says, but what he or she whispers that is important."

⁷ For a more extensive discussion of methodological issues surrounding estimating the effects of external examination systems see: Bishop 1996, Linn 2000, Koretz et al 2001, Hanushek and Raymond 2003 and Jacobs 2001.

⁸ This is one of the reasons why Amerein and Berliner's (2002) interrupted time series approach to measuring the effects of high stakes tests is flawed. The other weaknesses of the study are errors in dating the introduction of high stakes tests in many states, the use of national average scores as a comparison rather than states that did not implement high stakes testing and the arbitrary way of handling changing rates of exclusion. For a detailed critique see Hanushek and Raymond (2003). "Shopping for Evidence Against School Accountability" Education Next, 3(3), Summer 2003 unabridged version of "High Stakes Research."

⁹ Virginia had a MCT in the early 1980s but dropped it in favor of a 6th grade high school admission test introduced in 1990. The Standards of Learning were phased in at the end of the 1990s and became a Universal CBEEES with the graduating class of 2004. If Virginia is reclassified as a new Universal CBEEES, the coefficient on Universal CBEEES falls by about 20 percent and the coefficient on MCT/SBE is unchanged.