

POLICY BRIEF

The Impact of High-Stakes Exams on Students and Teachers

AUGUST 2004

The Impact of High-Stakes Exams on Students and Teachers:

A Policy Brief

INTRODUCTION

During the past several years, numerous states have developed mandatory high school exit programs in order to improve performance. Indeed, the movement towards high-stakes testing requirements before students can graduate from high school is one of the most prominent developments on the education policy landscape nationwide during the last decade. Furthermore, while this movement predates the recently passed federal education act known as No Child Left Behind (NCLB), legislative mandates announced with NCLB have served to give the high-stakes testing movement greater momentum and legitimacy. In particular, NCLB requirements of local educational agencies to use graduation rates as part of the determination of whether schools and school districts are making adequate yearly progress (AYP), has sharpened local accountability and insight.

In this policy brief we examine some of the consequences of this movement. More specifically, we attempt to highlight:

- The prevalence of high-stakes high school exit exams among the states;
- The impacts of high-stakes testing systems on both *students* (attitudes and motivation, dropout rates, and foremost, achievement); and
- *Teachers* (curriculum and instruction); as well as
- The *costs* of high-stakes' test implementation.

PREVALENCE OF HIGH SCHOOL EXIT EXAMINATIONS WITHIN THE UNITED STATES

Earlier we noted that the high-stakes testing movement appears to be gaining considerable momentum. As Table 1 shows, there are currently 19 U.S. states with mandatory high school exit exams as of 2003 (Center on Education Policy, 2003). These states are geographically displayed in the map in Figure 1, along with those four states that will be transitioning to high-stakes graduation/exit exams by 2008.

As Figure 1 makes clear, state participation in high-stakes, high school exams is dominated by states in the southern half of the country. As a result, children in states mandating the passage of a series of tests to graduate from high school are disproportionately black and Hispanic. They are also equally likely in states that have high levels of children in poverty. Table 2 highlights these important demographic

findings for those states now engaged in mandatory exit exams in contrast to their non-mandatory counterparts.

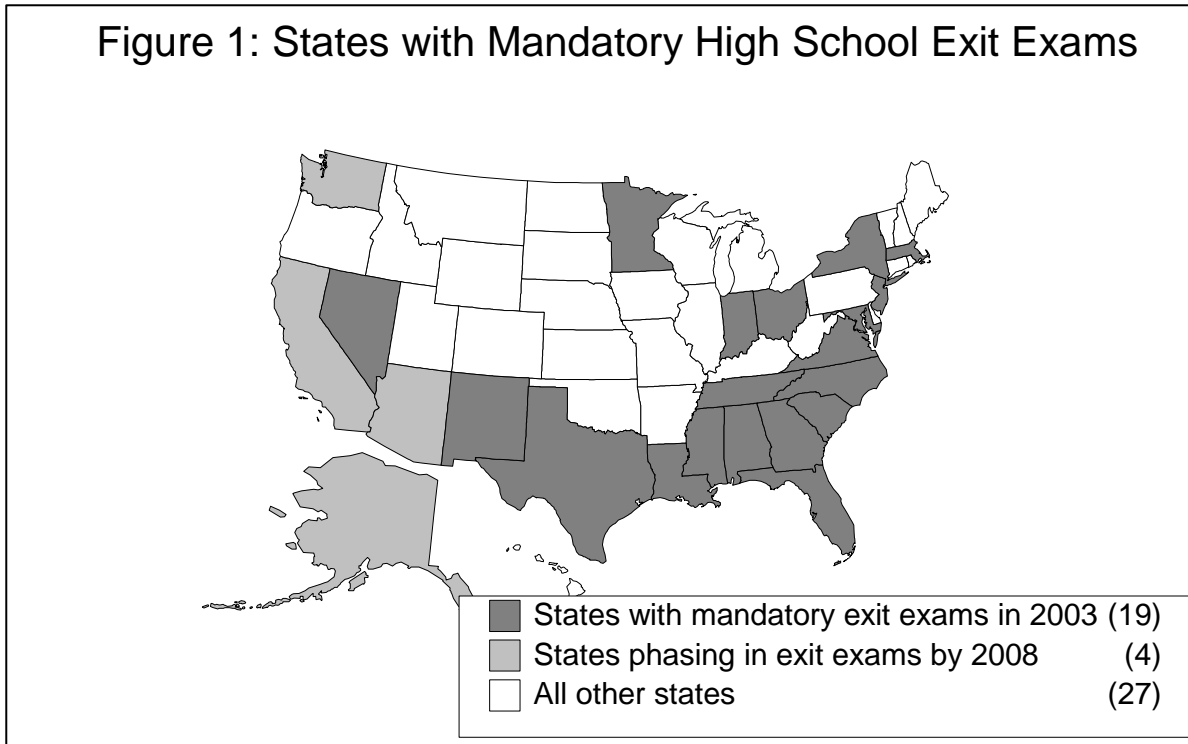


Table 1
States with Mandatory High School Exit Exams in 2003

Alabama	Louisiana	Mississippi	New York	Tennessee
Florida	Maryland	Nevada	North Carolina	Texas
Georgia	Massachusetts	New Jersey	Ohio	Virginia
Indiana	Minnesota	New Mexico	South Carolina	

Table 2:
School-age Child Poverty, Race/Ethnicity and Average Education Expenditure by States who Require High School Exit Exams and Those Who

	2002 Percentage of Children Aged 5-17 in Families w/ Income < Poverty	2002 Percentage of Population Black	2002 Percentage of Population Hispanic	2002 Average State K-12 Education Expenditure per Pupil
States w/ High School Exams	15.4%	17.8%	16.2%	\$8,135
States Without HS	15.4%	11.9%	14.0%	\$8,226

Sources: U.S. Census Bureau and National Education Association

IMPACT OF HIGH-STAKES EXAMS ON STUDENT ACADEMIC ACHIEVEMENT

The unambiguous targeting of high-stakes exams in high minority and poverty states makes perfectly good sense, assuming that the net benefits of such a policy outstrips its net costs. Evaluating the thesis that high-stakes tests have had an effect on achievement is a pivotal one since this is one of the critical “gap closing” assumptions on which the standards and high-stakes testing movement is based. The movement’s proponents believe that exams with consequences or stakes attached, such as the inability to graduate from high school, will provide an incentive to students to study and work hard in order to pass. Although the data are far from conclusive, the vast majority of studies have upheld a relationship such that we feel confident in concluding that the existence of **high-stakes exams do generally lead to higher student academic performance.**

Generally the research in this area attempts to demonstrate the effectiveness of state-level mandates not by comparing graduation rates or post K-12 outcomes (e.g., college attendance), but by examining elementary and middle grade performance. For example, Amrein and Berliner compare the test results on the National Assessment of Educational Progress (NAEP) of high-stakes test states (i.e., the treatment cohort) with those states without high school exit exams (the control group).¹

Their research received significant attention in the popular press with their finding that no clear overall relationship could be discerned between the existence of high-stakes exams and overall academic performance as measured by the NAEP (Amrein and Berliner, 2002). Several other researchers since then, while generally concurring with their choice of the NAEP to operationalize achievement, have nonetheless critiqued Amrein and Berliner’s methods² and using the same data have come to the opposite conclusion, that there is a statistically significant difference in the test results of high-

¹ The NAEP is an effective measure of academic performance for the following reasons: 1.) it is an exam that students generally don’t practice for and hence there is likely not to be a so called ‘training effect’, where competence is garnered by drilling repeatedly in the same items likely to be on the test (Braun, 2004); 2.) it is administered to a random sample of students Statewide, not just the best or college-bound students who could be expected to take the SAT, ACT or AP exams (Amrein and Berliner, 2002); 3.) it has been administered for many years; 4.) and across many subject or content areas; and 5.) since all but a few US states participate, it is a uniform indicator across states whose own state assessments may vary in content, rigor and testing approaches (e.g., open-ended vs. multiple choice items and criterion-referenced vs. norm-referenced tests). Finally another virtue of the NAEP as an assessment tool is that since the NAEP is designed to measure a broader spectrum of curriculum topics, rather than the standards a particular state or states may focus on, theoretically it should be a better measure of true learning or educational achievement.

² Like these other researchers we take issue with many of the research assumptions and methods employed by Amrein and Berliner. These criticisms are so serious as to cause us to doubt the validity of their findings. They include: the lack of any significance testing to ascertain if the differences in NAEP test scores between the control and experimental groups are meaningful and not the result of chance; needlessly selecting several states out of the sample to be analyzed; converting more robust continuous data (i.e. test scores) to nominal categories (i.e., ‘increased’, ‘decreased’), etc.

stakes and other states. Moreover, these researchers have found that the direction of this relationship is one which proponents of high-stakes exams would have hoped for: high-stakes test states have generally better--and significantly so, in statistical terms--academic performance.

Hanushek and Raymond in their analysis essentially “re-ran” Amrein and Berliner’s data by including states, they latter analysts needlessly excluded. When they compared the results of high-stakes and no-accountability states, they found higher NAEP test scores in high-stakes exam states. The average increase in 4th and 8th grade Math NAEP scores from 1992 to 2000 was 9% in testing states while it was only 4% in their non-testing counterparts. This 5% gain--more than double the rate of improvement--was statistically significant at the .05 level (Hanushek and Raymond, 2003). Rosenshine’s analysis which compared NAEP score change on the 4th and 8th grade math and reading exams between testing and non-testing states does not reveal the same magnitude of difference that Hanushek and Raymond’s does. Nevertheless, in every comparison, test score increase was greatest in testing states and overall the rate of improvement in the cohort of high-stakes states was roughly double that of the comparison group of non-testing states: 3.4% vs. 1.75% (Rosenshine, 2003).

Braun’s approach in his paper, *Reconsidering the Impact of High-stakes Testing*, is similar to that of Amrein and Berliner. Like Hanushek and Raymond, he re-analyzes the data by: including states Amrein and Berliner deleted; focuses on change over different years; and ‘normalizes’ the gains of states by accommodating standard errors in the analysis. Braun compared 4th and 8th grade NAEP scores from testing and non-testing states in terms of the relative gains over time. His findings strongly favor high-stakes tests in terms of academic performance (Braun, 2004).

Carnoy and Loeb take a somewhat different tack in their paper “*Does External Accountability Affect Student Outcomes?: A Cross-State Analysis*”. They focus on the period from 1996 to 2000 and with regard to NAEP scores, their concern is the proportion of students in grades 4 and 8 meeting the proficient standards. Moreover, they develop an accountability index independent variable, with each state assigned a score of 0 to 5 based on the estimated strength of their accountability regime. Other independent variables include a variety of political, demographic and educational variables. Their criterion of measurement is the change in percent of children meeting the proficiency standard over the period 1996 to 2000. They do not compare testing and non-testing states but rather fit a 50 state regression model. They find a relatively strong positive association between gains and the accountability index in grade 8, but a weaker, though still positive association in grade 4 (Carnoy and Loeb, 2003).

Another important set of research evidence is based on time-series data in a single state over time, while examining any changes in achievement after the establishment of a high-stakes regime. For an example of this type of more rigorous, quasi-experimental approach (an interrupted time-series analysis design), we can look to recent evidence from the state of Texas. Texas is a useful illustration for a number of reasons: it has received a great deal of publicity from both the popular press and the

education policy community; its testing system--the Texas Assessment of Academic Skills (TAAS)--has been cited as possibly contributing to improved student achievement; and President Bush made rigorous testing the centerpiece of his educational policy during his tenure as governor of that state.

Klein, Hamilton, McCaffrey and Stecher found that except for fourth grade math the gains by Texas students on the NAEP, although considerable, were comparable to those experienced by students nationwide during a period in the 1990s. Moreover, gains on the TAAS were several times greater than those on the NAEP. Therefore, how much students' proficiency in reading and math improved depends on whether the assessment is NAEP (based on national content standards) or TAAS (aligned to Texas state standards) scores. Accordingly, these researchers concluded that they were unable to prove that the high-stakes TAAS has had any significant effect on broader measures of learning as measured by the NAEP (Klein, Hamilton, McCaffrey & Stecher, 2000).

In addition to this analytical framework of comparing states with high-stakes tests with those who do not, some very important cross-national research on the same question has been conducted. Typically, these studies have compared the average test scores on international assessments of educational progress of nations who have curriculum based exit examinations with those who do not. Research using this analytical approach, which has generally been favored by John Bishop³ in his studies, indicates that nations with curriculum-based exit exams (CBEEs) have demonstrated greater achievement on standardized international tests than have their non-testing counterparts, after attempting to control for potentially confounding factors such as gross domestic product (GDP) and per capita spending on education. For example, Bishop found that on the Third International Mathematics and Science Study (1994), the existence of a curriculum-based high school exit exam was statistically significant in science (at the .01 level) and for math ($P=.08$). These effects indicate that nations with exit exams have average scores for 13 year olds that are equal in non-testing countries with children who are 1.3 grades higher in science and 1.0 grade higher in math. Bishop also analyzed the results of testing and non-testing nations in terms of scores on the International Assessment of Educational Progress, 1991. He found similar results insofar as nations with CBEEs had better scores on this cross-national measure. However, on this assessment the greater effect for CBEEs was associated with math scores: two grade level equivalents. In other words, the scores for 13 year olds in testing nations were as high as those students in non-testing countries two grades higher (Bishop, 1998).

Finally, researchers including Bishop have pointed out that the two states with the longest record of curriculum-based exit exams--New York and North Carolina--have historically performed higher than the rest of the nation after controlling for individual

³ Bishop has also taken another innovative approach in comparing the results of the International Assessment of Educational Progress (IAEP) for the Canadian provinces that have high school exit exams and those which do not. High school exit exam provinces did have statistically significant higher test scores than did their non-testing counterparts.

demographic, school and state characteristics. Bishop contends that New York was the only state in the nation with a CBEE in the early 1990s. Accordingly, in order to test the hypothesis of whether New York had any greater success on the NAEP, Bishop regressed several demographic variables including the percents of students in poverty, black or Hispanic or foreign-born and a dummy for NYS on the 8th grade math NAEP scores of the 41 states taking the exam in 1992. He found that although the other demographic variables generally had significant and strong impacts on test scores, New York's mean NAEP score was a statistically significant one grade level equivalent above that predicted by the model (Bishop, 1998).

The cross-sectional data then seem to support the contention that high-stakes exams result in greater performance, while the longitudinal studies are less persuasive. As an aside, cross-sectional approaches are the foundation of the average yearly progress calculation under NCLB. We think that they are to be favored also, because of the potential confounding effects of longitudinal research designs: usually high-stakes exams are part of a larger package of educational reforms that are usually simultaneous and which generally include establishing learning standards, consequences for teachers and/or schools, changes in instruction, etc. Thus, it is difficult to isolate the 'true' testing effects from these other factors.

STUDENT ATTITUDES AND MOTIVATION

While interest in the effects of high-stakes or mandated high school exit exams has been focused almost solely upon test performance and graduation outcomes, there are other outcomes of considerable interest and importance. Certainly an important outcome class concerns the impact of high-stakes testing upon student attitudes or motivations.

Much of the research on student response to high-stakes exams is actually based on teacher perceptions. For example, Smith and Rottenberg in their paper *Unintended Consequences of External Testing in Elementary Schools*, interviewed 19 teachers in 49 classes they observed in two low-income, mixed race/ethnicity schools, to garner their perceptions of how classrooms have changed since the implementation of high-stakes tests. They found that *teachers believed* that standardized tests (of any type-with or without stakes) cause younger students to experience 'stress ... burnout, fatigue, physical illness, misbehavior... and psychological distress' (Smith and Rottenberg, 1991). Moreover, Koretz et al., in their RAND Corporation study of teacher responses to Kentucky's implementation of a statewide exam system, found that one-third of teachers reported that student morale had deteriorated and none reported that it had improved (Koretz et al., 1996). And although most Kentucky principals found the testing system's requirements burdensome, most also thought that the benefits balanced or outweighed the costs. Moreover, although Kentucky teachers felt that the assessments increased expectations for all students, they were deemed to be of most benefit for high-achieving students. Finally, North Carolina teachers surveyed by the state's teacher association indicated higher levels of stress and lower morale among students after the introduction of high-stakes testing in that state. According to this

research, 63% of teachers and administrators reported that the focus on testing had added to students' level of stress (Hargrove, Jones and Jones, 2000).

A multi-state survey of teacher views of student behavior in response to high-stakes tests in three states was conducted by Clarke, Rhoades and others in 2003. These states were chosen because they represented low (Kansas), moderate (Michigan) and high (Massachusetts) responses to accountability needs based on a multi-criterion model. Interestingly, in all three states, that is, regardless of the level of stakes or consequences attached to an exam, interviewees reported more negative than positive consequences of testing, such as stress and too much testing (Clarke et al., 2003). For example, two-thirds of Massachusetts educators and two-fifths of those in Michigan--i.e., high and moderate stakes states, respectively--reported that their students were experiencing test-related stress, while only one-fifth of Kansas'--a low-stakes state--teachers reported this finding. Massachusetts' educators, who according to the authors' typology work under the constraints of a system of high-stakes or consequences for students, were most likely to note these negative effects.

The paradox in this study is this: while the Massachusetts study sample complained most heavily about the stress-related backwash effects, teachers in that state were also strongest in their assertions of the positive benefits of high-stakes exams: their focus on raising the quality of education and increasing student motivation to learn.

If as Clarke and Rhoades asserted, these high-stakes exams have beneficial, if stressful effects on teacher behavior, it is also true that students' views and attitudes are no doubt, affected by these processes. In many ways, they are the crucial actors in this process of educational reform. That is, it is their behavior that is being 'incentivized'. The proponents of the high-stakes movement believe that through the market mechanism of the reward of graduation and by extension, placement in the college of their choice (and it's corollary--the penalty of not graduating) students will be motivated to study harder, to learn more and to achieve greater cognitive gains than students not exposed to a high-stakes system. Massachusetts teachers in the example cited above, implicitly accept this market paradigm: it is their belief that the prospect of not graduating, as Massachusetts is a high-stakes state, has increased students' motivation to succeed. Nevertheless, another question worth asking is whether students also understand, or participate in expected ways in this market paradigm? Or, more importantly, does the research evidence provide any support for the view of those who oppose testing on pedagogical grounds. The latter for instance, argue that when rewards and sanctions are attached to performance, students lose their intrinsic interest in learning and cease to be independent, self-directed learners (Sheldon and Biddle, 1998). These and other countervailing arguments are listed in Table 3.

Positive Effects	Negative Effects
Provide students with clear info about their own skills	Frustrate students and discourage them from trying
Motivate students to work harder in school	Make students more competitive
Send clearer messages to students about what to study	Cause students to devalue grades and assessments
Help students associate & align personal efforts with rewards	

Source: Stecher, Hamilton and Klein, 2002.

The anecdotes of students themselves, as related by an Indiana teacher are illustrative and help to dispel the notion that a market model based on rewards and sanctions will work.⁴ For example, contrary to the market model’s assumptions:

- Virtually all (89 percent) students surveyed said that their parents were not worried they might fail;
- Students themselves did not value the test results;
- They thought it unfair that their accomplishments in terms of grade point average would be invalidated by a single high-stakes assessment; and
- The constant repetitious drilling in possible test items was frustrating for both poor and accomplished students (Hughes and Bailey, 2002).

Other studies came to very distinct conclusions. Roderick and Engel in their analysis of low performing students in Chicago found that children from disadvantaged backgrounds generally worked harder, which manifested itself in higher than average learning and promotion to the next grade level (Roderick and Engel, 2001). Nevertheless there is not a lot of information in the research literature on student perceptions of the effects of high-stakes tests on their motivations and attitudes.

THE IMPACT OF HIGH-STAKES TESTING ON DROPOUT RATES

If graduation is the ‘carrot’, then dropping out (or being retained in grade) is one of the ‘sticks’ in the implied incentive scheme that exists in a market framework approach to the phenomena of high-stakes testing.

Many of the same researchers who have analyzed the impact of high-stakes exams on achievement, discussed earlier, have also examined their impact on dropout rates. For example, Amrein and Berliner, using a time-series approach on 16 states with exit exams found that high-stakes tests increased dropout rates, retention in grade and

⁴ Indiana is an interesting case for us in New York because during the mid-1990s, this state transitioned from a minimum competency-based exam to a high-stakes, criterion-referenced, one, aligned to higher learning standards. In this way, they mimic New York State as it transitions to a Regents diploma for all.

enrollment growth in GED programs (Amrein and Berliner, 2002a). However as Table 4 demonstrates, the effect of high school exit exams was not universal across the board. Five states had decreases in dropout rates after the implementation of high-stakes exams, while slightly more--8 states--experienced increases in dropouts; in three states, the authors concluded that the effects at this point were unclear.

Table 4
Impact of High Stakes High School Tests on Dropout Rates
for 16 States Requiring Them, as per Amrein and Berliner

State	After HS Exit Exam Est'd. Dropout Rate	State	After HS Exit Exam Est'd. Dropout Rate
Alabama	Decreased	New Mexico	Increased
Florida	<i>Unclear</i>	New York	Increased
Georgia	Increased	North Carolina	Decreased
Louisiana	<i>Unclear</i>	Ohio	Increased
Maryland	Decreased	South Carolina	Decreased
Mississippi	Increased	Tennessee	<i>Unclear</i>
Nevada	Increased	Texas	Increased
New Jersey	Decreased	Virginia	Increased

Source: Amrein and Berliner, 2002a

Jacob, in his paper “Getting Tough? The Impact of High School Graduation Exams” using the National Educational Longitudinal Study (NELS) data set, found no appreciable effect of high-stakes exams on dropout rates after controlling for prior school achievement and other school, student and state characteristics for average students. However, low-achieving students in states with exit exams were 25 percent more likely to drop out of school than were their counterparts in states that did not require passage of high-stakes tests (Jacob, 2001).

Warren and Edwards’s approach (in their paper “High School Exit Examinations and High School Completion: Evidence from the Early 1990s”) is similar to that of Jacob in that they use the NELS. It attempts to tease out the association between exit exams and high school completion and the extent to which it varies by students’ socioeconomic status and prior academic records. They found that such exams are not associated with lower levels of high school completion or diploma acquisition (Warren and Edwards, 2003).

Moreover, a study of the impacts of Minnesota’s graduation exams found that the exam had no negative impact on dropping out: the rate of 11 percent dropouts for the four years prior to test implementation, remained the same after the exams were implemented in 2000 (Davenport, Davison, Kwak, et al., 2002). Similarly, Carnoy and Loeb found no effect of high-stakes tests such as exit exams on grade progression for

black or white students, although they could not rule this out for Hispanics (Carnoy and Loeb, 2003).

On balance, at this point in time, it appears that the evidence is not conclusive and **we are unable to say whether high-stakes exams are leading to higher rates of dropping out.** *Indeed, an expert panel convened to answer this question, concluded that there is only 'moderate, suggestive evidence, to date, of exit exams causing more students to drop out of school'* (Center on Education Policy, 2003). Decisions to drop out of school, like those related ones of whether to leave a job or career, are probably very dynamic, complex processes and therefore not simple ones to model or simulate.

CHANGES IN TEACHER BEHAVIOR

The opponents of high-stakes tests and perhaps by extension, of higher curriculum standards overall, argue that in order to maximize test scores for their students, teachers will be forced to deliberately focus on the limited content that the standards cover. This in turn will result in 'dumbing down' their curricula and 'teaching to the test', thereby leading to less of a focus on true education, in favor of the lesser standard of test passage.

Little systemic, national and comprehensive research of the change in teacher curriculum, instruction and other behaviors engendered by high-stakes testing has been conducted however. A recent study (Clarke, Shore, Rhoades, et al., 2002) of 360 educators in three states that are characterized by the authors as low (Kansas) to moderate (Michigan) to high (Massachusetts) in terms of the stakes or consequences for student failure was conducted. In it, the authors found among others, that:

- *Between half and three quarters of the educators in all states were **neutral to positive** in the effect of exams on aligning curriculum to the standards and creating a focus on problem solving and writing, with Massachusetts and Kansas teachers most positive in this regard;*
- Negative effects were cited by a minority of teachers but most frequently by (high-stakes) Massachusetts educators; these effects include a narrowing of the curriculum, inappropriate pace and material, and decreased flexibility;
- In all three states, teachers reported that test preparation did involve, to various degrees, excising, emphasizing or adding content in order to align to new curricula, with high-stakes Massachusetts teachers reported this trend most commonly; and
- Interviewees in all three states responded that instructional practices had changed as a result of state tests, with Massachusetts again leading the other three states in this regard. Teachers were quick to note important positive changes that have been engendered by testing (i.e., a focus on writing and

creative thinking, discussion and explanation) as well as negative ones (that include a focus on breadth rather than depth of knowledge, increased time spent on test preparation, and reduced instructional creativity).

Another similar study, which examined the teaching practices of a nationally representative sample of teachers, revealed very similar findings. Like the three-state study, it found that the intensity of classroom, instructional and teacher change was greatest in high-stakes test states. Not surprisingly, teachers in high-stakes test states found themselves: engaging in more test preparation; feeling under greater self-reported stress to have their students do well; and in aligning their instructional plans to the items and core content these assessments were designed to test for. Perhaps the most disturbing findings reported by this research team was the following: that a majority of teachers at each grade level found that “state testing programs caused them to teach in a manner which did not accord with their own views of what constitutes good educational practice” and *roughly three quarters of teachers, regardless of stakes or grade levels, found that the benefits of testing were not worth the costs and time involved* (Pedulla, Abrams, Madaus et al., 2003).

On the positive side, teachers overwhelmingly were laudatory in their comments regarding the effect of tests on focusing their states on developing curriculum standards and aligning their own curricula and instructional methods to the standards. Moreover, the majority of teachers across grade levels disagreed that tests were causing more students to drop out or be retained in grade (Pedulla, Abrams, Madaus et al., 2003).

If the self-reported impacts of high-stakes testing upon teachers is a key element in evaluating some of the potential “backwash” effects of high-stakes test, the views of national board certified teachers should prove especially important. One such study of these highly trained and experienced instructional practitioners has been conducted in Ohio. According to this sample the high-stakes testing model is seriously flawed. Indeed, an overwhelming majority asserts that changes engendered by testing have not been positive. For example, eighty percent believe that teacher autonomy has declined; close to all of them (98 percent) indicated that students spend too much time in test preparation; and more than 90 percent of respondents opined that high-stakes testing does not support developmentally appropriate practices for students. Perhaps even more disturbing is the finding that virtually all (97 percent) of the sampled teachers feel that tests negatively affected students’ love of learning and that 6 out of 7 sampled teachers said that the quality of education has declined in their classes since the institution of high-stakes tests (Rapp, 2001).

Some of the most recent research work in this area has been an evaluation of the implementation of the California High School Exit Exam (CAHSEE). The CAHSEE evaluation generally concords with other findings already discussed in this paper. Among these, are that: the exams have had the effects of focusing instruction on the standards and developing remediation and intervention for those students not yet reaching them; remediation has only limited effectiveness at mastery of the standards; and the lack of prerequisite skills have hampered many students from receiving the

benefit of courses that provide content relevant to the standards (Human Resources Research Organization, 2003).

In conclusion, the evidence of backwash effects of high-stakes tests on teacher behavior is mixed. On the positive side of the ledger, the consensus of the research is that the exams force instructors to infuse the curriculum with standards-related content, thereby creating a hierarchy of educational priorities. Moreover, contrary to the contention of some that test preparation will yield an emphasis on “drill and kill” pedagogical techniques, teachers relate that more labs, discussion and critical thinking are occurring in classrooms (Clarke et al., 2002). On the other--negative--side of the ledger, teachers themselves complain of: losing autonomy; of teaching inappropriate material too fast; focusing on test preparation to the detriment of other learning; and reducing pupils’ desire to learn.

Whether one views the net result as positive or negative probably depends on where one sits and what one values. For example, if one believes that the prime purpose of an education is to provide future workers with the tools to successfully compete in the nation’s economy, the fact that teacher autonomy and flexibility has been reduced in preparing students for high-stakes exams may not prove to be a compelling rationale for overturning such policy prescriptions. Moreover, the establishment of high-stakes exams appears to inevitably involve trade-offs: although academic achievement benefits in aggregate, it does so at the cost of probably more dropouts, and a generally disgruntled teaching community. So what one must ask is whether the 1 to 2 grade (depending on the study) improvement gains engendered by high-stakes testing, is enough to outweigh these costs.

THE COSTS OF HIGH-STAKES TEST DEVELOPMENT AND IMPLEMENTATION

Of the various questions that this policy brief attempts to answer, the question of the cost of implementing high-stakes high school exams has probably the weakest research base behind it. Few studies have examined this issue. The reasons for this dearth of information are that these standards-based exams (as opposed to the older, basic competency tests) are quite new. Moreover, although the direct costs of test creation are straightforward, the indirect costs are less clear. For example, any true accounting of the costs of high-stakes exams should include the costs of research and development, prevention, administration, remediation (for students not passing), and professional development of teachers who must prepare students for the exams. (See Table 5 for examples of activities and expenditures which would fit under these four categories). These broader, indirect costs may not become apparent until well after the test has been administered. This point is well illustrated by a recent cost analysis of Indiana’s assessment system. That study puts the total cost per pupil at around \$445 per child in grades K-12. However, the research team is quick to point out that direct test administration costs comprise a small percentage (18 percent) of this total expenditure as shown in Table 5. The largest costs in this Indiana study were attributable to remediation (29 percent of total test expenditures), while prevention

services and professional development were estimated at 28 percent and 25 percent of total test-related costs, respectively (Rose and Myers, 2003).

The Indiana study also addressed the question of what would be the cost of raising the current performance levels to a scenario in which the preponderance of students passed the Indiana Graduation Qualifying Exam (GQE). Table 5 indicates the marginal cost to bring current performance up to a state where: 80% of students are proficient on either the math or ELA exit exam; the percent of initial passers on both exams rises to 75% (from the current rate of 60 percent); and the increase for minority, low-income and special education students mimics that of the general population as a whole.

Table 5
Estimated Program Costs for Indiana Graduate Qualifying Exam

<i>Expenditure Category</i>	Share of Costs	
	<i>Current Performance</i>	<i>Improved Performance</i>
Prevention ¹	28%	43%
Professional Development ²	25%	31%
Remediation ³	29%	16%
Testing/Administration ⁴	18%	10%
Other	1%	1%
Total Expenditure per Pupil (Grades K-12)	\$444	\$685 *

Source: Rose, D. & Myers, J. (2003). Measuring the Costs of State High School Exams: an Initial Report

*Note that it would require an additional \$685, on top of the already expended \$444, for a total cost of \$1,129 per pupil

¹ Include expenditures to prevent school failure, such as revamping instruction, instruction techniques to better reach special ed and LEP students, instituting early learning programs, etc.

² Expenditures to train teachers: to teach in a standards-based environment; to teach students with learning challenges; to administer tests; or to use scores for diagnostic purposes.

³ Includes the costs of: summer school, after school programs, tutoring, academic intervention services, etc. for students failing to pass a test.

⁴ Includes: expenses to develop and disseminate information about exams; develop, research and write exams; keep records and analyze test effects (e.g., cut scores or passing points).

Not surprisingly the marginal or incremental cost of going from the current state to greater proficiency is higher than the base cost of the *existing* system. The additional

\$685 per pupil spending is not inconsiderable: it reflects an 8.5% increase in statewide K-12 spending. More importantly, the cost shares have changed significantly. In the current state, roughly equal shares of test-related expenditures were assignable to prevention, remediation and teacher development. To generate improved academic achievement a greater thrust is given to prevention and development. An implication is that as a testing system matures, less money may be required for traditional test administration and development. Moreover, we would have expected that the students who were on the margin of passing at first have been able to pass through remediation efforts. However, the problems of the non-passing students after some time has elapsed are more intractable and will require greater prevention efforts, focused on the primary grades.

CONCLUSIONS

In conclusion, the data support the findings that high-stakes exams have been associated with:

- Overall academic achievement gains;
- Both dropout rate increases as well as decreases;
- Mixed effects – both positive and negative- on students' motivation;
- Mixed effects on teacher behavior in terms of change in curriculum and instruction; and
- Significant cost increases to develop and administer tests, while also preparing teachers and students to teach and take them.

Whether one views the net result of these effects probably depends very much on one's view of the purposes of education and one's values or orientation toward issues of equity and educational access. Moreover, the extent to which one values teacher autonomy (in terms of choices about the curriculum to teach and the instructional methods to employ) and morale should also likely figure into one's assessment of whether high-stakes test are on balance a net benefit or cost to the educational enterprise.

Policymakers would do well then, to consider this last issue wisely. Teachers are the front line bureaucrats, in this process toward high-stakes testing, by administering exams and preparing students for their passage. Therefore, the success, or conversely the failure of this reform ultimately rests on their shoulders. As such, the extent to which the reform is implemented at least depends on their buy-in and willingness to do so. Educational decision makers then, would do well to engage in public education efforts, particularly with the teaching community, to carefully lay out the case for high-stakes exams, citing in particular, the overall academic achievement gains that, as this research has shown, these efforts have engendered.

WORKS CITED

- Amrein, A.L., and Berliner, D.C. (2002). High-stakes Testing, Uncertainty and Student Learning. *Education Policy Analysis Archives*, Vol. 10, no. 1 (March 28, 2002). Available from: <http://epaa.asu.edu/epaa/v10n18/>; Internet: Accessed March 3, 2004.
- Amrein, A.L., and Berliner, D.C. (2002a). An Analysis of Some Unintended and Negative Consequences of High-stakes Testing. Education Policy Studies Laboratory of Arizona State University. Available from: <http://www.asu.edu/educ/eps/EPRU/documents/EPSTL-0211-125-EPRU.pdf> Internet: Accessed March 3, 2004.
- Bishop, J. Do Curriculum-Based Exit Exam Systems Enhance Student Achievement? *Consortium for Policy Research in Education*, Research Report (RR) 40, 1998. Available from: <http://www.cpre.org/Publications/rr40.pdf>; Internet: Accessed March 3, 2004.
- Braun, Henry. Reconsidering the Impact of High-stakes Testing. *Education Policy Analysis Archives*, Vol. 12, no. 1 (January 5, 2004). Available from: <http://epaa.asu.edu/epaa/v12n1/>; Internet: Accessed March 3, 2004.
- Carnoy, M. and Loeb, S. Does External Accountability Affect Student Outcomes?: A Cross-State Analysis. *Education Evaluation and Policy Analysis*, 24(4) (2003): 305-331.
- Center on Education Policy. State High School Exit Exams Put to the Test. Available from: <http://www.ctredpol.org/highschool/1/exitexam4.pdf>; Internet: Accessed March 4, 2004.
- Clarke, M., Shore, A., Rhoades, K., Abrams, L., Miao, J. and Li, J. Perceived Effects of State-Mandated Testing Programs on Teaching and Learning: Findings from Interviews with

- Educators in Low-, Medium- and High-Stakes States. January 2003. Boston, MA: National Board on Educational Testing and Public Policy. Available from: <http://www.bc.edu/research/nbetpp/statements/nbr1.pdf>; Internet: Accessed March 4, 2004.
- Davenport, E., Davison, M., Kwak, N., Irish, M. and Chan. C. Minnesota High-stakes High School Graduation Test and Completion Status for the Class of 2000. September , 2002: Minneapolis, Minnesota. Office of Educational Accountability. Available from: <http://education.umn.edu/oea/II/Reports/CompletionStudy/ComplStudy.pdf> Internet: Accessed March 9, 2004.
- Hanushek, E. and Raymond, M. High-Stakes Research. *Education Next*, Spring 2003. Available from: <http://educationnext.org/20033/48.html>; Internet: Accessed March 3, 2004.
- Hughes, S. and Bailey, J. What Students Think of High-Stakes Testing. *Educational Leadership*, Dec. 2001/Jan. 2002. pp 74-76.
- Human Resources Research Organization. Independent Evaluation of the California High School Exit Examination (CAHSEE): AB 1609 Study Report-Volume 1. May 1, 2003. Available at: <http://www.cde.ca.gov/statetests/cahsee/eval/AB1609/volume1.pdf> Internet: Accessed March 4, 2004.
- Jacob, B.A. “Getting Tough? The Impact of High School Graduation Exams.” *Educational Evaluation and Policy Analysis*, (2001) 23(2).
- Jones, M., Jones, B., and Hargrove, Y. *The Unintended Consequences of High-Stakes Testing*. Lanham, MD: Rowan and Littlefield Publishers, 2003.
- Klein, S., Hamilton, L., McCaffrey, D. and Stecher, B. What Do Test Scores in Texas Tell Us? *Education Policy Analysis Archives*, Vol. 12, no. 1 (January 5, 2004). Available from: <http://epaa.asu.edu/epaa/v8n49/>; Internet: Accessed March 3, 2004.
- Koretz, D., Barron, S., Mitchell, K. and Stecher, B. Perceived Effects of the Kentucky Instructional Results Information System. (1996) Available from: <http://www.rand.org/cgi-bin/Abstracts/e-getabbydoc.pl?MR-792-PCT/FF> Internet: Accessed March 10, 2004.

National Education Association. Estimates of School Statistics Database: Table H-16 (page 75). Available at: <http://www.nea.org/edstats/images/03rankings.pdf>; Internet: Accessed March 4, 2004.

Pedulla, J., Abrams, L., Madaus, G., Russell, M., Ramos, M., and Miao, J. Perceived Effects of State-Mandated Testing Programs on Teaching and Learning: Findings from a National Survey of Teachers. March 2003. Boston, MA: National Board on Educational Testing and Public Policy. Available from: <http://www.bc.edu/research/nbetpp/statements/nbr2.pdf>; Internet: Accessed March 4, 2004.

Rapp, Dana. National Board-Certified Teachers in Ohio Give State Education Policy, Classroom Climate and High-Stakes Testing a Grade of F. *Phi Delta Kappan*, November 2002. Vol. 84, no. 3., p.215.

Roderick, M. and Engel, M. “The Grasshopper and the Ant: Motivational Responses of Low-Achieving Students to High-stakes Testing.” *Educational Evaluation and Policy Analysis* (2001).

Rose D. and Myers, J. Measuring the Cost of State High School Exit Exams: An Initial Report. Available from: <http://www.ctredpol.org/highschool/1/measuringcost/indiana.studyfeb03.pdf>

Rosenshine, B. High-Stakes Testing: Another Analysis. *Education Policy Analysis Archives*, Vol. 11, no. 24 (August 4, 2003). Available from: <http://epaa.asu.edu/epaa/v11n24/>; Internet: Accessed March 3, 2004.

Sheldon, K. and Biddle, B. “Standards, Accountability, and School Reform: Perils and Pitfalls.” *Teachers College Record*, (1998) 100(1), 164–180.

Smith, M.L., and Rottenberg, C. Unintended Consequences of External Testing in Elementary Schools. *Educational Measurement, Issues and Practice*. (Winter, 1991), Vol. 10: 7-11.

Stecher, B., Hamilton, L. and Klein, S. Making Sense of Test-Based Accountability in Education. (2002) Available at: <http://www.rand.org/cgi-bin/Abstracts/e-getabbydoc.pl?MR-792-PCT/FF> Internet: Accessed March 10, 2004.

U.S. Census Bureau. 2000 Census of Population and Housing: Profiles of General Demographic Characteristics. Available from: <http://www.census.gov/Press-release/www/2002/demoprofiles.html>

U.S. Census Bureau. Table ST-EST2002-ASRO-03-State Characteristics Estimates. Available from: <http://eire.census.gov/popest/data/states/ST-EST2002-ASRO-03.php>

Warren, J., and Edwards, M. High School Exit Examinations and High School Completion: Evidence from the Early 1990s. Paper presented at the Fourth Annual Undergraduate Research Symposium at the University of Washington.