

## **Testing the System: What passes and what fails in high school exit exam policies**

*By Sally Kilgore and Drew Johnson*

'Tis the season for caps and gowns, for *Pomp and Circumstance*, and for graduation dreams shattered by failed high school exams. As graduation ceremonies approach, accounts of diligent, hard working high school seniors deprived of their diplomas will surface in newspapers throughout the United States.

Confused readers may learn that a young man—despite excelling in calculus and physics—will not be able to march across a stage to receive his high school diploma because his score on an English exam is a few points shy of the state benchmark. Perhaps a story will emerge of a young woman admitted to several competitive colleges but denied her diploma (and thus college admission) because of one poor performance on the mathematics portion of her high school exit exam.

For the indignant reader of these troubling headlines, several questions emerge. How did such a testing system come to be? Why should states require these exit exams anyway?

### ***How did we get these exit exams?***

Most states began administering exams to potential high school graduates in the 1980s. They differed greatly from those used in many states today. The older exit exam, known as a “minimum competency exam,” responded to public concerns that America’s high schools were awarding diplomas to students who couldn’t read, write, add, or subtract.

As the label implies, these exams evaluated basic literacy competencies—most often competencies associated with an eighth grade curriculum—and the majority of students found them amazingly easy. States established passing scores at a level low enough to ensure that nearly all students passed these exams, and, with proper course credits in order, marched into their high school gyms to receive diplomas.

The exams in the newspaper headlines prohibiting students from graduating are proficiency exams. Proficiency exams emerged in the 1990s, as political and business leaders from all regions and political affiliations argued that the basic literacy captured by the minimum competency exams was insufficient for the challenges our young adults would face in the 21<sup>st</sup> Century. Not only is the substance of these exams more challenging, but also the passing level is set much higher than that of the earlier minimum competency exams. Testing standards are designed to raise the bar for all students, but not set a ceiling on what exceptional students can learn.

Initial calls for an improved exam system came as international comparisons of student performance in the early 1990s demonstrated that American students performed substantially below their cohorts in industrialized nations in mathematics and science. Concurrently, college professors reported their alarm at the lack of adequate preparation of entering freshmen. With over 25 percent of entering freshmen assigned to one or more remedial classes, few could disagree.

Consistent with the strategy used to revitalize and render competitive many of America’s corporations, business and political leaders called for establishing academic standards that reflected what a “world class” education looked like—i.e., an education comparable to the best in the world. The focus of many educational bureaucrats shifted from individual student performance relative to each other within a school, a district, or a state to national student performance relative to the rest of the world in mathematics

and science. Political leaders urged states to focus their instruction towards the economically viable objective of preparing students for the shifting career options emerging in the 21<sup>st</sup> Century. President George H. W. Bush moved quickly to give the effort national legitimacy, incorporating the need for states to establish academic standards into federal legislation relating to elementary and secondary education.

For academic standards to have any merit, one must determine if students succeed in meeting them. To address that challenge, states introduced proficiency exams—usually replacing minimum competency exams—that varied in their approach and rigor. Some states, like Georgia and New York, require students to pass end-of-course exams in such courses as Algebra I and biology. Most states require standards-based exit exams more generic in scope, such as language arts and mathematics. To date, 24 states require or will require by 2008 that students pass one or more proficiency exams to attain a high school diploma.

Determining a passing “score” for proficiency exams differs from minimum competency exams. A politically tolerable level of failures versus passes determines the acceptable score for a minimum competency exam and is thus established in a statistically-generated passing score. Conversely, collective human judgment sets the bar for passing a proficiency exam. Most states rely upon a panel of educators and other professionals to establish that bar, using the expectations embedded in the state standards. It is those standards, not arbitrary administrative decisions, which presumably reflect the knowledge students need in their subsequent life experiences—as students in post-secondary education, as participants in the labor force, and as valuable citizens.

#### ***What are the benefits of standards-based exit exams anyway?***

Theoretically, a standards approach to improving student learning can have profound and healthy implications. First, standards can improve equality of opportunity. Standards give students, parents, and teachers access to clear statements of what a student needs to know and be able to do to succeed at the next level of education or to attain a meaningful career (rather than just a job) after high school. The goal of practical preparation reflects the most common ambition states have adopted for their academic standards.

Americans have not been indifferent to public statements about what students need to know and be able to do. The Carnegie reforms in the early 20<sup>th</sup> Century set expectations for those aspiring to college in broad bands—two years (or two Carnegie units) of a foreign language, four years of English, three years of mathematics, and so forth. However, what students were learning during two years of foreign language at Hanover Academy could be vastly different from two years at Hamilton High, Anywhere USA. The difference, in large part, was a function of what teachers and parents knew about what students needed to prosper in their local communities and colleges.

Assessing students’ mastery of standards can also improve our evaluation of school performance. The long standing tradition of ranking schools and students as “best” or “worst” obscures what all students know. Such rankings shed little light on the extent of the differences existing between schools ranked first and second in average SAT scores, for example. Evaluating schools in terms of the proportion of students meeting academic standards theoretically allows all schools or students to perform well, or all can perform poorly. This is a profound and important change.

The standards-based approach also eliminates some illusions that inevitably emerge with percentile-ranked test reports. It is possible, for instance, for a student to answer only 60 percent of the items correctly, but place in the 95<sup>th</sup> percentile because everyone did worse than he or she did. While parents can feel good about their child performing in the 90<sup>th</sup> percentile, they will not know that their student—and most others—actually knows very little about the subject matter of the test. It’s the knowledge and skills that young people acquire, not a valueless ranking, that ultimately allow them—individually and

collectively—to prosper in social, political, and economic spheres. Proficiency exams based on capably-crafted standards constitute a shift toward valuing what students actually know, not simply just how they rank.

That’s the rosy part of the story.

***So, what’s the problem?***

The dark clouds are substantial. Faced with abysmally low passing rates on high school proficiency exams, some states are backing away from full implementation of them. This past summer, California’s State Board of Education voted unanimously to postpone applying penalties for poor results on its statewide assessment—a move that spared 92,000 students from failing. New York recently voided the results of its math exam based on disproportionate failure rates among minority groups. In these two cases, a central issue was whether students received a fair opportunity to learn: Were the students actually taught the material? Were teachers sufficiently prepared to teach the material? Were the exit exams really aligned with academic standards published by the State? No one really knows.

Other states committed to proficiency exams adjusted their ratings because of faulty questions: The Texas Education Agency gave almost 5,000 sophomores credit for an item on the mathematics portion of the Texas Assessment of Knowledge and Skills (TAKS) for which they were initially denied. The TAKS item had two “correct” answers, though most concede that the question as posed for students included an impossible set of mathematical relationships. This problem is not limited to the U.S. On an A-level 2001 physics exam in the United Kingdom, a question could not be answered with the information provided. These and many other cases reported in the recent report of the National Board on Educational Testing and Public Policy demonstrate that the skill of those developing test questions is also part of the problem.

Human errors also occur in statistical manipulations—norming scores, equating tests, and transferring data. In 2000, scorers incorrectly applied norming formulas to a number of individual student and school scores in California. These errors resulted in inflated scores for 22 schools—resulting in erroneously-given teacher bonuses. In Virginia, the cut scores for the 2002 writing test were set one point too high leading officials to raise the scores of over 5000 students in the fifth and eighth grade from fail to pass upon the discovery of the error.

The challenge of reducing errors in high school proficiency exams does not end with teacher preparation, statistical errors, test developers, or test administration. To the extent that their administration occurs concurrently with federally-mandated exams at other grade levels, it is important to consider the burden on corporations managing and processing student responses to exam questions.

The No Child Left Behind (NCLB) Act requires each state to align exams with their state standards in mathematics and language arts for grades three through eight, and three science exams—one each in the elementary, middle, and secondary levels. At the secondary level, one mathematics and language arts exam is required in at least one of the grades. As a result, the Education Policy Institute estimates that the number of annual state assessments administered in this country will more than double—from approximately 400 to 850 within three years.

NCLB also requires the results be made available to schools by the time schools open the subsequent fall. That’s reasonable. Teachers and administrators, in fact, do deserve access to information about student performance in a timely manner. In the past, many state regulations required educators to propose a school improvement plan in the absence of state data on their students’ prior successes and failures.

The current technology for grading these exams, however, makes such requirements challenging, to say the least. Most states administer exams in March and early April, leaving contractors less than five

months to complete test administration, data entry, scoring, and reporting. Consider the millions of bubble sheets to be bundled and loaded into Federal Express® and UPS® trucks making their way through Memphis warehouses to scanning machines in Minnesota, New Hampshire, Iowa, or New Jersey. Where needed, workers scan and electronically transfer open-ended responses via Internet to human markers throughout the country. Still more markers may convene at an abandoned barn in New Hampshire (this is not hypothetical). Conservatively, 40 million booklets are supposed to make their way through this system error-free.

The truth is that education testing systems have their own hanging chads—packages get lost, marks cannot be read, and Internet transfers can fail.

Most critics of proficiency exams assume that the current technology and system of testing will remain the same for the foreseeable future. Their assumptions may be justified: Federal funding for the new testing mandates is sufficient, according a recent Government Accounting Office (GAO) report, assuming that the technology for data management remains the same and that all states rely exclusively on multiple-choice items. Using a mix of open-ended and multiple choice items, however, increases test creation and scoring cost to the point that federal funding will meet, at best, only half of the development and administration costs. Obviously, the number of items released to the public each year will also affect costs, as will the system of field-testing used to develop those items—costs not reported in the GAO estimates. Given that high school exit exams will soon operate in the context of such a massive increase in testing, concern is justified.

### ***What to do?***

With all these hurdles, is it fair to have an essential requisite for entry into jobs and post-secondary institutions—the high school diploma—denied to students whose scores on an exam are below the bar set by experts? What public good or individual benefits can be realized with proficiency exams attached to a high school diploma?

To begin with, it isn't fair to have educators accountable for student performance in the absence of some incentive for students to do their best. Students must have some stake in their performance on proficiency exams—at least at the high school level, where students have a greater ability to discern what counts in their own lives, what does not, and perform accordingly. Linking proficiency exam performance to a diploma provides a fair incentive to achieve for students and a reasonable measure of the accomplishment of educators.

Second, students can benefit from some type of accountability. The evidence is accumulating that at-risk students are, in fact, more likely to complete high school when there are consequences of some type attached to their performance on state assessments. Kathryn Schiller and Chandra Muller found in a study of a national longitudinal sample of high school students that at-risk students were five percent more likely to graduate when they had assessments that had consequences attached to them than when they did not. John Bishop found that students in New York showed greater gains in achievement when consequences existed. That said neither of these studies succeeded in evaluating the relative value of diverse types of consequences.

Finally, then, is the claim of a larger public good possible through proficiency exams. The Center for Education Policy's recent report on exit exams aptly captures the most pervasive goal for proficiency exams: The new high school exit exams are designed to make a high school diploma "mean something." But a high school diploma already means something. Its value is embedded already in various job entry and college admission requirements. For most employers, a diploma means that the holder has, at the very least, demonstrated basic literacy skills and shown some tenacity in sitting still even when things are boring. Requiring students to pass a new high school exit exams, then, is not *giving* "meaning," it's

*changing* it. Changing the meaning of a high school diploma is much more difficult than creating a new signal. More importantly, it eliminates an operative signal in the labor market.

Students deserve, and employers could use, more discriminating signals about the knowledge and abilities gained by an individual as a result of their high school experience. Since high school diplomas already have meaning in the marketplace, why not *add* more information to the high school diploma? Delaware, for instance, has three levels of a high school diploma: basic, standard, and distinguished—partially determined by a student’s score on their statewide exams.

What if students were allowed to acquire a basic diploma the old-fashioned way: one credit at a time? A proficiency diploma (or in Delaware, a standard diploma) could be awarded to those students who earned the proper number of credits *and* attained passing scores on proficiency exams. Advanced or distinguished diplomas could be awarded to those students who perform exceptionally on the state exit exams.

This change may be especially important for minority students. Researchers at the Manhattan Institute find that among Hispanic students, the unemployment rate is practically the same for high school completers and high school dropouts—32 versus 33 percent. In contrast, Anglo students who complete high school have lower unemployment rates than those who dropout. The researchers argue that racism cannot account for this; rather it suggests that employers don’t believe a high school diploma conveys useful information about what Hispanic students know and are able to do.

Providing a differentiated diploma system, by itself, will not resolve all the concerns surrounding high school exit exams. Supporters of high school exit exams will remain fearful that students lack a meaningful incentive to work hard. Critics can still claim that marginal students are just as likely to become discouraged—either at their first or subsequent failures to pass the exam. Both critics and supporters must confront the technical and logistical issues that, unless resolved, will diminish the credibility of proficiency tests.

How can states maintain a meaningful incentive with less ominous penalties? Students could not only have the opportunity to have a higher octane diploma, but also earn the opportunity for scholarships to post-secondary institutions. Michigan, for instance, provides scholarships to students demonstrating proficiency rather than denying diplomas to those who cannot do so. If, however, states choose to follow such a strategy, they must prepare financially to support a system where nearly all students become eligible for them. After all, the goal is to have all students earn a score considered proficient by 2012. Michigan already faces compromises in their scholarships given improved student performance.

Guaranteed admission to public universities and colleges is also a useful motivator. Texas currently guarantees admission to some state universities based on class rank. Press reports suggest, though, that some students game the system by transferring to high schools where they can improve their chances of getting into the “top ten percent.” Yet, if scholarship eligibility required both high grades *and* strong performance on the state’s exit exams, every student would have a reason to excel *and* attend schools with challenging curriculums.

While only in the early stages, the American Diploma Project seeks to evaluate the degree to which high school proficiency exams actually predict what students need to know and be able to do for success in college and the workplace. Started just two years ago, the project brings together The Thomas B. Fordham Foundation, Education Trust, Achieve, Inc. and the National Alliance of Business—funded by a grant from the Hewlett Foundation. Collaborating with several states, they seek to link actual proficiency exams with specified needs for success in colleges and the workplace. If such a linkage occurs, then employers and college admission officers could more readily utilize the value of scoring at the proficient

or advanced level on a high school exit exam. That would add value—and thus provide incentives—for students to do their best on proficiency exams.

Rules governing assessments and accountability also need to change to remedy a variety of administrative challenges facing the current system. With standards as the organizing principle for assessments, it is no longer necessary to administer tests simultaneously throughout a school or a state. If standards define exam scores, rather than performance relative to others, then *when* a test is administered poses little problem. Already states are providing multiple opportunities for high school students to complete proficiency, as well as minimum competency, exams. Why not take it a step further by allowing students to sit for an assessment when they feel best prepared—whether it is earlier or later than anticipated by the state system. On-line testing with a large item pool—especially adaptive testing as some districts in Illinois are currently piloting—could minimize the intrusions into life in classrooms at all grade levels and reduce the likelihood that students could compromise the integrity of the exam. Such online testing will also serve to reduce the administrative costs associated with transporting bubble sheets from schools to processing centers and reduce the errors generated by transferring bubble sheets into electronic data.

Critical to reducing the likelihood of discouraged students, though, are the ways in which schools are accountable. High schools, consistent with Federal laws, should remain accountable for student performance. States, though, could hold them accountable not only for the proportion of students who pass the exam by, say, their 17<sup>th</sup> birthday, but also for the proportion of students who pass an exam at their *first* sitting.

Consider this: The Center on Education Policy reviewed data in three states to determine the proportion of students passing exit exams on their first attempt. Assuming that the students most likely to become discouraged and thus most likely to fall short of a diploma are those from low-income families served by free and reduced lunch programs, then, in Minnesota, for example, almost half of their at-risk students fail one or more of the exit exams on their first attempt.

Rewarding schools not only for increasing the proportion of students eventually passing, but also for ensuring that students are adequately prepared the first time they sit for an exam introduces incentives to prepare students adequately the first time. Such a plan would save money and time used administering exams and reduce the number of discouraging performances by ill-prepared students

A vital question remains: Are teachers adequately prepared to help students meet the standards established by states? In many cases, the answer is “no.” One obvious challenge is the absence of secondary teachers with certification in mathematics and science—especially in rural areas. It is no secret that, even among those certified, teachers vary substantially in their depth of knowledge.

Conscious of the problem, Congress appropriated almost a billion dollars for states and districts to use to improve the quality of teachers. These funds can be used for recruitment, to increase salaries, or provide for professional development. How are states using the funds? The data are not yet available.

The U.S. Department of Education provides substantial support to organizations that certify teachers almost exclusively on their discipline-based preparation—i.e., whether they majored in English, mathematics, science, and so forth. The notion, however, that subject-matter knowledge is nearly all one needs to be an effective teacher holds, at best, only if you are talking about serving students who are well prepared, eager to learn, and confident in their capabilities. Less prepared students—those with less confidence and eagerness—need teachers who have a wide repertoire of scientifically-based strategies to draw them into intellectual pursuits. Unfortunately, traditional state certification does not always

guarantee that new teachers come prepared with this wide repertoire any more than innovative discipline-based means of certification.

States address the challenge of poor educator performance in various ways. Virginia legislators recently proposed, for example, to exempt seniors at low-performing schools (i.e., where fewer than 70 percent of the students pass the six high school exit exams) from the state's graduation requirements. Legislators sensibly argue that it is unfair to hold students responsible for test results that more than likely reflect weaknesses in the preparation of educators. Abandoning student accountability at schools where pass rates are low does not, however, achieve fairness.

If many students are failing, why not conclude that teachers could use some help? States must give priority to using federal Teacher Quality funds to improve the substantive and strategic knowledge of teachers. For instance, states could focus on improving the skills of teachers at high schools where student performance in one or more subjects is more than, say, 20 percent below state expectations. Teachers in the relevant disciplines could attend a month-long summer seminar designed to increase the teachers' depth of knowledge of their subject, as well as to broaden their repertoire of strategies to help less motivated students acquire mastery of state standards.

### ***What happens to our students if things don't change?***

If things stay the same, the pressure of states and administrators to lower the expectations on high school proficiency exams will continue and the bar will be lowered until only the most unprepared students fail to make the grade. Evidence of such pressure is already apparent. In Texas, for instance, state officials posted on their Web site not only the likely pass rate on their 11<sup>th</sup> grade assessment exam if the passing bar (score) established by the state's panel of experts is used, but also the pass rate if their bar is statistically adjusted downward.

Statistically lowering the bar for what is considered "proficient" on a state assessment misses the point of the whole effort to establish standards-based education systems in our states. If we are to anchor educational goals in this country to academic standards, then it is necessary to establish a meaningful link between those standards and actual performance. Such manipulations are not much different from rankings where only so many students and schools can be "above average." Both compromise a society's ability to understand their students' knowledge and capabilities.

Standards-based assessments must differ from compromised targets and trite ranking systems. Instead, they should establish benchmarks for elementary and secondary students that parallel those embedded in the testing and licensing of professionals such as pilots and architects. The benchmarks for "passing" these exams are not set to guarantee that a predictable number of folks become architects or pilots, instead they are set to ensure the safety of the public. In the case of high school proficiency exams, benchmarks should reflect what is necessary for students to succeed in college, the workplace, or the armed services upon graduation from high school.

Standards-based expectations are an extremely important shift in our approach to education, one that deserves aggressive protection. Such protection can occur only if policymakers fearlessly consider alternative ways of structuring most state accountability systems. Future revisions of standards should occur after attaining sufficient information on their predictive validity—that is, empirical evidence on the degree to which performance on exit exams actually predicts successful completion of college, capacity to obtain meaningful career tracks if students choose to go directly into the labor force, or their ability to pursue a military career. That said, proficiency exams should never be viewed as a predictor of wealth or status, but rather a predictor of capacity to participate meaningfully in American society.

### *A vision of the future?*

Last, but not least, improving the quality of secondary education requires a long-range vision of how educational accountability should evolve. It must eventually include incentives and mechanisms to support the development of a much broader range of talent—much of it now remaining dormant—in our youth. Students need many more avenues to demonstrate excellence, not just adequacy. European countries have acknowledged this fact by creating a voluntary exam system that covers a broader range of academic disciplines as well as the commercial arts, thus allowing colleges and universities to identify exceptional talent in different combinations and special concentrations. Students opt for the exams that best demonstrate their talent and potential. A broader range of exams—inclusive of the commercial arts and humanities—can further reduce the proportion of students discouraged by existing exams.

The American tradition of finding and nurturing talent should encourage efforts to raise standards for all young people. Our children deserve to be challenged to achieve excellence in the arena of their choosing—be it academic, artistic, mechanical endeavor, or even public service. They also deserve to enter adulthood with the proficiencies that can allow them a meaningful career in the workforce. Without reconfigured state accountability systems of standards and testing, the pressure to lower standards will overwhelm the most committed reformers—leaving our youth ill-prepared for the economic opportunities and societal responsibilities they will face.

*Sally Kilgore is president of the Nashville-based Modern Red SchoolHouse and served on the advisory board for the National Center for Education Statistics for almost a decade.*

*Drew Johnson is president of the Tennessee Center for Policy Research, a nonprofit, nonpartisan research institute committed to responsible government.*