

High Risk or High Time?

A Critical Junction in Implementing Maryland's High-Stakes High School Assessment as a Graduation Requirement

History and background; problems and recommendations:

PUBLISHED BY
The Abell Foundation
111 S. Calvert Street, Suite 2300
Baltimore, Maryland 21202
www.abell.org

JULY 2003

Executive Summary



The Stakes

This June, nearly half of all public school students in the United States were required to pass one or more statewide exit exams in order to graduate from high school. In cases like this, where diplomas are contingent on the results of such assessments, tests are characterized as “high-stakes” for students.

In the next six months, the Maryland State Board of Education will decide whether to raise the stakes by making passing scores on the new more rigorous High School Assessments (HSAs) a condition for graduation. If the experience of other states is any indication, this decision could result in as many as 20% of high school students in the Class of 2008 being barred from graduation as a result of failing one or more of these tests. It will come as no surprise that denied students will disproportionately be disadvantaged students from Baltimore City or Prince George’s County or students who receive special education services.

The decisions around the HSA program will have irreparable consequences for Maryland’s children, yet neither the public nor educators has been privy to much of the dialogue or data surrounding such a critical move. Moreover, the environment in which these decisions will be made is very different, and more formidable, than a decade ago when the HSA program was first conceived. The Board’s charge is challenging: how will it raise standards for all students without over-penalizing those who are disadvantaged?

The Challenges

Maryland’s HSA program is a series of high school end-of-course exams designed to hold students accountable for mastering specific skills or content knowledge. Whether passing these will indeed become a graduation requirement in Maryland remains up in the air, but as now proposed, students entering high school in Fall 2004 would have to pass at least four HSAs to graduate from high school in 2008.

The trend toward state exit exam requirements is expedited by the federal No Child Left Behind Act of 2001, which requires all states to implement testing at the high school level. But as states pursue a high-stakes route, they are confronting tough challenges forcing them to re-think and delay testing implementation—and Maryland is no exception.

Educational concerns and financial uncertainties have forced the State Board to delay making the HSA tests a graduation requirement three times to date, most recently in May 2003. This summer and fall, the Board is slated to address the accountability issue by:

- establishing passing scores and proficiency levels for the HSAs;
- deciding whether to increase the number of end-of-course exams to 12, of which students would have pass 10 to graduate;

-
- exploring issues of school and district performance and accountability; and
 - finally, determining whether to make the HSA tests a graduation requirement.

Currently, all high school students must take HSA tests in five courses, and normative test scores (how a student performs against others) are reported on their transcripts—yet no proficiency levels to measure whether they have met course standards have been set. As a result, neither the public nor state educators have any indication of the proportion of students who failed, met, or exceeded state expectations of what they must be able to know and do. One outcome, however, is known: Maryland’s 2002 HSA data show predictable achievement gaps between low-income minority students and their more affluent counterparts as well as between special education students and regular education students.

Meanwhile, the State has yet to articulate how it will determine whether the HSA program improves learning, achievement, and post-graduation success for Maryland’s students. While it will be clear how many students eventually pass the HSAs, this measure is the means rather than the end of this reform. Scores on any test will increase as teachers become familiar with its format. Beyond scores on the HSAs, the State must clearly articulate its larger objectives for the HSA program and set out how it will determine progress towards those goals.

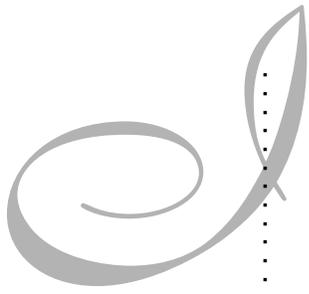
It is up to the Board to ensure that the HSA program is implemented with the highest probability of producing improved learning and achievement. If the Board decides it is committed to high-stakes high school testing, it must explore remedial support for those who fail, second chances for test-taking, and alternatives to the HSAs for certain populations. It must also move to reconsider its test format and scoring procedures to ensure timely dissemination of test results and data as part of building a collaborative system with districts, schools, students, and the public. An external evaluation tracking impact of the HSA program is critical.

A. The Role of This Report

Given the breadth and complexity of the high-stakes testing debate, there are, unfortunately, no easy answers to the dilemmas that Board members now face. This report aims to engage policymakers in, and focus public attention on, a critical evaluation of the state’s proposed HSA program, to ensure that further investments bring concrete improvements for student learning. Because there is much to be learned from other states, as well as Maryland’s own experience in high-stakes testing, the report has drawn on multiple sources. A series of questions are raised in an effort to guide the Board’s deliberations and to urge reconsideration of its earlier assumptions.

Finally, the report offers recommendations to provide additional or alternative strategies for consideration as the Board moves forward in its quest to enhance the value of a Maryland high school diploma.

I. Introduction



**DEFINITION OF
A HIGH-STAKES
TEST: “A TEST
FOR WHICH
THE RESULTS
HAVE IMPORTANT
AND DIRECT
CONSEQUENCES
FOR THOSE WHO
TAKE IT AND
FOR THOSE
WHO USE OR
SPONSOR IT.”²**

In May 2003, the state of Florida faced the dilemma of denying diplomas to 12,000 12th grade students for failure to pass the state’s new high-stakes graduation assessment.¹ Elsewhere around the country, nearly half of all public school students are now required to pass one or more exit exams in order to receive a high school diploma. Because results from assessments like these are used to assign students to classes, make promotional decisions, or award or withhold, high school diplomas, they are characterized as “high-stakes” for students. Nowhere are the stakes higher for students than the new mandatory high school exit examinations that Maryland is considering requiring as the bar for high school graduation.

Definition of a High-Stakes Test: “A test for which the results have important and direct consequences for those who take it and for those who use or sponsor it.”²

While high school diplomas in Maryland have been tied to passing scores on the Maryland Functional tests since 1987, these are 6th-grade-level minimum competency tests in reading, writing, and math that fall far short of evaluating the State’s new high school standards. This summer, however, the Maryland State Board of Education will set passing scores on the new, more rigorous High School Assessments (HSA); in September, the Board will begin deliberating whether to increase the stakes by making a passing score on these tests a condition for graduation.

Part of the stronger K-12 standards and accountability that Maryland has been moving toward in recent years, the HSA Program consists of a series of end-of-course exams in various subject areas. In addition to holding students, teachers, schools and districts accountable for meeting clear academic standards, the State Board created the HSA Program to:

- increase academic rigor;
- enhance the value of the Maryland high school diploma; and
- ensure that graduates have skills and content knowledge.³

Much remains to be decided about how much weight the assessments will ultimately carry— and whether passing them will indeed be a graduation requirement—but in the current proposal, students entering high school in Fall 2004 will need to pass at least four HSAs in order to graduate from high school in 2008.

Maryland’s High School Assessments mirror a national trend toward “standards-based reforms” that seek to make students accountable for mastering specific skills or con-

¹ Hegarty, Stephen. New math: 1,000 pass state test after all. St. Petersburg Times. May 9, 2003.
² Promissor, Inc. Retrieved from <http://www.promissor.com/cgi/glossary/>
³ Maryland State Department of Education, *Maryland High School Assessments*, October 2001, <http://www.msde.state.md.us/hsimprovement/pdf/HSAPPOct2001v1.ppt>

II. Maryland's High-Stakes Dilemma



The decisions the Maryland State Board of Education is poised to make about high-stakes high school testing will have momentous consequences both for students, and for the integrity and viability of the High School Assessment (HSA) program itself—and by extension, Maryland's overall efforts to raise and enforce K-12 standards. Board members must, in their deliberations, weigh important factors and ask tough questions for which there are few easy answers.

While the stakes are many and large, the Board's charge boils down to addressing the following key questions and issues:

- What will the passing scores and proficiency levels be for each of the HSAs, i.e., what will constitute passing for English I, Algebra/Data Analysis, Government and Biology?
- How can the HSA program hold students accountable for achievement without being so onerous that disadvantaged students are overly penalized?
- Should the Board proceed with plans to expand the program from four (or five) to 12 end-of-course exams?
- Should passing the HSA tests be made a graduation requirement? If so, when? Currently students who will enter high school in Fall 2004 are targeted.

Currently, Maryland students must pass prescribed coursework and perform 75 hours of community service in addition to performing at the 6th-grade level on Functional Tests to graduate from high school; clearly the standards must be raised. At the same time, according to at least one study, as many as 4 % of 12th grade students were denied high school diplomas for failure to pass the existing Maryland Functional Tests in 2000.⁶ With more rigid standards and assessment Maryland faces even higher failure rates, and the accompanying educational, workforce, and economic consequences.

A Changed Environment

It is a daunting charge, made more so by the fact that the environment in which these decisions will be made is very different—and arguably much more challenging—than when the HSA program was first conceived 10 years ago. At the same time, it is important that the Board keep in mind that the process of making these decisions presents numerous opportunities to improve learning for Maryland students as well.

Uncertain Financial Times

The fiscal environment in which the HSA program was launched held brighter prospects for education spending than the budget constraints of late. Insufficient

⁶ Amrein and Berliner, 2002

state funding has threatened the Board's ability to fully finance the anticipated remediation needs created by the HSA program. This, in combination with thinly-veiled concern over potential student failure, has contributed to the Board's three separate delays of the graduation requirement; originally intended for students entering 9th grade in Fall 2001, it is now slated to apply to those entering 9th grade in Fall 2004. As the Board advances the program this summer, it must realistically consider its feasibility and scope as state budget woes continue.

Delays: The Pros, Cons, and Prospects for More

Maryland's unstable economic climate has not been the only cause of delays in the HSA program. MSDE staff has acknowledged that it also miscalculated the time needed to construct, validate, and field-test a series of five content-focused tests. Given the delays to date, and the fact that there could be more, the Board must weigh the impact of delays thus far on implementation moving forward.

Due to controversies that have accompanied Maryland's past efforts with standards assessments, as well as the troubled implementation of high-stakes exit exams in other states, there may be advantages to the State continuing its cautious approach to HSA test construction and implementation. On the other hand, delays provide time for testing opponents to create organized opposition. Case in point: during the 2003 legislative session, the coalition Marylanders Against High Stakes Testing rallied behind the failed House Bill HB1166. Had it passed, this bill would have barred the Board from requiring a single assessment or series of assessments for graduation from public high school.

Aside from the strategic pros and cons of whether passing the HSAs should be a condition of graduation, the simple question for the Board remains: If Maryland decides to proceed with existing plans to implement a graduation requirement for students entering high school next fall, and/or expands the number of HSA tests to 12, can it roll these out in a thorough, timely fashion?

No Child Left Behind Act

Maryland's HSA program was well underway when the federal No Child Left Behind Act (NCLB) mandated its own testing and reporting requirements for high school students. Signed into law in January 2002, NCLB aims to raise student achievement and eliminate gaps among students of different backgrounds. Its goals and new high school assessment requirements may require Maryland to rethink the HSA program, particularly in the area of accountability.

NCLB requires states to adopt tests and define proficiency levels in reading, math and science that all students must reach by the end of the 2013-2014 school year. Furthermore, each school must meet annual improvement goals, called "average yearly progress," designed to bring each group of students—aggregated by race and eth-

nicity, income, disabilities, and limited English proficiency—up to those levels. Schools missing their targets over time will be labeled as “needing improvement” and required to target funds to underperforming students. Should the Board implement the HSA graduation requirement and set the bar high for passing, a likely consequence is that a large percentage of high schools would be identified as needing improvement. Furthermore, because NCLB is an under-funded mandate, it may well divert funding and human resources initially slated for the HSA program to the 10th grade NCLB tests in reading, math and, soon, science.

Finally, NCLB requires student groups within schools to hit average annual graduation rate targets, meaning states must ensure that their high stakes assessments do not lower high school graduation rates. This strikes at the heart of the Board’s HSA dilemma—whether to make the HSAs a graduation requirement, knowing that there is both research and recent evidence to show that more rigorous assessments could reduce the number of students graduating, particularly among disadvantaged and disabled youth.

Accountability and Alignment Lacking

Scores, But No Standards

For students now taking the HSAs, scores are reported on transcripts as a percentile ranking, meaning their performance on the test is compared with that of all Maryland students taking the same test. A student’s score does not, however, reflect the percentage of questions answered correctly. To date, students have not been measured against standards, the purported basis of these assessments. As a result, students, parents, teachers and schools have no idea to what degree student performance failed, met, or exceeded state standards in tested HSA courses. If students are to take these tests, the Board must ensure that the HSAs, and their results, are reported as performance against standards. Ultimately, this is the key to improving teaching and learning.

Missing Links with Higher Education

Furthermore, little has been accomplished with higher education officials to link high school performance on the new HSAs to admission and remediation decisions at Maryland’s colleges and universities.

Student Impact

Disproportionate Odds of Failure

Data show that low-income and minority students as well as students with disabilities are at greater risk of performing poorly on high stakes tests, and thus are at greater risk of not receiving a high school diploma under Maryland’s proposed graduation requirement. At the same time, state funding constraints threaten the likelihood that necessary educational supports will be in place to ensure that all students have increased opportunity to perform well on the HSA tests. Front and center in the Board debate that ensues this summer must be the intent and the means to give all students sufficient opportunity to succeed.

Not Enough Second Chances

A student who fails an HSA will be required by the State Board to participate in “remediation” before retaking the test. There are no graduation alternatives proposed to accommodate students who fail an HSA one or more times. When the Board considers its policies for retesting, test preparation, intervention, and remediation for students who do poorly on the HSA, it must strive to ensure that students are provided with the necessary supports and alternative assessment options to help students succeed.

High-Stakes Experience

More Research to Draw On...

While challenges for Maryland’s HSA program have mounted over the last couple of years, more and more research has accumulated, providing states with the opportunity to make better, more informed decisions about high-stakes testing.

As more states move toward high-stakes testing, their collective wisdom and experience offer valuable lessons to others, particularly those states like Maryland with less experience in the high stakes, high standards testing arena (see Appendix D). Also, as high-stakes testing has expanded, so has the number of related researchers, articles, and studies—providing a growing knowledge base to aid Maryland’s policy-making.

...Including Maryland’s Experience

In addition to the growing wealth of experience of other states engaged in high-stakes assessment, Maryland itself has acquired experience since the program was first conceived in the areas of public engagement, test development and field testing, and test administration. The lessons learned here at home are valuable, and the Board would be well-served by factoring them into its decision-making this summer and fall if the HSA program is to be more successful than Maryland’s past standards and assessment efforts.

Poor Public Engagement Revisited

The process through which the HSA program has evolved has been short on information-sharing by the State. Two years into the requirement that all students take HSA tests, they and their parents have yet to receive any meaningful interpretation of their scores. More importantly, the Board itself is preparing to make critical decisions about high-stakes testing without the benefit of any meaningful performance data on how Maryland students actually fared on the HSAs against standards. This lack of public access to information has long plagued public education reforms in Maryland, and it is up to the Board to change that as it considers the implementation of its “highest” high-stakes assessment system ever.

· **Tough Decisions Ahead**

· The consequences of the Board’s HSA decisions for students are monumental. More
· muted in the HSA debate are the consequences for the HSA program itself, and
· Maryland’s larger K-12 reform efforts by extension. If the intention is moving
· Maryland closer to providing all students with an adequate education, a fundamental
· question must drive the Board’s decision-making:

· *How will the State itself measure and be accountable for the degree*
· *to which the HSA program improves high school instruction and*
· *increases opportunities for learning, achievement, and post-gradu-*
· *ation success?*

· This question creates the framework for addressing the issues stated above and com-
· pels examination of the educational and moral dilemmas that intersect student,
· school and district accountability: How well will a system of sanctions and assistance
· improve education for all high school students, particularly if failing students are
· denied diplomas for not meeting standards? What if the graduation requirement caus-
· es failing students to drop out or causes schools to push out failing students, leading
· to higher HSA scores but higher dropout rates? Is increased rigor in awarding of
· Maryland’s high school diploma worth the probability that fewer students will be able
· to earn one?

III. Maryland's History of High School Standards and Assessment



The Maryland Functional Tests

Tying high stakes testing to high school graduation is not new to Maryland. In the late 1970s, the Maryland Functional Test program was designed to ensure that students acquired minimum levels of competency in 165 basic skill or “functional” areas prior to graduating high school. In 1983, the State Board of Education voted to tie high school diplomas to passing scores on these tests, beginning with the Class of 1987. By 1989, a passing score on four minimum-competency, or functional, tests—reading, mathematics, writing, and citizenship—was required for graduation, and remains so to this day for three of the four tests.⁷

The Maryland Functional Tests are problematic for two reasons. First, it is widely acknowledged that they test knowledge at a 6th-grade level, not the higher standards Maryland has adopted for its high school students. Second, there are few data to show their value or impact since implemented in the early 1980s. There has never been a published evaluation of the impact of the graduation requirement on retention rates and high school graduation rates. The 1989 *Report of the Governor's Commission on School Performance* (Sondheim Report) cited a 1987 graduation rate of 75% for the first class to graduate under the requirement, and according to Maryland State Department of Education (MSDE) calculations, Maryland has shown increasing rates of high school graduation since then. Yet recent reports by both the Manhattan Institute and the Maryland Kids Count Partnership show significantly lower reported graduation rates for Maryland.^{8/9}

**Percentage of entering Maryland 9th grade students
graduating 4 years later**

	1987	1990	1996	1998	2002
MSDE	75		79	80	84
Manhattan Institute				75	
Maryland Kids Count Factbook		73	75	73	74 (2000)

While passing rates on the reading and math Maryland Functional Tests rose during the first decade, data have demonstrated a pattern often seen in high-accountability tests: increased scores in early years, followed by a plateau in performance.¹⁰

⁷ The Citizenship test was deleted as a graduation requirement effective for 1999 graduates.

⁸ Greene, J. High School Graduation Rates in the United States. Manhattan Institute. April 2002 Revision.

⁹ *2003 Kids Count Factbook for Baltimore City*. Maryland Kids Count Partnership, Baltimore City Data Collaborative, The Safe and Sound Campaign, and the Family League of Baltimore City. March 2003.

¹⁰ Linn, R.L. (2000).

Maryland Functional Tests
Percentage of 9th grade students passing

9th grade	1984	1987	1990	1993	1996	1999	2002
Reading	93	92	93	97	97	97	97
Mathematics	61	66	68	79	83	85	82
Writing	NA	NA	88	97	97	97	97

The percentage of 11th graders passing all Functional Tests has also stabilized, despite the 1998 elimination of the citizenship test that reduced the number of tests students had to pass from four to three.

Maryland Functional Tests
Percentage of 11th grade students passing all Functional Tests

11th grade	1984	1987	1990	1993	1996	1999	2002
Reading, Mathematics and Writing	NA	NA	NA	93	92	92	92
Number of tests required	4	4	4	4	4	3	3

National studies on minimum competency exams have found that, by and large, these assessments have not brought about the envisioned changes in public education.¹¹ Data about Maryland's experience come largely from external sources—because the state did not collect comprehensive, district-level data until the 1990s—and appear to mirror these national findings. The controversial Amrein and Berliner (2002) study looks at the impact of high-stakes testing on student performance and finds that student achievement in Maryland decreased following introduction of the Functional Tests, as measured by SAT, ACT, and AP test participation and scores. Yet it remains unclear how many Maryland 12th graders are actually denied a diploma as a result of not passing all Functional Tests. Amrein and Berliner claim that approximately 4 percent of Maryland's 12th graders were barred from graduation in 2000 for failing to meet test requirements;¹² meanwhile, Gary Heath, Assistant Superintendent for MSDE's Division of Planning Results and Information Management, maintains that this number is not known.¹³

Raising the Floor: Demand for Higher Standards in the Early Grades

A growing clamor for higher standards in the 1980s led then-Gov. William Donald Schaefer to appoint the Commission on School Performance in 1989. That same year

¹¹ U.S. Office of Technology Assessment, 1992.

¹² Amrein, A. and Berliner, D. (2002). Educational Policy Archives, p19.

¹³ Telephone interview, May 2003.

the commission issued the Sondheim Report, which called for a new accountability system to hold schools and districts responsible for greater rigor and equality, and thereby push K-12 students beyond minimum-competency levels. Specifically, the Commission recommended that Maryland implement a more demanding, comprehensive, and coordinated system of assessments, performance standards, and accountability, paving the way for various programs that have since emerged in grades K-8:

- the Maryland Learning Outcomes curriculum standards [in grades K-8];
- the Schools for Success program that expanded support to early education and low-performing schools;
- the now-defunct Maryland School Performance Assessment Program (MSPAP) of elementary and middle school testing; and
- the Maryland School Performance Program, a comprehensive system of indicators, monitoring and interventions to measure and improve school performance.

Focus Shifts to High Schools

The Sondheim Report also had implications for reform at the high school level; its recommendations for standards-based tests were the impetus for the High School Assessments.

In the mid-1990s Maryland increased the number and level of courses required for high school graduation, most notably in mathematics and science starting with the class of 1997. To graduate, students had to pass four, 6th -grade-equivalent Functional Tests and earn a requisite number and type of course credits. Even today, however, there is no instrument to determine whether teaching and learning in these courses has met the same academic expectations from school to school, and/or district to district. High school accountability continues to be based on school completion, dropout rates, attendance, and passing the Functional Tests—a far cry from the high standards the Sondheim Report had envisioned.

The Sondheim Report recommended new criterion-referenced¹⁴ tests to measure academic standards within an overhauled state accountability system—tests that would measure “important, fundamental things in the curriculum” with “more than a narrow, low level set of objectives” in a manner that would “not encourage rote teaching of material covered by the test.” These tests would also replace the Maryland Functional Tests as a demonstration of student competency for graduation.¹⁵

¹⁴ Criterion-referenced tests are tests that measure student achievement against a specified body of knowledge. They are typically contrasted with norm-referenced tests which measure student performance in comparison with the other students who took the same test.

¹⁵ *The Report of the Governor’s Commission on School Performance*, August 1989, p. 8-10, 15-17.

With the launch of the MSPAP well underway in 1996, the State Board of Education approved K-12 learning standards in English, math, science, and social studies and the development of new high school assessments to measure student progress toward these “Core Learning Goals.” The HSA program was intended to increase expectations for academic performance at the high school level by tying new, rigorous assessments to high school graduation. In MSDE’s own words, it was intended to ensure that students “have the higher level skills demanded by the modern workplace and higher education and to gauge both school performance and individual student performance, providing valuable information to students, parents, and school improvement teams.”¹⁶ MSDE has published its intended purposes for the High School Assessment Program:

- to increase academic rigor;
- to enhance the value of the Maryland high school diploma; and
- to ensure that graduates have skills and content knowledge.¹⁷

With this as a framework, it was envisioned that the HSA program would make coursework more rigorous and hold students accountable through, in the language of standards-based reform, a test “worth teaching to.” A rarely stated premise of the program as it has developed is that high-stakes testing can and should improve high school instruction and learning for all students, particularly in content areas and among students where achievement has typically lagged.

How the HSA Program Evolved

Developing Standards to Increase Rigor

Approved in 1996, Maryland’s Core Learning Goals outline the skills and knowledge that Maryland high school students are expected to master, and thus provide the basis for all state assessments. MSDE believed this attempt to “raise the floor” would provide students at poorly performing schools with a more demanding curriculum, without inadvertently imposing an artificial ceiling on students who were high achievers.¹⁸ Core Learning Goals in each of the four major academic subject areas provided an outline for the course of study for the entire high school program, rather than for individual courses or grades. In theory, this approach would allow individual districts and schools to develop their own curricula and formative assessments to meet the Core Learning Goals. As implementation progressed, however, more direction was deemed necessary.

¹⁶ Maryland State Department of Education, “Raising Expectations for Maryland Students,” http://www.mdk12.org/mspp/high_school/what_is/raising_expectations.html

¹⁷ Maryland State Department of Education, *Maryland High School Assessments*, October 2001, <http://www.msde.state.md.us/hsimprovement/pdf/HSAPPOct2001v1.ppt>

¹⁸ Telephone Interview with Ronald Peiffer, MSDE Deputy State Superintendent for Academic Policy, October 24, 2002.

In 1998 the State Board of Education, concerned that some schools and school districts were not able to prepare students to pass the HSAs, asked MSDE to develop an “academic intervention plan” and more specific content standards. Again, the original intent of the content standards was to help all schools and teachers—but particularly those with fewest resources—by detailing what was to be taught, when, at each level of schooling, thus drawing a road map for meeting standards and identifying the need for early intervention.

How High Are Maryland's Standards?

In standards-based reform, the assessment can only be as effective in raising the bar as the standards themselves. In 2001, MSDE commissioned Achieve, Inc., a national organization that is clearly pro-standards-based reform, to conduct a general “policy review” of Maryland’s reform strategies. Achieve’s review praises Maryland’s overall efforts, but acknowledges that it did not conduct a full analysis of Maryland’s standards. It notes that Maryland’s content standards, initiated in part because of criticism by the national organizations cited below, “represent a significant improvement from the learning outcomes” and goals.¹⁹

Two other evaluations of Maryland’s standards find that, in general, they excel in specificity and clarity. In its *Quality Counts 2002* state report card, *Education Week* awarded Maryland its highest score of 98 for standards and accountability, based on the following criteria: clear and specific standards in core subjects, a regular timeline for revising standards, a variety of types of test items, the use of criterion-referenced assessments aligned to state standards, and a wide range of accountability measures. In its *Making Standards Matter 2001* report, the American Federation of Teachers also awarded Maryland high marks for the clarity and specificity of its standards, performance indicators, and state-developed tests based on standards. The group did, however, find the standards lacking in grade-specific instructional strategies, lesson plans, and resources to help teachers align instruction to standards.

Meanwhile, in its *State of State Standards* report issued in 2000, the Thomas B. Fordham Foundation was more critical of Maryland’s system, even though it did increase Maryland’s overall rating from an F in 1998, before the content standards were in place, to a C+ in 2000. In general, Fordham—which conducts more detailed reviews than other organizations—advocates standards that are specific and measurable, and that emphasize content, comprehensiveness, and rigor. With regard to Maryland, it placed the state among those with “shaky foundations” where vague standards and high stakes accountability can lead to difficulty with implementation in the classroom.

¹⁹ Achieve, Inc., *Aiming Higher: The Next Decade of Education Reform in Maryland*, January 2002.

Developing the HSA Tests

One of Maryland’s earliest decisions in developing the HSA program was to make its tests end-of-course exams rather than the more global, standards-based competency tests favored by many states. This decision has significant implications.

Standards-based exams are aligned with a generic set of standards across a subject area, rather than specific course content.²⁰ Because standards-based exams are aligned with state standards at the high school level rather than the middle school level, they are deemed more rigorous than the earlier generation of minimum-competency exams, such as Maryland’s Functional Tests. Standards-based exams are usually offered for the first time in 10th or 11th grade. They tend to focus on language arts and math skills (although science and social studies are increasingly being added), and test a range of skills and knowledge that students encounter during their high school career.

End-of-course exams, on the other hand, are administered to assess what students have learned in a specific course. By adopting these, Maryland made the choice to test students immediately upon completion of key academic courses, beginning in 9th grade or earlier.

MSDE tapped CTB/McGraw-Hill, a private testing company that reportedly accounts for 40 percent of the test design market, to develop the HSA tests. The process has worked thus: since new test items must continually be developed and tested for each new round of testing, CTB/McGraw-Hill trains high school teachers to develop test questions with MSDE’s oversight. Expert content teams composed of secondary and post-secondary educators refine the questions, then items are field-tested to define the level of difficulty based on the proportion of students who correctly answer each question. Finally, testing professionals ensure that the tests meet validity standards (whether they accurately measure student knowledge in a given content area) and reliability standards (the consistency with which they generate the same responses).

How the HSA Program Looks Today

Current HSA Tests and Subjects

Under MSDE’s original plan, 12 end-of-course exams would accompany high school level courses in the four major academic disciplines—English, social studies, science, and math. Students would have to take and pass 10 (with a choice of two of the four science HSAs) to graduate. Thus far exams have been developed for English I, Government, Biology, Algebra/Data Analysis and Geometry, with students taking them upon course completion, typically in either 9th or 10th grade although some students

²⁰ It is important to note that end-of-course exams such as Maryland’s are also standards-based in the sense that they are designed to align with state standards.

take Algebra I as early as 7th or 8th grade. (The Geometry HSA test has also been adopted to fulfill the NCLB 10th grade math requirement; what accountability it will have as an HSA test remains to be clarified.)

Maryland’s High School Assessment Program

Subject Area	Course Credits Required	HSA Tests in Place to be Required for Diploma	HSA Tests Projected
English	4	English I	English I English II English III
Mathematics	3 Algebra I Geometry Other	Algebra/Data Analysis Geometry (Used for NCLB 10th grade math test)	Algebra/Data Analysis Geometry
Science	3 Including one lab experience	Biology	Take two of the following: Earth/Space Science Biology Chemistry Physics
Social Studies	3 Government World History US History	Government	Government World History US History

Test Format

Maryland adopted a mixed-item format for the HSA tests, incorporating both “selected response” (multiple choice) items and “brief” or “extended constructed response” items (open-ended; requiring a written response by the student). This way tests can be graded more quickly and objectively than the performance-oriented MSPAP tests (though not as quickly and objectively as an entirely multiple choice format), while also requiring students to demonstrate critical analysis and evaluation skills.

According to Ron Peiffer, Assistant Superintendent, MSDE’s Office of Academic Policy, the more multiple-choice items, the more efficient the grading process and the faster the results, by a margin of four to six weeks. Such a format would be more objective

and less expensive as well. Yet Maryland opted to include a substantial proportion of constructed response questions out of questionable concern that a purely “objective” test would force instruction to focus on factual information and rote learning, versus higher-level thinking skills.

Samples of Question Types from the Algebra 1/Data Analysis Assessment (MSDE)

Selected Response Question

Mary graphed the system of equations below.

$$y = \frac{3}{2}x + \frac{7}{2}$$

$$y = -\frac{2}{3}x + \frac{7}{3}$$

Which of these best describes the relationship between the two lines?

- A. They have no point in common
- B. They have one point in common
- C. They have two points in common
- D. They have infinite points in common

Extended Response Question

The table below shows the sales for a greeting card company for the years 1990 through 1998

Year	Sales (in thousands)
1990	\$205
1991	\$230
1992	\$245
1993	\$270
1994	\$295
1995	\$320
1996	\$340
1997	\$350
1998	\$365

Complete the following in the Answer Book:

- Write an equation for the line of best fit for this data. Let x represent the years since 1900 and y represent the sales, in thousands of dollars. (If you choose to draw a graph, use the grid provided in the Answer Book.)
- According to your equation, what were the sales in 1999? Use mathematics to explain how you determined your answer. Use words, symbols, or both in your explanation.

Test Scoring

Maryland contracts with a private company, Measurement Inc., to score the tests. While multiple choice items are scored mechanically, the open-ended responses and essays are scored by pairs of educators using set criteria to judge student answers; if a pair disagrees on a student's score, the item in question is turned over to a supervisor to score. According to estimates by the U.S. General Accounting Office, a test with only multiple-choice items that can be scored by computer is one-third the cost of the mixed-item test that Maryland has adopted.²¹ Based on current costs, estimated at \$24 per student per HSA test,²² this shift could save Maryland more than \$2 million annually at the onset and savings would increase exponentially if HSAs are added.

Testing Schedule

The HSA tests are offered twice a year: in May and in January, in order to accommodate schools on a block schedule where students take four courses each semester (eight courses in a year). Summer administration of the tests is in the offing, but more frequent administration may be required.

Reporting of Test Scores

Once the HSA program becomes fully operational, Maryland expects to release student scores from the mid-May testing in late August. This timeline, adopted in part to comply with federal NCLB regulations, will represent a significantly shorter timeline for results than during the 2001- 2002 school year, the first year the tests were officially administered, when results of May testing were not released until December 2002. Ideally, the HSA would play an integral role in determining student promotion and other instructional decisions. This scoring procedure will not, however, allow schools to use the exam for promotion or course grade purposes, or permit students to take a remediation course and retest during the summer.

Accountability

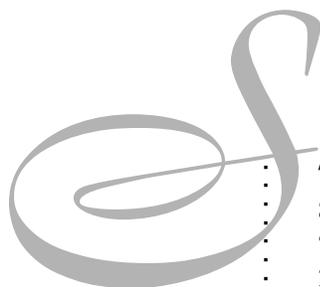
Student accountability: Students currently enrolled in HSA courses are required only to take, not pass, the HSAs to graduate. While scores are reported in percentile rank form on student transcripts, passing the HSAs is not currently a graduation requirement, nor has the State Board of Education set proficiency standards or passing scores for HSA performance.

School and district accountability: The Maryland Report Card that the state currently uses to evaluate high schools is based on attendance rates, dropout rates, and Maryland Functional Test pass rates. As it phases out the Functional Tests, the State

²¹ Title I: Characteristics of Tests Will Influence Expenses; Information Sharing May Help States Realize Efficiencies GAO-03-389, May 8, 2003.

²² Estimate provided by Ron Peiffer, MSDE, April 16, 2003.

IV. High-Stakes Testing: An National Challenge



The High-Stakes Trend

State testing and accountability programs in general—and what is now known as “standards-based reform”—originated with the 1983 publication of *A Nation at Risk: The Imperative For Educational Reform*. This report, commissioned by then-Secretary of Education T. H. Bell and authored by the National Commission on Excellence in Education, raised the red flag that students in the United States were falling behind their international peers. What followed was a wave of state education reform efforts that centered on setting higher standards for student learning, measuring student progress against those standards, and holding administrators, educators, and students accountable for performance.

In the early years of standards-based reform, many states, including Maryland, adopted “minimum competency exams” that sought to ensure that all students acquired basic skills in math, reading, writing, and, in some cases, other subject areas such as citizenship and science. Over time, standards-based reform has emphasized the twin goals of increasing academic rigor and ensuring equal educational opportunity—that “the only way to be fair to all students is to expect the same thing of all students.”²³ As a result, New York State, which had traditionally offered a two-tiered testing and diploma system, is now moving to adopt a single Regents exam that all students must pass in order to graduate. Other states, including Maryland, are eschewing minimum competency tests in favor of end-of-course or standards-based exams.

While the individual states have distinct strategies and policies driving their high school reforms, the common trend is clearly toward greater student, school, and district accountability. Eighteen states currently require students to pass exit exams before they graduate from high school; within the next six years, such tests are expected to be required for graduation in at least 24 states.²⁴ More specifically, of the 24 states profiled, 10 had minimum competency exams, seven had standards-based exams, and two had end-of-course exams; by 2008, only three of the 24 states will use minimum competency exams, 16 states will have standards-based exams, and five will use end-of-course exams.²⁵

²³ “Making Standards Work: An Interview with Lecturer Paul Reville.” *Harvard Graduate School of Education News* (March 21, 2002).

²⁴ Center on Education Policy (CEP). (2002).

²⁵ CEP. (2002). CEP excludes from its analysis states that have tests but do not require them statewide, that have differentiated diploma systems, or where student accountability is not a part of the system.

Because this new generation of high standards, high-stakes testing is still very much in its infancy, evidence of its impact is mixed and inconclusive, providing few, if any, clues to how Maryland's HSA program should proceed. But in making key decisions about the new tests, Maryland cannot afford to ignore what is known today (and accumulating daily) about how high-stakes testing can best be implemented to increase student learning. Given that the various states are at different points along the standards-based reform continuum and have faced a range of consequences and challenges, their experiences would be instructive to the State Board of Education as it seeks to advance the HSA program this summer.

Furthermore, as more and more states have entered the high-stakes testing arena, numerous studies and reports with conflicting evidence give the Board even more fodder for its decision-making this summer and beyond. It is clear from the lack of concrete answers to the questions posed to and by the Board that high-stakes testing is still an arena for pioneers.

The Graduation Requirement Dilemma

Maryland is not alone in delaying its new graduation requirement, nor is it to be admonished for doing so, given the potential impact it will have on many students. Indeed below is a sampling of states at similar impasses in implementing graduation requirements:

- **Alaska** has postponed requirement due to high failure rates.
- **Arizona** has postponed implementation of the requirement four times.
- **California** has just decided to postpone its requirement for two more years.
- **Florida** lowered the passing scores for math and reading, and waived requirement for special education.
- **Massachusetts** also lowered scores for passing and issued waivers.
- **Michigan** offers multiple types of diplomas for different levels of achievement.
- **New York** has added a waiver/appeal process.
- **Wisconsin** has made the requirement optional for school districts.

Different Impact on Different Students

Data show that low-income, minority and otherwise disadvantaged students have traditionally performed worse on high-stakes tests than their white, more affluent counterparts. During the early years of minimum competency tests in the 1970s, 20% of African American students failed graduation tests in the state of Florida and were denied high school diplomas, compared to 2% of white students. More recently, researchers have asserted that high-stakes testing not only hurts and holds back low-income and minority students, but that it is also linked to high dropout rates among these groups. According to Gary Orfield, professor of education and social policy at the Harvard Graduate School of Education and Kennedy School of Government and co-director of Harvard's Civil Rights Project, African Americans and Hispanics are three to four times more likely to be held back as a result of high-stakes test scores than whites. Tests, Orfield warns, can lead to "the punishment of innocent victims of unequal education."²⁶

Nationwide, outcomes have cultivated opposition to high-stakes testing, and support for the notion that standards that are too high lead to large failure rates. Court challenges to high-stakes testing have argued that this type of testing unfairly penalizes groups of students who have not received an adequate education or opportunity to learn. The landmark case, *Debra P. v. Turlington* ruled that such tests are legal, but must be a fair measure of what students have been taught; that students must have adequate advance notice of high-stakes testing requirements; and that the burden of proof is on the state to show that students are being taught the appropriate material.²⁷ This legal "test" of high-stakes has become known by the phrase "opportunity to learn." Ensuing court decisions have validated the premise that not only must test questions correspond to the required curriculum, but tests must correspond to material that was actually taught to the students. States have met this mandate simply by showing a strong connection among content standards, the courses students take, and the testing system.

A legal challenge to the new high-stakes Massachusetts Comprehensive Assessment System (MCAS) exam and its use as a graduation requirement, now before a U.S. District Court, argues that minority, disabled, bilingual, and vocational students, and students in the Holyoke school district have not had adequate time and opportunity to learn. The MCAS exam is a high-stakes test that public school students must pass to receive a high school diploma. Beginning with the Class of 2003, all Massachusetts

²⁶ American Youth Policy Forum Presentation. *Impact of High Stakes Testing Policies on Minority and Disadvantaged Students*. Washington: June 7, 2000. Retrieved from <http://www.aypf.org.forumbnets/2001/fb0107700.html>.

²⁷ *Debra P. v. Turlington*, 474F.Supp.244(M.D.Fla, 1979), aff'd in part, rev'd in part; 644F.2d397(5th cir,1981), on remand; 564F.Supp.177(M.D.Fla,1983), 730F.2d1405(11th Cir,1984), aff'd. <http://www.faculty.piercelaw.edu/redfield/library/Pdf/case-debrap.dct.pdf>

In Massachusetts, meanwhile, where activists blame MCAS for fueling the dropout rate, education officials point to a declining dropout rate between junior and senior year (from 3.9% to 3.6%) as evidence that the graduation requirement is helping students stay in school.³¹ Critics, however, contend that the more likely factor is pressure in earlier grades to retain under-performing students before they drive up schools' failure rates or drive down measures of progress required for state and federal accountability. In Boston, 16% of the Class of 2002, which was not subject to the MCAS exam, was retained in 9th grade, compared to more than 25% of the Class of 2004, which must pass the MCAS to graduate.³² In theory, identifying failing students early can push schools to help those students meet standards, but research has shown strong evidence that retention in a school grade is more likely to lead to dropping out.³³

Cut Scores: A Volatile Mix of Science and Politics

One of the most critical—and politically charged—decisions state educators make in designing high-stakes tests is where to establish the initial passing, or “cut,” score. Unless standards are set fairly low there will likely be many students who do not initially meet them, raising the question: Should standards be set low so fewer students fail (a more politically acceptable scenario than massive failures) at the expense of increased rigor? Or, should standards be set high to increase academic rigor, with the likely result that a politically unacceptable number of students will fail?

To illustrate the controversy, when New York State set its initial Regents scores low to avoid massive failures at disadvantaged schools, parents in the more affluent town of Scarsdale complained that it caused teachers “to narrow and simplify instruction.”³⁴ The state proceeded with its plan to raise its original cut score of 55% to 65% over two to four years. The impact of this incremental increase in the pass rate is seen in the percentage of students who actually passed the lower initial cut score in the first year versus those who would have passed under the longer-term goal. As a result of this phased-in stakes, 72% of students in schools with largely minority students passed in the first year versus 44% who would have passed had the initial rate been set at 65%.

³¹ Kevin Rothstein, “MCAS credited for state’s decreasing dropout rate.” *The Boston Herald*, (February 14, 2003). BostonHerald.com

³² Rothstein.

³³ Cites in Dan Goldhaber, “What Might Go Wrong with the Accountability Measures of the ‘No Child Left Behind Act?’” Paper prepared for *Will No Child Truly Be Left Behind: The Challenges of Making This Law Work, A Conference sponsored by the Thomas B Fordham Foundation* (February 13, 2002), p. 6. Also cites in Christopher Edley Jr. and Johanna Wald, “The Grade Retention Fallacy,” *Boston Globe* (12/16/02), p. A17.

³⁴ Lynn Olson, “States Turn to End-of-Course Tests to Bolster High School Curriculum,” *Education Week* (June 6, 2001).

Percentage of general education students from the 1996 cohort who passed the Regents examination in English after 4 years

	Current pass score 55 or higher	Eventual pass score 65 or higher
General Education students	90	75
Students with disabilities	63	36
Students in schools with 0-20% minority populations	98	88
Students in schools with 81%-100% minority populations	72	44

Source: Center on Education Policy, 2002: based on information from <http://www.nysed.gov> (July 2002)

Many states have taken a similar approach to phasing in higher standards and/or revising cut scores over a period of time. Virginia adopted cut scores in Fall 1998 and, after the first administration of the test, showed that only 2% of schools met accreditation requirements; following the second administration, only 7% met standards. Public alarm ensued, and state education officials made mid-course corrections; for example, the history exam was amended after teachers complained about the amount of content being tested. Acknowledging the difficult standards of its own exit exam, Massachusetts adjusted its cut score so that students have to answer only 40% of the questions correctly to pass. Texas, meanwhile, raised cut scores for passing from 60% to 70% over a two- to three-year period.

The Narrowed Curriculum

As illustrated in Scarsdale, critics of statewide testing have charged that it has a profound effect on the curriculum: what is taught and how it is taught. A new study by J. J. Pedulla et al. (2003) finds that state testing programs have caused a “narrowing of the curriculum,” particularly in states with high-stakes for students and schools.³⁵ 80% of teachers in these states report that the pressure to produce high scores means that they teach little content beyond the test. Such unintended consequences of large-scale policy at the classroom level are often the most difficult to identify and the most critical to success.

³⁵ Pedulla, J., Abrams, L., Madaus, G.F., Russell, M.K., Ramos, M.A. and Maiao, J. (2003). "Perceived Effects of State-Mandated Testing Programs on Teaching and Learning: Findings from a National Survey of Teachers." Chestnut Hill, Mass: National Board on Educational Testing and Public Policy, Boston College.

·

· **Student Stakes and Performance**

· The student accountability of many high-stakes exams is based on the premise that
· student, and thus school, performance will increase with higher levels of individual
· stakes. John Bishop (1998) notes that student performance in countries with medium
· and high stakes for students is higher than in countries without stakes. He finds, for
· example, higher overall achievement levels in New York State with its high-stakes
· Regents exams than comparable states without such tests. The Achieve Benchmarking
· Institute points to Massachusetts, where 82% of the Class of 2003 passed MCAS
· English and 75% passed math on the first try once it became a graduation require-
· ment, versus 66% and 55% respectively, before the stakes were attached.³⁶

· Yet an older study cautions policy makers about the impact of these stakes on poorly-
· performing students. Kellaghan, Madaus, and Raczek (1996) posed the questions: “Do
· high-stakes tests motivate all students?” and “What do tests motivate students to do?”
· They found that high-stakes testing can actually undermine motivation, especially
· among students who are struggling with school.

·

· **Testing Models and the Need for Alternatives**

· *Conjunctive Testing*

· The American Educational Research Association has asserted that no decision or char-
· acterization that will have a major impact on a student should be made on the basis
· of a single test score.³⁷ Although students have numerous opportunities to take and
· pass high-stakes graduation tests, the bottom line remains that they must pass a sin-
· gle test (or series of tests) to graduate, despite performance on other measures.
· Others argue that various tests and grades factor into graduation decisions; while this
· is true, it ignores the “conjunctive” nature of the graduation model—that is, students
· must perform on all indicators to graduate, such that failure on any single measure
· can prevent graduation.

· Using end-of-course exams as graduation requirements is a conjunctive model, and as
· such, it raises concerns. Without a careful balance between school and student
· accountability, students in poorer areas may be less prepared, academically and other-
· wise, to pass these very specific content exams. For students who fail several tests
· early on, the high school years may amount to a cycle of constant test re-taking, try-
· ing to catch up. Critics also question whether the high level of detail in end-of-course
· exams is a fair measure of what students really need to know and do as a result of
· high school education.³⁸ Should a student who cannot pass a statewide chemistry test
· be denied a diploma despite otherwise satisfactory course grades?

· ³⁶ “Aiming Higher: The Next Decade of Educational Reform in Maryland.” Achieve’s
· Benchmarking Institute, January 2002, p.37-38.

· ³⁷ American Educational Research Association et.al., 1999, Standard 13.7, p.146

· ³⁸ Lynn Olson, “States Turn to End-of-Course Tests to Bolster High School Curriculum,”
· *Education Week* (June 6, 2001).

There are also good reasons for using end-of-course exams to improve the quality of education. A range of tests provides schools the opportunity to cover content areas beyond reading and mathematics. Perhaps more importantly, they can serve to highlight inequities in course curricula and instruction between schools in poor areas and wealthier areas, and provide an impetus for rectifying these.

Compensatory Testing

A “compensatory” testing model, on the other hand, allows poor performance on one measure to be offset by good performance on another. In the high-stakes environment, this could translate into alternative modes of evaluation that give students more flexibility in attaining a diploma by providing a safety valve. In reality, unacceptable failure rates on high-stakes tests have compelled many states to seek alternatives to graduation.

Providing Alternative Pathways to Diplomas

Requiring passing scores on multiple end-of-course exams for graduation raises serious concerns about how students who fail will have adequate opportunity to relearn and retest. The state of Massachusetts launched its graduation requirement with the Class of 2003 without alternative options in place. Though only two standards-based exams (English and math) are required for graduation, students have had five opportunities to pass the exams, starting in the 10th grade. As of May 2003, three years into the requirement, 8% of Massachusetts’ high school seniors had not yet passed the high-stakes test a month before their planned graduation dates.

Many states have taken steps to avoid what happened in Massachusetts and ensure that students all along the achievement spectrum have alternative means of earning diplomas. The following strategies have been implemented to accommodate students with multiple test failures as well as those at the highest levels of achievement:

- **Provide multiple options to retake the test several times annually during and after leaving high school.** In many cases, students are permitted to simultaneously pursue a higher-level course in the designated discipline. Some states with graduation requirement exams, such as Massachusetts, allow students to retest anywhere from two to 11 times, some up to the age of 21.³⁹
- **Offer a lower-level diploma** such as a Certificate of Completion. Alabama, Georgia, New York, New Mexico, North Carolina, and Vermont do this now.
- **Offer a proficiency assessment** for those who have failed high-stakes exams. Arizona has plans for a comparable equivalent demonstration test titled Arizona’s Instrument to Measure Standards.

³⁹ CEP, p.48

-
- **Establish a waiver or appeals process.** Students and their parents can make a case for graduating based on grade point average or other specified criteria in Indiana, Minnesota, and Mississippi. Massachusetts has granted more than 1,000 waivers to 12th-graders in its first year.
 - **Provide multiple accommodation options for students with disabilities and English language learners** (a majority of states, including Maryland).
 - **Substitute an alternative curriculum in lieu of tests.** Indiana’s Core 40 curriculum is a rigorous high school-level curriculum that can be substituted for traditional graduation requirements provided students pass each course with a “C”.
 - **Allow substitute test scores.** Some states accept other test scores, though this typically applies to high-achieving students. In New York and Virginia, students can substitute Advanced Placement, SAT II, and International Baccalaureate scores for relevant content areas on exit exams, thus relieving concerns about tests that “dumb down the curriculum.”

V. New Federal Mandate Complicates the HSA



The various requirements and timelines imposed on states by the No Child Left Behind Act of 2001 have created yet another set of issues for Maryland's State Board of Education to consider as it weighs the future of the HSA program this summer.

Signed into law by President George W. Bush in January 2002, NCLB aims to raise achievement among all students and eliminate gaps among different groups of students—not unlike the goals of Maryland's HSA program. Moreover, NCLB is designed to allow states flexibility in setting standards and designing assessments that meet those standards. NCLB is less flexible, however, in its goal to hold schools and districts accountable for ensuring that no child and no group of children categorized by race, income, or ability level—is left behind.⁴⁰

Assessments: Maryland on Track

In the area of assessment, it appears that Maryland is on track to satisfy NCLB requirements at the high school level.⁴¹ By 2005-2006, NCLB will require states to assess all high school students in reading and math once in the 10th, 11th or 12th grades; beginning in 2007-2008, states must administer a science test at least once during that same time frame. Unlike MSPAP (which has been replaced by the new Maryland State Assessment Program in the elementary and middle grades), the HSA tests were designed to report individual student data with test results disaggregated by race, ethnicity, income, and special services as required under the new federal law. Maryland uses the existing HSA in Geometry to meet the NCLB 10th-grade math mandate and plans to use the Biology HSA in 2005 to meet NCLB's science requirement—although in both cases the test may be administered before the 10th grade. To comply with the federal high school reading requirement, Maryland has again contracted with CTBS/McGraw-Hill to create the NCLB reading test that was administered to 10th graders for the first time in March 2003.

Accountability Means Progress at Aggressive Pace

For Maryland and most other states, the greatest potential impact of NCLB is likely to be in the area of school accountability. States must now define standards in reading, math and science that all students must reach by the end of the 2013-2014 school year. Schools must also set "average yearly progress" goals and test students regularly to bring groups categorized by race, ethnicity, income, disabilities, and English proficiency up to those standards at designated intervals. Groups that perform poorly on these assessments early on will have farther to go to meet the long-term standards, and far more aggressive targets each year as a result.

⁴⁰ Neill, M. "High Stakes, High Risk," *American School Board Journal* (February 2003).

⁴¹ No Child Left Behind. Education Commission of the States. Retrieved from <http://nclb.ecs.org/nclb/>.

Schools and districts where any group of students fails to meet targets on either math or reading for two consecutive years will be designated as “needing improvement.”⁴² Most likely to fall into this category are Title I recipients,⁴³ more than half—58%—of the nation’s public schools. For their failure to stay on track, these schools will be subject to corrective action such as revising instruction, receiving technical assistance, providing public school choice, offering students supplemental education services, reorganizing or replacing staff, and restructuring the school—all of which become more severe if schools fail to show improvement over multiple years.⁴⁴ After three years of failure to make yearly progress, for example, schools must allow students to use Title I funds for private tutoring and other supplemental services. Meanwhile, in some cases, districts have already been forced to offer transfer options to students in failing schools.

Accountability May Also Mean More Failures and/or Lower Standards

Because NCLB requires that all groups of students show annual progress, many state educators anticipate large numbers and ranges of schools will be classified as “needing improvement,” even those in relatively affluent areas. According to a simulation by the Council of Chief State School Officers using data from eight states, between 49% and 88% of schools would be categorized as needing improvement after two years.⁴⁵ As a result, some state education officials are warning that NCLB’s lack of flexibility in assessing performance standards will eventually lead to a majority of American schools being defined as “needing improvement.”

A more immediate concern is that states with truly challenging standards may have the largest number of under-performing schools, because their students will have farther to go to become “proficient.” According to Robert Linn, Distinguished Professor of Education at University of Colorado, “The ambitious adequate yearly progress targets required by NCLB . . . will pose major challenges to all states, but the challenges will be greatest for states that have adopted or are putting in place rigorous assessments aligned with challenging content standards with demanding performance standards for students to be identified as proficient.”⁴⁶

⁴² Center on Education Policy, *From the Capital to the Classroom: State and Federal Efforts to Implement the No Child Left Behind Act* (January 2003). There are “safe harbor” provisions that secondary schools can be considered to make AYP if the percent of students below proficiency is at least 10% less than the previous year and the school makes its target for improved graduation rates.

⁴³ Business Week, 11/25/2002.

⁴⁴ Center on Education Policy, 2003. States may elect to apply these strategies to non-Title I schools.

⁴⁵ Center on Education Policy, 2003, p.30. Citations- Smole and Riddle, 2001; CCSSO, 2002b.

⁴⁶ “NCLB: Interview with Robert Linn about Assessment,” National Governors’ Association, www.nga.org/common/issue/BriefDetailPrint/1,1434,4154,00.html.

The concern isn't unfounded. Of the 8,652 Title I schools identified in 2002 as having failed to make progress on test scores for two consecutive years, the greatest number—1,513—were located in Michigan, considered by experts to be among those states with the highest standards.⁴⁷ As a result, Michigan recently adjusted its NCLB targets, lowering from 75% to 42% the percentage of students who must pass its English test in order to “certify” a high school.⁴⁸ By contrast, Maryland had 118 Title I schools that failed to make “average yearly progress.”⁴⁹

Under NCLB, High School Completion Rates Count

Upping the ante even more, NCLB requires that each group of students within a school meet average yearly targets on graduation rates, such that a school where a specific subgroup makes its assessment target can still be designated as “needing improvement” if that group does not meet the graduation target.⁵⁰ In other words, states must ensure that their high-stakes assessment system does not negatively impact the rate at which students graduate from high school. This means schools must find solutions to the current practices of pushing failing students out the door or retaining them in lower grades, given the effect such actions would have on dropout and graduation rates.

The concern, even among high-stakes testing proponents, is that the federal law will have a dampening effect on states that were attempting to raise standards, providing them instead with incentives to set more attainable progress targets and low performance standards overall. Connecticut, for example, set proficiency standards so low that 80% of students are slated to pass its tests, and Wyoming's goals ensure that no school will be designated as “needing improvement.”⁵¹ U.S. Secretary of Education Roderick Paige acknowledged this trend in an October 2002 memo admonishing district administrators who sought to lower their standards.⁵²

Clearly, lower performance standards on the HSAs will lead to fewer Maryland schools designated as “needing improvement”. Thus, NCLB offers strong incentive—atop existing concerns about disproportionately low achievement among minority, disadvantaged, and special education students—to set low HSA performance standards.

⁴⁷ Dillon, Sam. *The New York Times*, 2/16/03.

⁴⁸ Dillon, Sam. *The New York Times*, 5/25/03.

⁴⁹ USDOE, “Paige Releases Number of Schools in School Improvement in Each State,” <http://www.nochildleftbehind.gov/media/news/070102.html> (July 1, 2002).

⁵⁰ Center for Education Policy, 2003, p. 23.

⁵¹ Alexandra Starr, “Why Johnny Can’t Fail,” *Business Week* (November 25, 2002).

⁵² US Department of Education. “Letter Released from US Education Secretary Paige to State School Chiefs on Implementing No Child Left Behind Act.” press release (October 23, 2002).

Many Rules, But Little Funding

Heightened by the uncertain economy, a very real concern in Maryland is that implementing a graduation requirement while retaining high standards could result in a large number of schools being categorized as “needing improvement.” Under NCLB, Maryland would potentially have higher numbers of failing schools and greater responsibility for providing educational resources to improve these schools, yet no additional funding to do so. Despite its many and difficult mandates, NCLB allocated just \$490 million a year nationally in federal funding to cover state testing and other costs. A U.S. General Accounting Office report estimates that Maryland will receive an NCLB appropriation totaling \$35 million for K-12 assessments between fiscal years 2002 and 2008. The report further projects that Maryland will spend \$91 million on state testing during that same time frame, based on its choice of format for assessments: a combination of multiple choice and open-ended response items,⁵³ the latter being significantly more costly.

This scenario raises a host of difficult questions for the State Board of Education: Will large numbers of schools categorized as “needing improvement” under NCLB spread resources so thinly that the additional funding for a truly needy school is inadequate? Will the financial burdens of implementing the many other federal requirements of NCLB further reduce the resources currently available to needy schools and students? Will the state’s focus on and funding for the HSA program be eclipsed by concerns about ensuring success on the NCLB 10th grade reading and math tests?

⁵³ Title I: Characteristics of Tests Will Influence Expenses; Information Sharing May Help States Realize Efficiencies GAO-03-389, May 8, 2003.

VI. What We Know from Experience with the High School Assessment Program



Achievement Gaps Are Already Evident

The most recent HSA percentile rank scores from 2002 indicate performance gap patterns among students, schools and districts based on race, family income and special education status. Given the link between high-stakes tests and dropout rates, particularly among at-risk students, this is a critical gap for Maryland education officials to confront in order to improve high school learning. These gaps are important because they demonstrate precisely the inequities that standards and accountability measures allegedly seek to alleviate. At the same time, they highlight the difficulty of setting performance standards that hold all students accountable without penalizing those students the program purports to help.

Median percentile scores were reported for all Maryland schools and districts, and broken down according to racial and ethnic groups, students with disabilities, English language learners, and low-income students who qualify for free and reduced lunch and Title I funds. The median is the middle score in a set of ranked scores. A score of 50 is an average score; if a group has a median percentile score above 50, that group is above average compared to all other Maryland students.

“Regular” education students outperformed recipients of special education services by an overall median percentile of 53 versus 19.

2002 HSA Performance – Student Receiving special education services

HSA score by median percentile	Algebra	English	Biology	Government	Geometry
Regular education students	52	54	53	53	52
Students receiving special ed services	20	16	19	18	22

Low-income students who qualified for free and reduced-priced meals performed on average at the 31st percentile, while more affluent students who did not qualify for assistance performed at the 54th percentile.

2002 HSA Performance by free/reduced lunch status

HSA score by median percentile	Algebra	English	Biology	Government	Geometry
Non free- and reduced-meal students	54	55	55	54	54
Free- and reduced-meal students	32	32	31	32	30

Asian-American and white students noticeably outperformed Hispanic-, African-, and Native- American students, with the greatest disparity between Asian-American students who performed at the 73rd percentile and African American students at the 32nd percentile.

2002 HSA Performance by ethnicity

HSA score by median percentile	Algebra	English	Biology	Government	Geometry	Overall Mean%
Asian	75	71	74	68	76	73
White	63	61	63	61	61	61
Hispanic	39	41	41	41	40	40
American Indian	35	36	37	36	45	38
African American	28	34	30	34	33	32

Baltimore City and Prince George’s County populations are predominantly African-American and, in the case of Baltimore City, both poor and urban. Students in these jurisdictions performed below students in suburban and rural counties, regardless of the poverty level.

HSA Performance by County with Free/Reduced Lunch Rate

HSA score by median percentile	Free/reduced lunch rate high school	Algebra	English	Biology	Government	Geometry	Overall percentile
Baltimore City	45.8%	23	34	24	37	20	28
Prince George’s County	26.8%	28	34	31	32	25	30
Somerset County	45.8%	43	46	40	45	31	41
Montgomery County	11.1%	72	67	70	70	70	70

Even in higher-performing schools and districts such as those in Montgomery County, there are significant achievement gaps among races when controlling for income by deleting the free- and reduced-lunch population.

These scores for middle and upper income students suggest that the achievement gap may not be determined by income alone.

2002 HSA Performance – Montgomery County Non-Free/Reduced Lunch Population

HSA score by median percentile	Algebra	English	Biology	Government	Geometry	Overall Percentile
Asian	85	79	83	78	83	82
White	82	79	80	81	77	80
African-American	48	48	49	52	47	49
Hispanic	46	46	48	44	47	46

Females generally perform slightly better than males on HSA tests, most notably in English where they scored at the mean 58th percentile compared to the 41st percentile for males. The gender gaps become more accentuated when disaggregated by race and income, raising red flags about the overall performance of African-American males:

2002 HSA Performance by Gender

HSA score by median percentile	Algebra	English	Biology	Government	Geometry	Overall Percentile
Female	51	58	52	53	50	53
Male	48	41	48	47	50	47
African American males free/reduced lunch eligible	22	22	23	25	21	23

Two of these performance gaps demand the most critical attention.

First, testing accommodations on the HSA already exist for students with learning disabilities,⁵⁴ yet their relative performance is extremely low. The state must set performance standards that treat these students fairly. In Virginia, for example, special education students are allowed to earn a “modified standard diploma” that does not require them to pass its exit exams. While this idea has been floated in Maryland as well, many parents and advocates are concerned that a different diploma would consign special education students to a second-class education and weaken efforts to ensure that students with disabilities receive an equal education. Whether this differentiated diploma would affect a student’s employment prospects is unclear.

Second, the achievement gaps among races are stark. Poor performance trends among minority and low-income populations within districts and schools, and for the majority-minority districts of Prince George’s County and Baltimore City, are longstanding and appear to continue unabated under the new HSA system.

Still in the Dark: How Did Maryland Students Perform vs. Standards?

The limited reporting of HSA scores to date confirms predictable achievement gaps in performance, but offers no insight into whether students are meeting state expectations in terms of academic achievement. While such information is slated to be available later this year, its absence now severely limits the opportunity to learn from the HSA implementation thus far in a way that impacts classroom instruction.

⁵⁴ “Accommodations will be consistent with those provided during classroom instruction and testing and designated in the IEP for a student with a disability, or in the plans for eligible 504 or limited English proficient students.” Message to Maryland Parents from Dr. Nancy Grasmick, December 2002. Retrieved from <http://www.msde.state.md.us/testing/Dec2002HSAletter.pdf>.

During the last five years, between 49% and 56% of the Maryland students who attended a Maryland community college needed math remediation.⁵⁵ While the University of Maryland System, MSDE and the Maryland Higher Education Commission have established the structure for higher education institutions' participation in devising the HSA program through the K-16 Partnership for Teaching and Learning and have participated in content standard reviews, there is little evidence that the state's high school graduation standards have been aligned with college admissions requirements. For example, most Maryland colleges require Algebra II and some Trigonometry as a pre-requisite for college math, but Maryland's K-12 system requires only Algebra I and Geometry for high school graduation. Nor have Maryland's post-secondary institutions embraced the HSA tests thus far as an indicator of content mastery, although the University of Maryland system allows, but does not require, its institutions to use HSA scores as an admission/scholarship criterion.

Researchers have advocated using HSA scores instead of college placement tests for post-secondary entry arguing that it would reduce the testing burden and motivate students to meet state standards. As an example, post-secondary institutions could consider HSA performance as a criterion for admissions, financial assistance and/or course placement. A study of secondary and post-secondary alignment in Maryland recommends that the Algebra and Geometry HSAs be used to exempt students from remedial college placement tests, while the English HSA might also be a viable alternative to college admissions (Le, 2002). In New York, officials at the City University of New York (CUNY) have replaced the college's own English placement exams with the English Regents exam, making the connection between high school and college performance real for students. Any such coordination would require planning and coordination that has yet to occur between Maryland's K-12 and post-secondary institutions.

A Critical Lesson in Public Relations

Data measuring student performance against state standards are slated for release in the next few months, but a pattern has already emerged suggesting a reluctance on the State's part to disclose student performance fully. Much of the MSPAP testing process was clothed in secrecy even as scores stopped increasing, and there are discrepancies in Maryland's reporting of its high school graduation rate versus national statistics. Even today, the State is unable to pinpoint the number of students barred from graduation by the Functional Tests.

Laudably, the HSA process began with over 200 public sessions, but this has slowly devolved into numerous decisions made behind closed doors without adequate public dialogue and seemingly without adequate information. The State intentionally planned to release comparative percentile data only in the first year so that educators'

⁵⁵ Student Outcome and Achievement Report(SOAR). Maryland Higher Education Commission, November 2002.

demand for results could be assuaged without rushing the Board in its standards-setting process. Matthew Gandal of Achieve points out the paradox of using normative score reporting for criterion-referenced tests (comparing student against student rather than student against standard), and notes that it serves a purpose by buying the State the additional time it needed to set standards.⁵⁶

The normative HSA data released to date, however, raises questions: the 2002 HSA percentile and scale score data available on the MSDE website (www.msde.md.us) demands further explanation. While psychometricians and statisticians understand the science of adapting raw scores to scaled scores, and the tenuous quality of first-year scores, how should the general public interpret the fact that the 99th percentile on the English I assessment equates to a scale score of 473 on a scale of 0 - 800? Does that relatively low score mean that, generally, Maryland students did not fare well as measured by the standards? As it stands, the public has received insufficient information regarding decisions that will profoundly impact its children.

⁵⁶ Telephone interview with Matthew Gandal, Executive Director, Achieve, Inc., December 18, 2002.

VII. Key Issues the Board Must Resolve



The major decisions the State Board of Education is charged with making are as follows:

- What will the proficiency levels and passing scores be for the HSAs, i.e., what will constitute passing on each HSA?
- How can the HSA program hold students accountable for achievement without making failure so onerous that it over-penalizes disadvantaged students?
- Should it proceed with plans to expand the program from four (or five) to 12 end-of-course exams?
- How will it build school, district, and state accountability into the HSA program and comply with NCLB?
- Should passing the HSAs be made a graduation requirement, and if so, when? Students entering high school in Fall 2004 are currently targeted.

To ensure that these decisions lead to an effective HSA program, the Board must also address a series of underlying issues that are critical to ensuring academic achievement and accountability at the high school level. These range from logistical issues (testing process and timing) to questions about the feasibility of the HSA program as planned, its inherent equity for all participants, and Maryland's ability to maximize the opportunities the HSA program presents to improve public education for all its students.

Funding Limitations—Past, Current and Future

The HSA program and No Child Left Behind Act of 2001 are intended to provide the education community with data and legal impetus to help reduce educational inequities among students. Yet, even before the NCLB accountability requirements were enacted and the state budget crisis arose, inadequate state funding threatened Maryland's ability to both raise standards and reduce the achievement gap, prompting the Board to delay the HSA graduation requirement on three occasions.

In addition to the costs of HSA development and administration (including the additional expense of the new 10th grade reading test), there are also issues of what academic interventions become necessary when standards are raised. Questions about the fiscal feasibility of the HSA program date to 1997 when the Board first proposed use of HSA tests as a graduation requirement. The Board's then-vice president, Dr. Edward Andrews, called for making the HSA a graduation requirement contingent on two conditions: remediation for failing students before high school, and teacher training. MSDE, which is to "provide models of curriculum, program alignment, and policies, and assure adequate funding to support local school systems' comprehensive

intervention plans,”⁵⁷ responded with an academic intervention plan, *Every Child Achieving: A Plan for Meeting the Needs of Individual Learners* (1999).

The plan noted that the approval of exit exams “raised serious concerns regarding responsibility, equity and accountability, and thus created a new imperative for monitoring individual student progress and for instituting a comprehensive program of prevention and intervention...early in students’ academic careers.”⁵⁸ It outlined a comprehensive approach to improving teacher capacity, to preparing students for meeting standards, and to intervening early where students are not meeting standards. Among the interventions recommended were after-school tutors, whole-school reform models, and class size reductions to 15 students in the 1st and 2nd grades; this item alone is projected to cost the State \$17,495,000 a year.

When former Gov. Parris Glendening provided only \$11.5 million of the \$49 million requested for FY ’01 in the *Every Child Achieving* plan, the Board again voted to delay the graduation requirement until Fall 2003, making the Class of 2007 the first to be required to pass the tests in order to graduate. As a result, students entering 9th grade in Fall 2001 were required to take all 2002 and 2003 HSAs, although they were not required to pass them. Individual scores were to be reported, however, as percentile rankings on their transcripts.

Looking ahead, state funding commitments remain unclear. For FY’03, \$19.1 million was allocated for the Every Child Achieving plan and included in the FY ’04 budget as part of the Bridge to Excellence master plan. Additional FY’04 funding amounting to \$242.1 million was recommended by the Commission on Education, Finance, Equity, and Excellence (Thornton Commission) and approved by the General Assembly in 2002. Aimed at reducing inequities among the state’s school systems, this funding formula targets greater resources over the long term toward the state’s poorest areas so that all students can better meet standards. The majority of funding for the \$2.4 billion plan occurs in the later years (2006-2008) and is dependent on re-approval of joint resolutions by the House and Senate in 2004. If this does not happen, the funding will instead amount to an increase of 5% per year, threatening many of the State’s educational programs.

The Board has pledged to review academic support programs this summer prior to considering the graduation requirement. One strategy MSDE has identified for providing intervention is the development of a voluntary state curriculum at every grade level in every content area. This curriculum would be designed to help teachers more efficiently and effectively translate state content standards into classroom practice, incorporating both enrichment and remediation as necessary. As the Board conducts its review of

⁵⁷ “*Every Child Achieving: A Plan for Meeting the Needs of the Individual Learner.*” A Report to the Maryland State Board of Education, October 1999. p 5.

⁵⁸ Executive Summary, *Every Child Achieving* (MSDE, 10/27/99).

such options, an analysis of the State's ability to provide the necessary funding and support to help districts support under-prepared high school students will be critical.

Past and Present Delays May Derail the Original HSA Program

State education officials concede that they underestimated how long it would take to construct, validate, and field-test four content-focused tests, yet just this spring the Board decided for the third time to delay the graduation requirement by a year, such that it will now apply to students entering 9th grade in Fall 2004 and graduating in 2008.

High School Assessment end-of course exams	Original 1997 Timetable⁵⁹	Actual timetable
Field tests	1999-2000 field testing	2000 field testing - selected schools 2001 field testing - all schools
First phase testing (English I, Algebra, Biology Government)	2001 – 9th grade students Required for graduation	2002 – 9th grade students Not required for graduation; scores reported on transcripts only
Graduation requirement applied	Class of 2004	Class of 2008
Future Phases HSAs required for graduation 7 HSAs 10 of 12 HSAs	Class of 2005 Class of 2006	Undecided Undecided

Among other substantive issues that factor into the graduation requirement decision, such as providing students with the necessary supports to succeed, the Board must consider the impact of another delay on the momentum of the HSAs. Given the controversy MSPAP engendered and the public outcry in other states over implementation of high school graduation requirements, there may be advantages to MSDE continuing its cautious approach. These delays can help ensure that schools are able to respond to higher standards (informed, ideally, by student performance against those standards), and that students and parents are made fully aware of an impending graduation requirement.

On the other hand, delays allow time for high-stakes testing opponents to organize, opening the door to potential backlash. As mentioned above, some activists in Maryland have already taken advantage of the extended timeline: Marylanders Against

⁵⁹ Mary Maushard, "Education Officials Seek Input on Tests," *Baltimore Sun* (October 15, 1996) and "State OKs New Tests for Graduation," *Baltimore Sun* (December 11, 1997).

High-Stakes Testing supported proposed House Bill HB1166 which would have prohibited the Board from making passage of a single assessment or series of assessments a requirement for graduation from public high school. The three delays have potentially lowered the commitment of educators and students to the HSA program as well – how serious can one be about tests that, at best, won't impact graduation until 2008?

Finally, delays to date raise questions about the advisability of expanding the HSA program as originally envisioned: Can the State implement the full range of HSA tests in a timely manner, and does it still make sense to do so?

Data Is The Key to Winning Points With the Public

Two issues related to HSA data should factor into the Board's deliberations: Maryland's disinclination to create a test and/or scoring process that would allow quick release of data to schools and the public, and its ongoing reluctance to release data that could spark public controversy. These are important considerations not only for practical purposes, but also because more forthright information would help build statewide goodwill with an unengaged and skeptical public.

The utility of the HSA program in improving instruction would be greatly enhanced if scores were released in less than three or four months. Because Maryland high schools cannot use HSA scores for promotion or final course grading, many school districts have had to create or continue to use additional end-of-course final exams that can be administered and scored within the school—an additional and undue burden on students and teachers alike that is not necessarily in concert with the goals of the HSA program.

Meanwhile, some states using end-of-course exams as high-stakes assessments are accessing test results and putting them to use quickly. New York State's Regents exams are typically administered the third week of June, and scores are reported the following week thanks to a system of on-site grading. Tests and essays are graded by teams of teachers at each school during "Rating Day"—a staff day scheduled immediately following Regents tests administration. To compare school raters to state raters, 10% of schools are randomly selected each year for external evaluation, and results are returned to the New York State Education Department by September. Surely Maryland could establish a similarly efficient process.

The limited HSA data released for 2002 have been of little use to either the public or educators highlighting another issue the Board must address: its relationship with the Maryland public. While it is true that detailed performance data and public response to it may have unpleasant consequences, such openness could also inform the Board's ongoing debate about inequities that have contributed to the delays in the graduation requirement thus far. Student data, measured against the specific learning standards the HSA tests were designed to assess, could provide schools,

teachers, parents, and students with useful information on how to better prepare students to meet standards; how to address barriers to success; and whether those barriers are found in the tests themselves, the scoring process, or a lack of resources at the school, school system, or state level.

In Virginia, early release of negative data sparked the predicted public outcry, but it also created an opportunity to improve the state's standards and assessment system.

The Virginia Standards of Learning Story A Study in Public Engagement

Virginia's state standards were adopted in 1995; the "Standards of Learning" (SOL) tests were piloted in 1997, with the first full administration in 1998. The Class of 2004 will be the first required to pass six SOL tests for graduation.

Unlike Maryland, Virginia adopted cut scores in Fall 1998 after the first administration of the test; scores released in January 1999 showed that only 2% of schools met accreditation requirements. From one perspective, release of this kind of damning information was a public relations disaster, resulting in media coverage and alarmist reactions among the public. From another perspective, Virginia's open approach has enabled the public to judge student performance against standards, as well as the value of the testing program. Moreover, public confrontation of disappointing results has led to immediate and long-range public policy changes including:

- the Governor's immediate request not to include SOL results on student transcripts;
- establishment of an advisory council and expert panel to provide continuous review of both standards and testing;
- changes to the program's world history standards, exam and cut score; and
- accommodations for learning disabled students and recent immigrants; for students in advanced study programs like AP or International Baccalaureate; and for the first three classes faced with the graduation requirement, whose students had received much of their schooling before the SOL tests were in place.

Opposition to the SOL tests in Virginia has not completely quieted as a result of making the assessment process public. Recently, in fact, several school districts complained about the state education board's rejection of a proposal to allow local school boards to award credits to students with high grades but failing SOL scores. Overall, however, openness in Virginia appears to have contributed to the impetus for continuing discussion and program improvements, while bolstering public support.

This summer, Maryland has a similar opportunity to engage the public in its standards-setting process. Because the Board's decisions in setting proficiency levels involve cultural and political judgments about what performance or score constitutes "passing," it is critical that the State include participants who represent the diversity of Maryland's students in the process—and openly communicate the outcomes of that process. Not only could the HSA program likely benefit from public participation and scrutiny, but such debate could engage the public in a way that makes the largest stakeholders in the HSA program—students and their families—feel counted.

The Need for Alternatives to the HSA Tests

Based on the recent experience of other states introducing high stakes testing, the Board, as part of its decision on whether to implement an HSA graduation requirement, would be wise to consider proactively alternative routes to a high school diploma rather than responding hastily to unanticipated student failure and public outcry.

With plans to expand to 12 HSAs, Maryland is positioning itself among those states with the most onerous graduation requirements in the nation. This proposal requires Maryland high school students to pass 10 of the 12 tests in order to receive a diploma. Under this “conjunctive” plan, a student who fails just one of the HSAs would be barred from graduating.

Retesting is not simply a matter of retaking the test. More importantly, how will students with multiple HSA failures overcome deficiencies and advance to new, more challenging coursework? Maryland’s HSA program stipulates that these students must participate in some form of “remediation” before retaking the exam, though what this entails is unclear. This remediation requirement could potentially throw another cumbersome obstacle in the path of graduation. It is critical that MSDE and the districts collaborate to insure that remediation is easily accessible and responsive to student need and that it works. Furthermore, the Board may need to administer the HSA more than twice a year so that both enrolled students and out-of-school youth have multiple opportunities to retest.

Because concerns about students who fail multiple HSA tests will no doubt arise and intensify if and when Maryland phases in all of its projected exit exams, the Board could establish guidelines for remediation and other pathways toward a Maryland high school diploma upfront.

Where’s the Accountability?

Maryland’s implementation of its HSA program thus far lacks accountability on several fronts. Missing are both measurable outcomes and accountability to an independent evaluation of the impact of this initiative, particularly its unintended consequences. The Board has not yet established the framework to answer the single most important question to its purpose: Is the HSA program succeeding in its effort to improve high school instruction (and hence post-secondary opportunities) for students throughout Maryland?

Maryland’s original goals for the HSA program are not clearly and specifically articulated and, for the most part, are not attached to established, measurable outcomes. Similarly, there is no evaluation system in place to measure the success of the program against important outcomes on a continuing basis. Though the state uses language like “improving high school instruction” and “raising standards,” there is no indication

of how progress toward these critical goals will be measured. How will the State ensure that increased learning, not increased test scores alone, are driving the reform?

Incorporation of a well-conceived independent evaluation would address this concern by holding MSDE accountable not only for standards and test scores, but also for actual classroom instruction and learning in Maryland's high schools. It would also inform the State and public on a continuing basis where Maryland students stand, identify unanticipated consequences, and allow for mid-course corrections.

California, for example, commissioned an outside evaluation before its final decision on implementing the graduation requirement, providing a basis for informed policy-making.

Staying on Track: California's Feedback and Accountability Plan

California contracted with the Human Resources Research Organization to provide a longitudinal evaluation with a series of annual reports before the state requires passing of its high school exit exam (CAHSEE) for the graduating class of 2004. California students started taking the math and English tests in 2001, with accountability looming for the class entering 9th grade that year. Each year, the evaluation has analyzed trends in achievement, retention, graduation, dropout rates, and college attendance rates, as well as qualitative information from teachers and students. The annual reports have provided both information and recommendations to state educators and the public about the relationships between performance, outcomes, and implementation within four months of CAHSEE administration. The May 2003 report updates passing rates for the class of 2004 and provides an impact statement for instituting a compensatory approach that would allow students with achievement above the minimum in one subject to make up for deficient achievement on another with a combined total score.⁶⁰

Percent of Students Passing the CAHSEE in March 2003

Student groups	English % Passing	Math % Passing	Both % Passing % with total score >700	Compensatory approach:
All students	66	46	43	56
African American	52	25	23	37
Economically-Disadvantaged	48	27	23	35
Special Education	42	14	10	16

This evaluation has also been critical in offering evidence of "dramatically increased" coverage of California's academic standards in classrooms, and predicting that 20% of the first officially accountable graduating class under the system will be denied a diploma.⁶¹

⁶⁰ *Independent Evaluation of the California High School Exit Exam: AB 1609 Study Report.* Human Resources Research Organization, May 1, 2003, p.96. Retrieved from <http://www.cde.ca.gov/statetests/cahsee/eval/AB1609/chapter6.pdf>

⁶¹ Olson, L. States Debate Exam Policies for Diplomas. *Education Week*: 14:36, May 14, 2003, p.14.

VIII. Next Steps



The goal of this report is to engage policymakers in, and focus public attention on, a critical evaluation of the state’s proposed HSA program, to ensure that further investments bring concrete improvements for student learning. The State Board of Education is expected to make critical decisions in the upcoming months about the future of high-stakes testing in Maryland. Because there is much to be learned from other states, as well as Maryland’s own experience in high-stakes testing to date, it is important that the Board draw on the many resources currently at its disposal to ensure that its decisions are thoughtful and well-informed.

Key questions

Following are questions that frame the ensuing recommendations that flow from collected information on high-stakes testing in Maryland and elsewhere. Together, it is hoped these will focus the Board and public on how best to leverage Maryland’s HSAs to increase student achievement:

- How can Maryland set pass scores and proficiency standards on the HSA that promote academic rigor but do not result in unacceptable numbers of failures that reduce the already limited opportunities available to poor, disproportionately African-American and Hispanic, and special education students?
- What should the stakes for students be? Is it time for a graduation requirement or will this doom many low-income minority students to failure? Conversely, can demanding tests without any high stakes accountability for students achieve the stated purpose of improving instruction and learning for all students?
- What testing format (multiple choice objective vs. open-ended constructive items) will yield valid and reliable outcomes while allowing for cost-effective and expedient scoring?
- How should the State implement the next phase of the HSA testing program? Given the resources needed and requirements of NCLB, does it make sense to expand the HSA as originally planned? Alternatively, do the existing five HSAs (including Geometry) make sense without the full-scale implementation of 12 end-of-course exams?
- How will we know if the HSA is improving high school instruction and rigor as opposed to simply increasing test scores – a discernible pattern with any new test? What is the State trying to achieve, and how will the State Board know it has succeeded?
- How can the State achieve the above while re-engaging educators and the public and working constructively with stakeholders so that criticisms and concerns lead not to public alienation and political backlash, but the best possible solutions for Maryland’s students?

Recommendations

The Sondheim Report supported the concept of high-stakes high school testing in Maryland as far back as 1989, but recommended a thorough investigation of the following issues prior to implementation that are still valid today:

-
1. adequate prior notice to students;
 2. complexity of retesting; and
 3. additional legal constraints that apply to a test that can grant or deny a high school diploma.⁶²

More recently, the Center on Education Policy made some general recommendations about instituting high-stakes testing that are useful in framing recommendations for Maryland in its deliberations this summer.

The Center on Education Policy recommends:

1. Use exit exams in conjunction with other measures of student performance
2. Ensure that students have an adequate opportunity to learn the material being tested
3. Provide sufficient opportunities for remediation
4. Provide adequate funding
5. Conduct outreach campaigns to build public understanding and support
6. Monitor the consequences
7. Phase in new exit exams over several years

— *State High School Exit Exams: A Baseline Report, August 2002*

In light of the above principles, this report offers the following recommendations for consideration by the State Board:

1) Set high standards immediately, but raise the stakes for student performance on the High School Assessments gradually over time.

There is evidence that high stakes testing may have a negative effect on graduation rates for the very students it is designed to help. Based on what already appear to be significant achievement gaps in HSA scores, imposition of a graduation requirement (accompanied by challenging performance standards) would likely have a disproportionately negative impact on precisely those at-risk students who need the most help. If the Board were to institute the graduation requirement this year, it might be tempted to set lower performance standards in an effort to reduce the negative impact upon graduation rates. Such an action would undermine the program's mission to enhance academic rigor and promote improved outcomes for those students least likely to graduate.

There is a strong likelihood that State funding for academic intervention for failing students (not to mention general support for curriculum and instruction in HSA courses) will not be available for some time. Moreover, the idea that even a fully-funded targeted intervention plan will magically provide equal opportunity for all students to learn is flawed, given the long-standing deficits in many students' school and

⁶² The Report of the Governor's Commission on School Performance (The Sondheim Report). (1989). p.17.

home environments. The HSA program, if instituted today in its entirety with a graduation requirement, is likely to create insurmountable burdens on schools and students to complete remediation and retesting in unrealistically short time frames, especially for students who fail multiple tests.

To address and minimize these outcomes, the Board could do one of the following:

As an intermediate step in the HSA program evolution, The Board might adopt a medium stakes alternative to the graduation requirement that would require all school jurisdictions to incorporate HSA scores as final exams that comprise a significant (at least 30%) proportion of a student’s grade in that course. This option would require a significant change in scoring procedures to facilitate faster reporting.

Or, alternatively,

The State Board might set a composite score for performance on all the High School Assessments, creating a compensatory model whereby a student’s good performance on one HSA would offset a lower performance on another.

Both options would enable the Board to set performance standards that bolster the HSA Program’s mission of academic rigor while recognizing that, in the short term, many districts and schools will be unable to offset long-term students deficits in both home and school environments so that all students have an opportunity to master high level content.

Either option would give the State and its educators the ability to focus more energy on the true drivers of high school improvement – the development of a strong program of curriculum and instruction, intervention and remediation, and professional development and classroom observation. Finally, under either option, the State could institute the remainder of the HSA program, if it so chooses, without imposing complex and perhaps insurmountable retention, remediation and retesting burdens upon students, teachers, and schools.

2) Put adequate and feasible remediation and alternative paths to a high school diploma in place before HSA-linked graduation requirements go into effect.

As evidenced by its initial delay of the graduation requirement, the Board is concerned about imposing such high stakes on students without providing adequate support for all students to prepare, remediate, and retest. At the same time, the State has yet to provide individualized performance data against standards, nor have schools developed and begun to institute plans for remediation. The Board would be wise to ensure that multiple pathways to attain a Maryland high school diploma exist.

To retake an HSA, Maryland stipulates that students are required to participate in “remediation.” Because the graduation requirement is not in place and such remediation is not now required, it is not clear what the remediation entails or whether it will be effective. In fact, such a stipulation might present yet another obstacle to success. To this end, it is critical that the Board follow through on its commitment to review remedial programs that have been effective in Maryland high schools prior to making final decisions on high stakes testing. It is also recommended that the HSAs be administered up to four times each year and offered to students who have left high school.

Like other states embarking on high-stakes testing, Maryland may find it necessary to provide alternative pathways to a diploma for students who have failed the HSA tests multiple times. The Board should therefore consider offering more testing opportunities, alternative assessments, a waiver process, alternative curricula, and differentiated diplomas, among other options. Along these lines, it should also consider whether the current Independence Mastery Assessment Program diploma will still be awarded to some students with disabilities.

On the other end of the spectrum, Maryland might combat concerns about a narrowed curriculum by accepting passing scores from nationally recognized tests typically taken by higher-achieving students as a substitute for the HSAs. These include appropriate SAT II, Advanced Placement or International Baccalaureate tests.

3) Ensure that the HSA test design and the scoring process will be useful in improving high school instruction. At the very least, demand a faster turnaround for test score reporting.

The HSAs should be an integral component of promotion and other instructional decisions as well as function as a true cumulative end-of-course test (rather than one administered five weeks before the end of school with scores released over four months later). This will require a revised testing format and/or scoring process that will allow the test to be taken at the true conclusion of the course and make test scores available within 10 days. The New York Regents provides such a model without arguably compromising the integrity of the assessment. In order to achieve this tight time frame, the State might also reconsider simplifying the test format to include only multiple-choice items or revisit the decision to have the tests externally-scored. Either of these choices would also be less costly. At a minimum, the Board should clearly communicate to the public the higher costs and extended scoring time that result from its current choice of HSA format versus a purely objective test.

4) Assign similar student accountability to the two NCLB tests given in 10th grade.

While student accountability is not a component of NCLB requirements, research suggests that attaching stakes to tests used for school and district accountability will

increase the likelihood of improved school performance measures, particularly among disadvantaged youth. Maryland should consider a student accountability measure in all NCLB high school tests as well as end-of-course exams. The medium stakes (or composite score) alternative proposed above stands a good chance of spurring student effort without decreasing the graduation rate. In addition, the State might integrate the 10th grade reading test for NCLB with the proposed 10th grade English II HSA requirement, and do it sooner rather than later.

5) Renew the level of public information and awareness that marked the campaign to institute the HSAs with a more transparent and participatory process for setting goals, establishing cut scores, measuring progress, and evaluating outcomes. Begin with releasing data on how Maryland students actually performed against standards.

Maryland has not provided adequate and useful information about student performance on the HSAs to administrators, teachers, students, and parents. Field tests for the HSAs were used to improve the tests but not used to improve public understanding of student performance on the tests. Lack of performance standards, of information about how students performed against specific learning standards that the tests are designed to measure, and of specific remediation plans at the school level mean that teachers, students, and parents are not prepared to hear potentially bad news about failure to meet standards.

The Board's examination of the HSA program this summer presents a timely opportunity to re-engage both public and educators. The process of scoring tests, scaling scores, and setting performance standards should be made more transparent to the educators, parents, and students whom it affects the most. Drawing upon Virginia's experience, it would be advantageous for the public and decision makers to understand how Maryland students actually performed against standards rather than against each other. We recommend a public release of students' criterion-referenced scores on the 2002 (and 2003 if possible) HSAs as a critical component of establishing cut scores. We also urge the State to use community forums again to gather feedback and provide information about the graduation requirement. Finally, members of panels formed for establishing cut scores should represent the wide diversity of Maryland's stakeholders.

6) Institute a longitudinal study by an independent evaluator to set, measure and continuously evaluate progress toward clearly articulated educational outcomes, as well as measure unintended consequences. This evaluation should be disseminated to educators and the public at regular intervals and should begin with publishing criterion-referenced scores on the 2002 and 2003 HSAs.

Maryland should institute a process whereby it sets quantifiable objectives, measures desired student outcomes, and evaluates the HSA program in light of student, school,

and district progress toward these outcomes. It is critical that the study identify unintended consequences of Maryland's high stakes testing (i.e., dropout rates) and analyze the impact of the HSAs on the scope of the high school curriculum. An obvious way to accomplish this is to contract for a longitudinal study with an independent evaluator. While HSA scores may increase over time, the more important questions are whether all students have received a better education, whether their post-secondary prospects are improved, and whether they are better prepared to take advantage of those opportunities. The state must articulate these goals and institute a comprehensive system that establishes and evaluates progress toward measurable standards.

Equally important is the opportunity afforded by such an evaluation for the State to make mid-course corrections in its assessment plan and to improve its supports for schools and students. Such a study would evaluate the impact on the classroom as well as the test scores – on determining how the program has affected the quality of high school instruction and student learning and, in turn, informing the instructional techniques of teachers. It would also serve to measure other important outcomes for student achievement and future opportunity, such as graduation rate, rigor of high school curriculum taken, SAT scores, and preparation for post-secondary opportunities including, but not limited to, college matriculation and achievement.

7) Understand the impact of the first phase of the HSA program and the new NCLB tests before introducing additional Phase II tests.

Maryland, unlike some other states using end-of-course exams, debuted four high stakes exams in one phase. This major project was further complicated by the mandatory implementation of reading and math assessments for all 10th grade students required in 2003 by NCLB. The Board may also consider rendering no decision on pursuing Phase II of the HSA program until findings from an initial external evaluation of Phase I are communicated.

8) Work with Maryland's higher education community to articulate performance on the HSA with college admission and placement.

Higher education institutions should ensure that HSA course and test content is well articulated with the requirements of the state's postsecondary institutions. A stronger and better coordinated program, as outlined in these recommendations, may inspire willingness among state higher education institutions to reward strong HSA performance with preferences in admissions and financial assistance, which would provide another incentive to students to perform well, and remove additional barriers of placement testing, particularly at Maryland's community colleges.

IX. Conclusion



Maryland is at a critical junction as policymakers decide how high to set the bar for graduating from high school: whether to make passing scores on the HSAs a graduation requirement; what exactly constitutes passing the tests (i.e., what proficiency standards students must meet and how those translate into passing scores) and whether to proceed with plans to introduce up to seven more tests.

While the Board faces many difficult decisions over the next few months, the decision it must make first—the one that will drive the rest of the decision-making process—poses two possible outcomes. The Board can decide to move quickly to raise the standards and stakes for a diploma, which will likely result in thousands of 12th grade students being unable to graduate in May 2008, particularly those who are minorities, impoverished, disabled, or otherwise at risk or disadvantaged. Or it can choose a more judicious path of relying on high standards with gradually-increasing student consequences, more efficient scoring, an external evaluation with annual reporting, mid-course corrections, and increased public and higher education engagement to improve the quality of high school learning across the board.

Will the HSAs as currently envisioned achieve the State’s goal of producing high school graduates with the skills and content knowledge to greatly enhance the value of Maryland’s diploma? It is up to the Board to answer this question, choosing what is best for Maryland’s students, not necessarily what makes the State and its schools look best. It is high time to put high school high-stakes assessment at the center of our collective attention.

References

- Amrein, A., and Berliner, D. (2002). High-Stakes Testing, Uncertainty, and Student Learning. *Education Policy Analysis Archives*. 10(18). Retrieved from <http://epaa.asu.edu/epaa/v10n18/>.
- Amrein, A., and Berliner, D. (2002). *An Analysis of Some Unintended and Negative Consequences of High-Stakes Testing*. Tempe: Arizona State University.
- Bishop, J. (1998). The effect of curriculum-based external exit systems on student achievement. *Journal of Economic Education*, 29(2), 171-82.
- Carnoy, M., and Loeb, S. (2002). Does External Accountability Affect Student Outcomes? A Cross-State Analysis. *Educational Evaluation and Policy Analysis*, 24(4), 305-321.
- Chadowsky, N., Kober, N., Gayler, K., and Hamilton, M. (2002). *State High School Exit Exams: A Baseline Report*. Washington, D.C.: Center on Education Policy. Retrieved from <http://www.ctredpol.org/pubs/statehighschoolsexitexams.html>
- Clarke, M., Haney, W., Madaus, G. (2000). High Stakes Testing and High School Completion. NBETPP Statements, 1(3), Chestnut Hill: Boston College, Center for the Study of Testing.
- Greene, J. (2001; April 2002 revision). *High School Graduation Rates in the United States*. New York: The Manhattan Institute for Policy Studies.
- Hamilton, L., Stecher, B., and Klein, Stephen P (eds.). (2002). *Making Sense of Test-Based Accountability in Education*. Santa Monica: Rand Publications.
- Haney, W. (2000). The Myth of the Texas Miracle. *Educational Policy Analysis Archives*, 8(41). Retrieved from <http://epaa.asu.edu/epaa/v8n41/>.
- Horn, C. (2003). High Stakes Testing and Students: Stopping or Perpetuating a Cycle of Failure? *Theory into Practice*, 42(1), 30-41.
- Jacob, B.A. (2001). Getting Tough? The Impact of High School Graduation Exams. *Educational Evaluation and Policy Analysis*, 23 (2), 99-121.
- Kellaghan, T., Madaus, G., and Raczek, A. (1996). The Use of External Examinations to Improve Motivation. Washington, DC: American Educational Research Assoc.
- Le, V. (2002). *Alignment Among Secondary and Post-Secondary Assessment in Five Case Study States*. Santa Monica: Rand Publications. Retrieved from: <http://www.rand.org/publications/MR/MR1530.0/>

Linn, R.L. (2000). Assessments and Accountability. *Educational Researcher*, 29(2), 4-16.

Maryland State Department of Education. (2002). *Aiming Higher: The Next Decade of Education Reform in Maryland*. Washington, D.C.: Achieve Inc. Retrieved from http://www.msde.state.md.us/pdf_files/aiming.pdf

McDermott, K. (2001). *State Testing and Graduation Requirements: A Framework for Analysis*. Amherst: Center for Education Policy, University of Massachusetts.

Raymond, M., and Hanushek, E. (2003). High Stakes Research. *Education Next*, (3) 48-54.

U.S. Office of Technology Assessment (1992). *Testing in American Schools: Asking the Right Questions*. Washington D.C.: U.S. Government Printing Office.