Briefing Paper on the 2008 Maryland School Assessment Results

August 26, 2008

Contents

Executive Summary

The Maryland School Assessment (MSA) program has been in place for five years and was first administered in 2003 as the State's accountability tool for elementary and middle schools under the federal No Child Left Behind Act. In order to assure that Maryland could produce viable test results in the first year of the new tests, the State used an "Augmented Norm-Referenced Test" design, which included a small number of test items from a national test that were augmented by a much larger number of State-developed test items. The assessments, peer-reviewed and approved by the United States Department of Education, have served the State well as valid and reliable tests that are annually studied and approved by a panel of national experts. This panel has assured that the tests have held to the same level of accuracy and difficulty over the six administrations of the test.

The 2008 administration of the assessments differed slightly from prior administrations of the test in that the State no longer required students to take Norm Referenced Test (NRT) items that did not count toward the Maryland School Assessment scores. In addition, the State replaced a small number of NRT items that *did* count toward the MSA score in each of the reading and mathematics tests with substitute items that Maryland developed and field-tested before use. The replacement items were built to exacting standards to be of the same difficulty and item type (selected response) as the original NRT items and to test precisely the same content. The review by Maryland's National Psychometric Council affirmed that the 2008 reading and mathematics assessments maintain the same level of difficulty as previous forms of the test—from 2007 back through 2003.

Analyses by Maryland's National Psychometric Council and by Maryland State Department of Education staff have found that students in 2008 continued the five-year pattern of gains in both reading and mathematics. Gains were particularly strong among African-American and Hispanic students as well as among students receiving special services such as English language learners, students with disabilities, and students living in poverty. Improvements among many White and Asian populations were somewhat smaller as those students reached closer to the 100% proficiency level.

This paper provides a brief history of the Maryland School Assessments and an overview of the design of the assessment. It describes the gains experienced in the 2008 administration and outlines numerous reasons that the Maryland State Department of Education has celebrated the results. Student performance in 2008 is indicative of the improved instruction seen across the State, particularly with populations of students who have benefitted from the consistency of exposure to a high-quality, aligned curriculum delivered by increasingly better qualified and better educated teachers.

Introduction

On July 15, the Maryland State Board of Education received a full briefing on the results of the 2008 Maryland School Assessment administration, which occurred in April 2008. The results showed impressive gains over 2007 and continued the long-term improvement patterns that have characterized the results since the initial year of assessment in 2003. Since the release of the assessment, several questions emerged from the media and the public about a design change in the assessment that was initiated in 2007 and implemented with the 2008 test administration. This paper will provide basic facts about the history of the assessment, its design, and the Maryland State Department of Education's analysis of the results as provided to the State Board of Education.

MSA Historical Development

The Maryland School Assessment (MSA) was first administered in 2003 in grades 3, 5, and 8 in response to federal accountability rule changes that came with the 2001 reauthorization of the Elementary and Secondary Education Act (ESEA). The President signed the reauthorization into law on January 8, 2002, under the newly minted title, No Child Left Behind (NCLB). Maryland was compliant with ESEA rules for accountability and assessments in the prior reauthorization of the law because the Maryland School Performance Assessment Program (MSPAP) provided individual student assessments at grades 3, 5, 8; however, the assessment produced no individual student *scores* in the six content areas it tested. At the high-school level, Maryland complied with the law by having its long-running graduation test in place—the Maryland Functional Testing program (MFT).

No Child Left Behind required grade-by-grade assessments in grades 3 through 8, and, at the high-school level, an assessment in reading/language arts and mathematics. Science assessments were to be included by 2008, which Maryland also implemented with the 2008 assessment administration; however, the science assessments were not intended to be incorporated into accountability calculations for schools.

In 2001, Maryland engaged Achieve, a nonprofit organization based in Washington D.C., to conduct a full policy review of Maryland's accountability system. This review ultimately helped pave the way for the state's transition into the implementation of NCLB. The review indicated that local school systems wanted grade-by-grade standards and wanted to have grade-level tests aligned to those standards that would produce individual student results—all of which were to become NCLB requirements in 2002.

Transition from MSPAP

Maryland, like many states with existing accountability systems, studied the federal requirements for tests and determined that the quickest way for the State to produce an assessment that met the new federal requirements was to develop its new assessments as "Augmented Norm-Referenced Tests." Norm-Referenced Tests (NRTs) have been in use nationally for decades and have been administered in Maryland as a part of a legislated school accountability program that had been used in previous years, and which simply involved testing and subsequent reporting. The NRT most recently used was the "Comprehensive Test of Basic Skills," a product of CTB/McGraw Hill. Each student at grades 3, 5, and 8 took the assessment, and in some school systems, the assessment was also purchased by the local school system for students in other grades.

NRTs generally are designed around a testing blueprint that includes content that the test publisher decides is likely taught at each grade level. These assessments are not generally accepted by the U.S. Department of Education (USDE) as suitable for federal accountability purposes in that there is seldom a good correlation between the test content of any publisher's test and any particular state's standards. In fact, only two states nationally use NRTs for federal accountability.

NRT Selection. Augmented NRTs have been used in eight states, including Maryland, up until the 2008 assessment. For Maryland to use the Augmented NRT design, the State went through several meticulous steps. First, MSDE brought in nearly 100 educators from around the state to review the NRTs produced by national test publishers who responded to a Request for Proposal. The Maryland educators looked at Maryland standards at their assigned grade and subject and came to consensus on the assessment that most closely matched the Maryland standards. From that

selection process, tests from two different vendors were selected. Harcourt, publisher of the Stanford Achievement Test (SAT 10), was selected for the reading assessments, and CTB/McGraw Hill, publisher of TerraNova, was selected for mathematics.

Alignment Studies. Maryland then engaged Achieve to conduct a detailed alignment study of the selected assessments with Maryland standards. Achieve's experts were tasked with mapping out in detail the items in the selected vendor tests that covered the Maryland standards. In both the mathematics and reading assessments, it was determined that there were enough items to take advantage of the psychometric qualities of the NRTs and establish the federally required definitions of proficient and advanced student performance on the new Maryland School Assessments. MSDE then worked with the vendors to produce the additional items that were aligned to content standards that were not adequately covered by NRT items, and thus the assessment instrument was completed.

Standard Setting. The results of the 2003 MSAs at grades 3, 5, and 8 were used in the summer of that year to set standards for the tests. CTB/McGraw Hill was the vendor identified to oversee the standard-setting process, which involved approximately 100 educators from around the State who were knowledgeable about the performance and instruction of students in the tested grades.

The procedure used for setting the passing scores at that time and in all subsequent standard-setting exercises is termed the "Bookmarking Procedure." The vendor reorders items used in each of the assessments by difficulty, beginning with those items that students most frequently answered *correctly* through those items that students most frequently answered *incorrectly*. Through a series of presentations of facts to participants and discussions and votes, a recommendation for the passing (or "proficient") score is generated. The same steps are then repeated to set the score for "advanced."

Participants in this exercise were posed the question, "What is the most challenging item on this assessment that all students should be expected to answer correctly if this test were administered in 2014?" This question drove standard-setting because it was important to recognize that students would need more time to build requisite skills, and teachers would need more time to bolster instruction and fully implement all of the standards. According to NCLB, 2014 is the year when all students are expected to achieve proficiency in all assessments.

Recommendations for the proficient and advanced standards were passed on to a Review and Articulation Committee—a committee of educators and administrators—which was asked to review the recommendations from the teachers and produce a recommendation for the State Superintendent. The State Superintendent reviewed all of the proceedings and made her recommendations to the State Board. The State Board's decisions on the standards are determined to be final and are applied to student results.

Grades Added. In 2003, the MSAs were administered only in grades 3, 5, and 8. To be fully compliant with federal law, tests in grades 4, 6, and 7 were added to the program in 2004, following all of the same test-development procedures used the year before for grades 3, 5, and 8.

Maintaining the Standard. Each year, test items are introduced and others are removed to allow for public release of the test, the maintenance of test security, and the fine-tuning of the test for accuracy. However, MSDE must assure that the MSAs continue to test the very same content every year and that the passing standard is maintained at the same level of difficulty. One fine-tuning activity undertaken over the past five years was the linear alignment of the scale scores in 2004—that is, all tests were placed on an identical scale, one that begins at 240 points and ends at

640 points. This is a standard practice; it adjusted the passing scores for some of the assessments but did not change their difficulty. The passing scores (or "cut scores") have been the same since 2004, and the difficulty of the scales has been maintained since 2003.

In 2007, Maryland dropped the TerraNova mathematics NRT from the MSA and replaced it with the SAT 10 mathematics test. Some shifts occurred in timing, but the scores of students on all mathematics tests were watched carefully by the Department and its advisors to ensure that the transition maintained the level of expectation used from 2003 through 2006.

National Psychometric Council. Maryland's National Psychometric Council has advised the State from the inception of its assessment programs, which date back to the 1990s. This panel of independent, nationally recognized testing experts is engaged by the State to advise on the design and building of the assessments, the analysis of each year's results, and the updating of the assessments over time to ensure the highest quality controls for the assessment program.

The assessment review conducted annually by the panel involves a standard protocol of reviews at the item and test levels. The Council looks at a variety of item and test characteristics, such as how well individual test items measure the skills and concepts they are written to test. The Council also engages in a review of all the reports and statistics from the vendor to determine if its analyses and procedures are of the highest quality. In some states—if indeed they have such a council of advisors—the technical advisory committee role may *not* necessarily include the review of the vendor's work. This is a step Maryland feels is critical to assure the highest quality standards are maintained.

From the outset, Maryland felt it was important to have the eyes of experts on the State's testing procedures and on the vendors' work to assure the highest quality of both. The Council frequently asks Harcourt to re-run analyses, explain findings, etc. In the end, the National Psychometric Council in Maryland is given wide latitude to challenge both MSDE and the test company when it feels the integrity of an assessment is in question.

2008 Adjustments to the MSA

MSDE has had ongoing discussions over several years with local school system staff, including superintendents, assistant superintendents for instruction, and local accountability coordinators, on ways to enhance the assessment program over time. Among the issues discussed are:

- **Length of Testing Time:** There was some interest in reducing the amount of time involved in the administration of the tests, even though the MSA requires significantly less testing time than MSPAP, which was administered from 1991 to 2002.
- **Test Score Turnaround:** The superintendents asked MSDE to consider strategies for returning scores to schools and parents more quickly.
- **NRT Content:** There was concern that some of the content on the NRT that did not contribute to the MSA scores may be miscuing teachers and students about what actually counts toward the MSA score and what actually represents Maryland standards.
- **Public Release of Forms:** There was interest in publishing more complete public-release forms of the MSA, which could not occur as long as Maryland used some NRT items that were held secure by the vendor.
- **Earlier Identification of AYP Status:** School and school system staff expressed interest in knowing earlier their AYP status for the next year so that appropriate planning could begin.

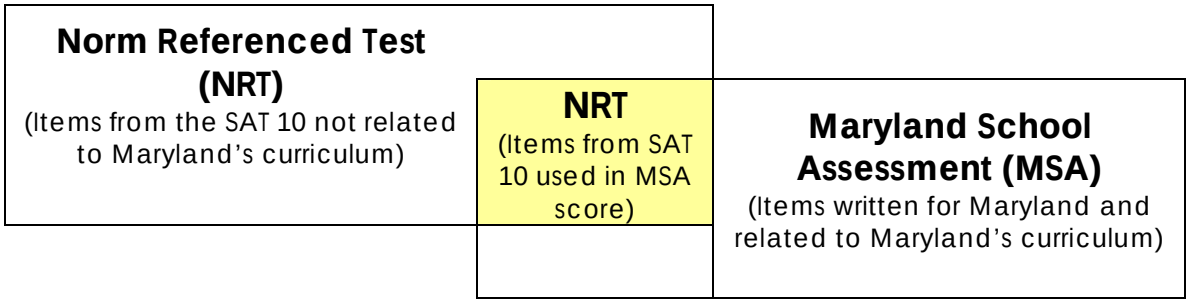
On May 11, 2007, at the final monthly meeting with local superintendents for the 2006–07 school year, Dr. Grasmick and Dr. Leslie Wilson presented a plan for eliminating the NRT component of the Maryland School Assessments as early as the 2008 administration (Attachment A). Eliminating the NRT had been discussed in prior years, and votes had been taken among superintendents. However, a sizeable number of superintendents supported the use of the NRT because they felt their publics had an interest in the NRT scores. That support, however, dwindled over time. The public seems to understand that NRTs have limited value when they do not directly assess student performance on State standards and what is taught in the classroom. By the 2006–07 school year, more superintendents supported dropping the assessment. However, no formal vote was requested until the May 2007 meeting.

Local superintendents were given a memorandum from the State Superintendent asking for a response on the proposed change (Attachment B). The response was to be submitted by the local accountability coordinators at their meeting on June 5, 2007. The June 5 submissions showed support for the change from all but two school systems—Howard and St. Mary's County Public Schools. Because of the overwhelming support for the elimination, MSDE decided that the State would move forward with the change, and the decision was announced to superintendents via a June 13, 2007, memorandum (Attachment C). Subsequent to the announcement, MSDE staff began preparations to eliminate the NRT component of the MSA. The change would take effect with the spring 2008 MSA administration.

How the NRT Was Eliminated

The implementation plan involved replacing the small pool of items initially pulled from the NRTs for the MSA. The MSA reading score was initially calculated using all of the MSA items plus up to 25 reading items from the NRT, depending on the grade level. The mathematics assessment was structured a little differently. Approximately five items from the NRT were incorporated into the score computations of the mathematics MSA. As a regular component of test development, new items are written, edited, evaluated, field-tested, and analyzed every year. Items with acceptable characteristics are entered into the item bank for future use. Therefore, the items selected to replace NRT items actually measured precisely the same content as the original items.

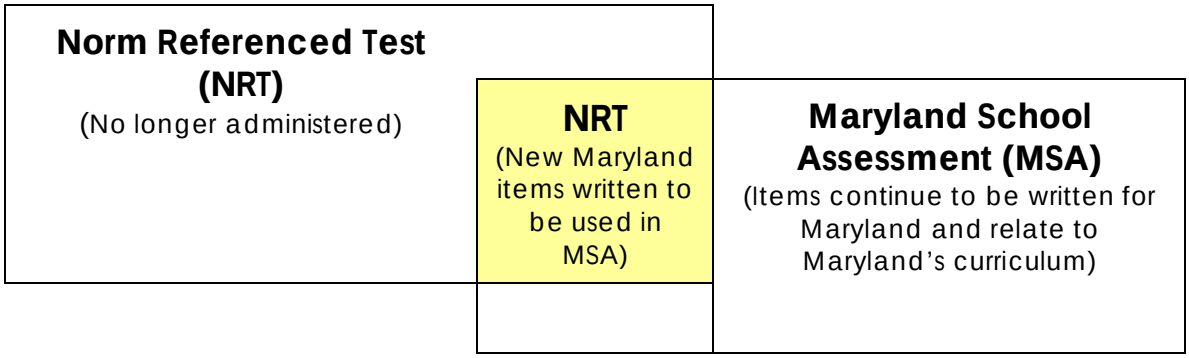
The MSA Augmented NRT design appears below:



The transition was possible in 2008 because MSDE had actually field-tested replacement items in the 2006 MSA administration due to concern that year that, given the required rebidding of the MSA contract, there was no way to know if Harcourt (publisher of the NRT) would win the contract for the next cycle. The items were field-tested, and the resulting data on the performance of those

items were put aside for future use. However, because Harcourt *did* win the rebidding of the contract, there was no need for the items at that time.

The MSAs moved away from the Augmented NRT design beginning with the 2008 MSA administration, as diagrammed below:



Reconfiguring Testing Time. For each of the two tests—reading and mathematics—the original design of the MSA involved approximately two-and-a-half hours of testing on Day 1 and a little less than two hours on Day 2. The reading MSA was administered over two days during the first week of testing, and the mathematics MSA was administered over two days during the second week of testing.

The NRT was completed by students on the first day of testing, involving approximately a half-hour block of work time during which the student completed an array of multiple-choice questions. The NRT items had to be administered as a set of questions in a particular sequence because the behavior of those individual items as measurement tools could be influenced by questions appearing immediately before or after any one of them.

Reconfiguring Testing Items. From 2007 to 2008, 50–70 multiple-choice items (depending on grade level) were removed from the MSA reading exams. However, those items were used only to generate norm-referenced scores. As is shown in the table on page 8, the total number and type of items used to generate MSA scores remained the same. MSDE simply developed some multiple-choice items that were previously taken from the SAT 10.

In 2007, students worked approximately three hours over two days on the reading MSA and, during the following week, a similar amount of time over two days on the mathematics MSA. The NRT consumed about a half-hour of student work time for each of the exams during Day 1 of testing. With the 2008 administration, students were able to complete the assessment in less time. (In reading, the 2008 testing time was reduced from 15 to 29 minutes by eliminating the NRT, pending the grade level.) Through 2007, once the NRT was completed, students were provided directions for the MSA itself, and students then worked the remainder of the first day on certain sections of the test.

Items Contributing to MSA Scores in Reading

	2007	2008	Change in Number of Items from 2007 to 2008
Multiple Choice			
Developed by MSDE	8	33	+25
From SAT 10	25	0	-25
Total	33	33	0
Constructed Response	4	4	0
Total Number of Items	37	37	0

Note that the NRT differed from the MSA in several ways. The items were all selected-response or multiple-choice items, unlike the MSA, which includes both constructed-response and selected-response items. Further, the content of most of the NRT was frequently mismatched with Maryland standards, and questions were not always of the style and quality expected on Maryland tests. Hence, there was some concern that these items were misdirecting teachers about the real Maryland standards on which their students were tested. They were also frustrating to some students who were faced with items assessing content they had not been taught.

In 2008, unencumbered by the NRT, Maryland carefully incorporated the previously developed substitute questions into the MSA and ensured that their placement did not disrupt the measurement qualities of other items. It is important to note that Maryland employs a rolling test development procedure in which new test items are continually developed and field-tested in each year's tests. Those field-test items that are proven to measure accurately are placed in a pool of test items that may be used in future test forms. In order that students and teachers do not become too familiar with the items, the MSA is rebuilt each year with either some newly field-tested items that have never counted in the test or with items used two or more years previously. This procedure helps the test to look fresh and new each year without compromising the standard or design continuity.

Pre-Release Analysis Procedures for the 2008 MSA Results

The results from the 2008 MSA were analyzed using the same procedures that had been followed in previous years—with one small added step. Each of the NRT items that had counted in previous MSAs were now gone. The substitute items measured the same skills and knowledge. The vendor and the Psychometric Council analyzed student results on those items to determine if these new items worked. It was important that they be fair for all students and not functionally easier or more difficult. Additionally, the overall performance of students on the remainder of the MSA needed to mirror the performance in previous years. In both instances, the Psychometric Council was confident that the test operated as planned (Attachment D).

Initial results showed somewhat higher than expected performance in grades 5 and 7, and the Council asked the vendor to re-run its analysis of the data. The re-run data confirmed the initial

analysis. In one of several conference calls, the Council briefly discussed the grades 5 and 7 performance and noted that the scores were still within range of what it would expect as acceptable, especially given the performance of those groups in grades 4 and 6 the previous year. But the Council conjectured that the students may have experienced a little less fatigue by not having to complete the NRT on Day 1 of the test. In the final analysis, the Council determined that the results were suitable to publish and confirmed that the difficulty of the assessment and the performance standard had been successfully maintained from 2007 to 2008.

The National Psychometric Council annually confirms the work of the vendor and the Department for all tests and directs the vendor to conduct additional analyses if they are warranted. Typically, the Council will recommend to the Department the publishing of scores if its conclusion is that the scores are valid measures of student performance in that administration. However, if the Council were to find that any year's assessment had not functioned properly, it could advise the Department to invalidate the scores and not publish results for that year or order the vendor to go back and reassign passing scores that maintain the difficulty level of the test from the previous year. Thus far—including the 2008 administration—the Council has endorsed results each year and has not had to invalidate any year's results.

Limits on the Psychometric Council's Work. Unless the Council confronts serious issues that would interfere with the publication of the results, its discussions are not typically shared with Department leadership—except as they appear in the final Technical Reports produced after each assessment. If a problem surfaced in the Council's work, the State Superintendent would be notified immediately.

To be clear, the Psychometric Council's work with test results is limited to the validation or invalidation of test scores. The Council has no knowledge of the instructional factors that might affect test results, nor is it aware of the funding improvements, master-planning, technical assistance, or any other related changes that are at play. Its expertise is specific to the test instrument and its administration and scoring. It is the work of MSDE and policy analysts to put the pieces of the puzzle together with the assessment staff.

Analysis of Grades 5 and 7 Reading. For example, the Psychometric Council noted that the 5th- and 7th-grade reading scores rose more than was expected and more than other grades. Their analysis and the analysis of Department staff found that the amount of total test-time reduction at those grades was actually in the 15- to 16-minute range—among the smallest reductions in time across the six grades. It is also true that 5th-graders' reading performance improved only 0.7 percentage points over their 4th-grade performance a year earlier. The 7th-graders had improved 4.6 percentage points over their performance in the 6th grade.

Test Time Differences, by Grade: Reading

Grade	Reading		Difference
	2007	2008	
3	178	150	-28
4	178	149	-29
5	173	157	-16
6	173	157	-16
7	173	158	-15
8	173	158	-15

This data indicates that the 5th- and 7th-graders were actually poised to perform at higher levels in 2008 as evidenced by their scores the previous year. Furthermore, 3rd- and 4th-grade students improved their performance more modestly from 2007 to 2008, yet these two grades each experienced a 30-minute reduction in testing time from 2007 to 2008. See the table below:

Student Performance Changes, by Cohort

Year of Testing	Grades	Reading (% proficient+)	Grades	Reading (% proficient+)
2007	4	86%	6	76.7%
2008	5	86.7%	7	81.2%
Difference		+0.7%		4.6%

The lack of consistency in the gains across grades raised doubt about the possible impact of the reduction in testing time; the patterns lead one to believe that additional factors are responsible for the improvements. Because the Psychometric Council's charge is to ensure the structural integrity of the assessment itself, its conjectures are limited to the impact of the tests' timing changes. However, Department staff have noted other, more compelling reasons for the rise in scores in 2008: instructional improvements; a reform climate in Maryland that has driven accumulated changes in public school quality; extensive training of teachers and principals by MSDE and local school systems, training that ensures teachers understand State standards and are able to incorporate them into their lessons.

Policy Analysis Associated with the Gains. The improvements in services for young children in Maryland are unprecedented, and among all of the policy changes that have an impact, this is one of the most important. Prekindergarten programs, Judy Hoyer Centers, the Maryland Model for School Readiness, the consolidation of early childhood services at MSDE, and Reading First programs have cumulatively ramped up services for all children.

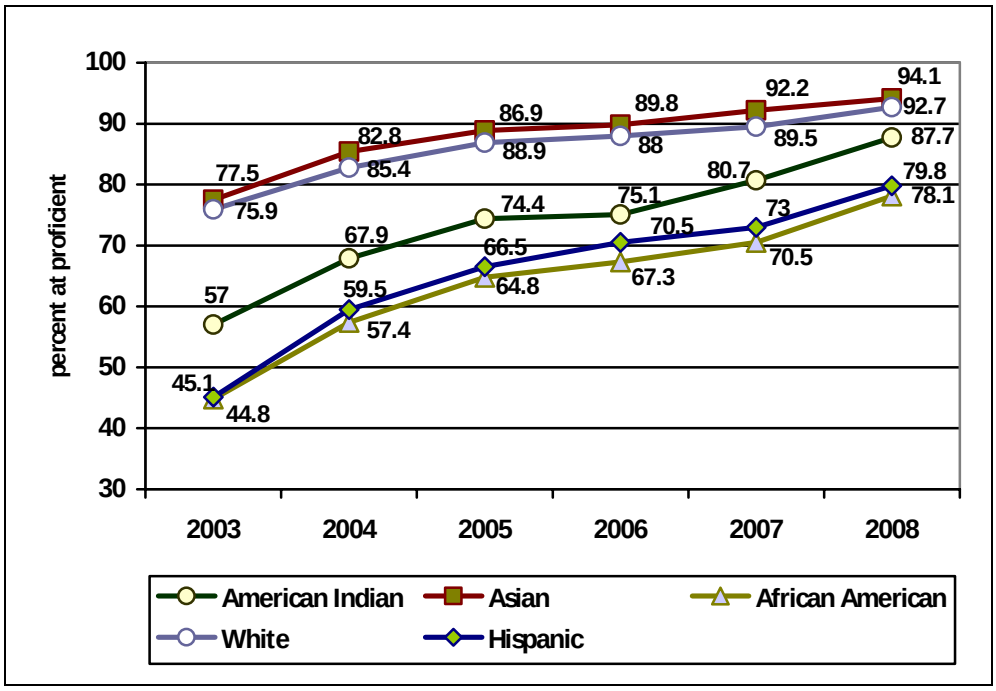
The infusion of $3 billion in additional State aid to local school systems through the Bridge to Excellence Act of 2001 has had a huge impact. With the additional aid came the elimination of perhaps 25–30 separate programmatic funding streams to local school systems, not all of which were well-coordinated. These funds were consolidated into a single aid package, and local school systems have unprecedented latitude to allocate these funds as they see fit. This departure from past equity-based funding programs brought new accountability to local school systems.

The Department requires that each school system develop a master plan that explains, every year, how all of its state, local, and federal aid is being used to improve student performance. These plans are typically reviewed by panels pulled from a statewide pool of more than 100 experts. The panels produce an analysis of each school system's master plan, identifying gaps and strengths. The review panels' work often leads to school systems shifting their emphasis and strategies to ensure a strong focus on the student groups and schools in greatest need.

The gains in 2008 were strongest in elementary schools, among African-American and Hispanic students. They were strong among students in special education and students learning English. Gains were weaker among White and Asian students. Therefore, the gaps that divided White and Asian students from their African-American and Hispanic counterparts shrunk in 2008. This generally followed the pattern of gap reduction that we have seen over the past several years. Clearly, the gains were made where MSDE leadership expected them—largely in groups where

resources were focused and, indeed, where there was significant room for growth. See the graph below.

Student Performance, by Race: Elementary Reading



Long-Term Trend Lines. In two decades of accountability and testing in Maryland, we have studied results from year to year for the Maryland School Performance Assessment Program, the Maryland Functional Testing program, the Maryland School Assessment program, and the High School Assessments. We have also studied State NAEP results, and results on the SAT, ACT, and AP. A few observations accumulate from these testing studies:

- **Change sometimes takes time**. Initial slow growth is sometimes followed by more aggressive gains over time. It is hard to locate a one-to-one relationship between any one year's program changes and changes in test scores.
- **Curricular specificity has a greater impact on scores**. During the 1990s, the general grade-band instructional frameworks gave teachers imprecise information on what was to be taught each year, and therefore did not lead to aggressive gains in student performance in many schools. With the MSA, specific grade-by-grade standards help teachers by clarifying what will be tested and what should therefore be included in instruction. Gains in MSA scores have been very strong over the five years of the assessment program.
- **The trend line provides more insight than any one year's uptick or decline**. In all of the assessment programs over time, school-level data have been more volatile, and MSDE and local school system staff will be examining school-level data to see what additional information can be gained to explain the volatility.
- **The score trends at specific grade levels focus on a different group of students every year**. Each group of students has a different sequence of support, and as years go by, each group has benefited from increasing amounts of support. By following cohorts, it is increasingly evident that the skills accumulated by students over time help their performance in upper grades and on into high school.

Maryland Standards Hold Up Well

A Fordham Foundation report released in October 2007, entitled *The Proficiency Illusion*, is incorrectly referenced in the June 17 *Baltimore Sun* article that addressed the 2008 MSA adjustments. Contrary to the *Sun*'s claims, the report does not rank the MSAs 26th in difficulty when compared to other state assessments. The report's methodology and findings are very problematic and do not aid in understanding where Maryland standards actually stand.

The report compares results from an assessment developed by the Northwest Evaluation Association (NWEA), called MAP, to results on state tests. The report does not actually examine each state's assessment or standards to check them for quality or difficulty. The report bases its conclusions on the performance of often small, unrepresentative samples of students on the MAP. States with better results on the MAP than on state assessments are said to have more challenging assessments, while states that do better on state assessments are said to have less challenging assessments.

The more likely explanation is that differences in state assessment results and MAP results are a function of how well the specific standards tested by MAP happen to overlap state standards. This is precisely the same problem that emerges when one compares proficiency levels on state tests with NAEP results.

Numerous flaws limit the value of the study. The MAP is, in fact, used by only a few districts in 26 states, and the conclusions are therefore based on a small, unrepresentative sample from each state. Furthermore, the report is written by the developers of the MAP and begins with the assumption that the MAP is a high-quality assessment that is aligned to states' standards, but offers no evidence of the real quality of the assessment. When NWEA calls the MAP a state-aligned test, this does not mean that the test has been devised with state standards in mind or that the MAP is customized for each state. This means only that NWEA has calculated what scale score on the MAP corresponds to proficiency cut scores on state exams—which is *not* how most people would define alignment.

Maryland is ranked 18th out of the 26 states studied, and that ranking is based on the fact that Maryland students perform better on the MSAs than they do on the MAP. However, since the MAP does not cover the same standards as the MSAs, there is no reason to assume that our students would do as well on the MAP as they do on the MSAs.

The report says that states have more demanding expectations in later grades than in earlier grades. NWEA's purported evidence is essentially that fewer students pass the exams in the later grades. The organization believes that if 40% of students reach proficiency in 3rd grade, 40% of students should reach proficiency in 8th grade. NWEA characterizes their failure to do so as a failure of test design. There is no way any state will win this flawed argument when the study's authors frame their conclusion in this way. If more students were reaching proficiency in 8th grade than in 3rd grade, the authors would contend that the test is too easy in 8th grade, rather than contending that education is improving and adding value to students' learning the longer they are in school.

The report also states that tests appear to be getting easier over time. Maryland is one of the states singled out. NWEA's evidence for this is that MAP performance has stayed the same over time, while more students are reaching proficiency on the MSAs. If students are being taught the standards on state tests, they *should* improve over time as instruction and alignment improve. Since MAP standards are *not* what students are being taught, there is no reason why MAP scores should improve. It is also worth pointing out that Maryland's NAEP scores are improving at the same time

that MSA scores are improving—so, by that measure, our tests are not getting easier. We're just doing better.

The report contends that state math tests are harder than state reading tests. However, poorer performance on math tests does not mean they are harder. It could simply mean that students are not as well prepared in math. That makes sense when, in comparative studies with other countries, the U.S. spends more time on reading instruction and less time on math instruction than most other countries do.

NAEP Data Confirm Maryland Improvements. There is actually quite a bit of national data, especially from NAEP, showing that achievement in Maryland is higher than in most states. The NAEP results also verify that the growth in MSA achievement is not an illusion—it's been mirrored in NAEP results. For example, Maryland students improved their scores across the board on NAEP's 2007 mathematics and reading tests, when compared with 2005 results. Fourth- and eighth-grade students made steady gains between 2005 and 2007. This continues a record of improvement on most of the NAEP exams for the past seven years.

The mathematics scale score for Maryland's 4th-grade students increased from 238 in 2005 to 240 in 2007, while the mathematics scale score for 8th-grade students jumped from 278 to 286. The reading scale score for 4th-grade students jumped from 220 to 225, and the scale score for 8h-grade students climbed from 261 to 265. In 2007, Maryland's scores were above the national average in all categories.

Only five states scored statistically higher in 8th-grade mathematics than Maryland (Kansas, Massachusetts, Minnesota, North Dakota, and Vermont). Only five states scored statistically higher in 4th-grade reading (Vermont, New Hampshire, Massachusetts, New Jersey, and Rhode Island).

The gap between the scores of White students and African-American students has diminished over time, even as scores for all students have improved. In grade 8 reading, for example, African-American students have closed the gap with White students by five points since 1998. Hispanic students have closed their gap with White students by seven points over the same period.

Maryland's special-education students have posted higher scale scores in grades 4 and 8 than the nation since 2000 in mathematics and since 2002 in reading.

For Future Maryland Study

Among the issues to be studied in the future is the distribution of scores between the Advanced and Proficient categories. In most grades, more students are scoring at the Advanced level than originally expected. Initial conjecture attributes these changes to the higher quality of instruction brought about by the Voluntary State Curriculum and the effect of that instruction on students at all performance levels. The strategies employed by schools seem to be focused on improving instruction across the board, rather than on simply moving students in the Basic category to Proficient. An investigation of this issue will help determine whether the Advanced cut score is appropriately set (and improvement is merely an outgrowth of higher quality instruction), or whether the Advanced category is mislabeled. The National Psychometric Council has not been asked to delve into the issue of Advanced scores because these scores are strictly informational at this time and do not affect accountability.

A second issue of greater importance is an examination of the standards to determine two things. First, it is important to examine the content taught at various grade levels up through high school to assure that the trajectory of learning is appropriate and that students are being prepared adequately for the next phase of learning. The second is to examine the Proficient cut score to assure that it is appropriate. The standards were set in 2003 (grades 3, 5, and 8) and 2004 (grades 4, 6, and 7) under the assumption that the standards are achievable by all student subgroups by 2014, as required by No Child Left Behind. A reexamination of the standards at this time can be better informed by the past five years of performance. However, MSDE has been very cautious when making incremental changes to the assessment program and will continue to make any further changes in a deliberate and careful manner.

Summary

The results of the 2008 Maryland School Assessment are fully comparable with the prior forms of the assessment administered from 2003 through 2007. The small adjustments implemented in 2008 were carefully orchestrated over time and with the guidance of Maryland's National Psychometric Council. The Maryland State Department of Education published the scores in July 2008 with full confidence that the results are valid and reliable.

The results were fully adequate for reporting student performance at the state, school system, and school levels and were sound for the purpose of making Adequate Yearly Progress determinations at all three levels. The National Psychometric Council's discussions regarding the possible impact of test-time reductions on student performance were fully warranted and necessary, and their conclusion that the 2008 assessment was scaled at the same level of difficulty as prior test forms was based on an accurate analysis of the results.

The Department's analysis of the policy changes associated with instruction over the five years of No Child Left Behind—increased local resources via the Bridge to Excellence Act, mandated local master planning, a focus on low-performing student groups, and local efforts to increase the number of qualified classroom teachers—supports the validity of the results. Essentially, better, more focused assistance channeled to children living in poverty, English language learners, and students with disabilities has improved results. Accompanying the instructional improvements is an increased emphasis on assuring a stronger start for young children through expanded preK programs, school-readiness measurement tools, and curricular improvements in kindergarten. Furthermore, the early identification of students with learning challenges has assured earlier intervention on their behalf.

Future gains on the Maryland School Assessments may be more challenging for those subgroups now frequently approaching 90% proficiency. However, with gaps shrinking—yet still remaining—for students receiving special services and for African-American and Hispanic students, the continued commitment to all children will continue to yield gains. Cohort analyses tell us that students who perform well in lower grades continue that performance in higher grades—better than their predecessors. This sustainability of performance will fuel changes in the future. It will be important for the Maryland State Department of Education to monitor progress, continue successful strategies for low-performing students and schools, and develop with local school systems further tools to assure that teachers can be even more successful in the future. ■

Attachment A

Agenda

May 11, 2007

State Superintendent's Meeting

Strathmore Hall, Room 402
Montgomery County

I. Discussion

9:30 a.m.	A.	Accountability/Assessment Issues ♦ Algebra II Field/Pilot Testing ♦ HSA October and April Administration Options ♦ Graduation Rate/AYP ♦ NRT	Ron Peiffer Leslie Wilson
10:20 a.m.	B.	AED Draft Regulations	Donna Mazyck

II. Announcement

- ♦ Budget Update
- ♦ Online Course

III. Information

051107AGENDA

Attachment B

MARYLAND STATE DEPARTMENT OF EDUCATION
Achievement Matters Most

Nancy S. Grasmick
State Superintendent of Schools

200 West Baltimore Street • Baltimore, MD 21201 • 410-767-0100 • 410-333-6442 TTY/TDD

TO: Local Superintendents
Local Accountability Coordinators

FROM: Nancy S. Grasmick

DATE: May 11, 2007

RE: Considerations for Eliminating the Norm Referenced Test from MSA Reading and Mathematics

---

Background

When the No Child Left Behind legislation was enacted, Maryland needed to move from the MSPAP testing program to a program that would test students in Grades 3-8 and provide individual scores. One efficient way to make the transition was to utilize an intact standardized test, choose items that align with the Maryland's Voluntary State Curriculum (VSC) and augment them with additional items to ensure coverage of VSC. This presented an opportunity to combine the advantages of a norm-referenced test (the intact test) with a criterion-referenced test (the Maryland School Assessment (MSA)). However, it did mean that students had to complete the entire NRT even though only some of the items would count on the MSA.

Over the years, as the MSA scores have gained their own meaning, the norm-referenced test scores, while somewhat useful to some parents, have been used less and less by local school systems. During testing, teachers often question items that are not aligned with the VSC. These are the items that are part of the NRT only, but cause confusion, and some teachers may even begin to teach the content of these items, even though the items will not count on the MSA.

Current Considerations

We have been reassessing the original decision to continue the NRT, and we are requesting input from your local school system about this consideration. To improve testing for the 2008 school year, we need the feedback from your school system so a final discussion can take place at the June 5 meeting of the Local Accountability Coordinators (LACs). The LACs need to be prepared to represent your system's views in the June 5 discussion, as the outcomes of the meeting will be used as the basis for any recommendation that is made to the State Board for consideration.

Here are some advantages to eliminating the NRT that each school system should consider:

- The NRT is not aligned with the VSC. Eliminating the NRT will eliminate the items which teachers say are not aligned with the VSC.
- The NRT requires almost two full days of MSA testing for reading and mathematics. Eliminating it would reduce MSA testing time.
- The NRT items are not the property of MSDE.
    - Their use has inhibited us from releasing a complete MSA form on our web site, making the MSA-release process differ from that followed for HSA and MSA Science.
    - The items cannot be modified as we construct the Modified MSA tests in reading and math. Without the NRT items, we would have more flexibility in creating the Mod-MSA and would be able to create a test of appropriate length for the students who are eligible to take the Mod-MSA. (To ensure the content is covered for students taking the Mod-MSA and to produce a valid and reliable test, we would need to increase the length of the Mod-MSA beyond the current length of the MSA.)
- The NRT requires standard administration directions that cannot be edited by MSDE. This makes some of the directions that test examiners give on the MSA confusing to teachers and students. Eliminating the NRT would allow MSDE to create examiner materials that are completely appropriate for Maryland students.

Next Steps

1. Please review the background information and current considerations and discuss this information with your staff.
2. Determine your system's position regarding the NRT portion of the MSA.
3. Complete and sign the attached system preference form and have your LAC bring it to the June 5 LAC meeting.

We greatly value your input on this important issue. If you have questions, please contact Leslie Wilson (410) 767-0027 or email her at lewilson@msde.state.md.us.

Response Form

School System Preference on the Continued Inclusion of the NRT on the MSA
(Please complete and send with your LAC to the June 5 LAC meeting.)

School System Name _____
LAC Name _____

Preference for NRT (check one)
____ Keep the NRT on the Grade 3-8 reading and mathematics MSAs
____ Eliminate the NRT from the Grade 3-8 reading and mathematics MSAs

Superintendent's Signature _____

Attachment C

MARYLAND STATE DEPARTMENT OF EDUCATION
Achievement Matters Most

Nancy S. Grasmick
State Superintendent of Schools

200 West Baltimore Street • Baltimore, MD 21201 • 410-767-0100 • 410-333-6442 TTY/TDD

June 13, 2007

MEMORANDUM

To: Local Superintendents

From: Nancy S. Grasmick

RE: Elimination of Norm-Referenced Testing

I want to thank you for your thoughtful consideration and cooperation to provide me with your recommendation concerning the continuation of norm-referenced testing as part of our Maryland School Assessment Program (MSA). We received feedback from all school systems in a very timely fashion. The overwhelming response was that we should proceed with elimination of the norm-referenced test. This will provide us with the following advantages:

We will be able to:

- Ensure that all items administered on the MSA are aligned with the Voluntary State curriculum.
- Develop test instructions that are more appropriate for Maryland students
- Shorten MSA testing time.
- Release, for public review, entire forms of the MSA.
- Produce a reliable and valid modified version of the MSA without increasing test length.

The assessment staff will begin work immediately to make the required adjustments to the testing materials so that we can take advantage of these benefits for the spring 2008 testing. With this decision, I think we have experienced an example of a significant way that we can work together to improve how we assess our students and monitor their progress. I hope that we continue this collaboration and improve assessment and instruction for all of our students.

As always, thank you for your dedication and hard work on behalf of Maryland's public school students.

Thank you.
NSG:l

Attachment D

National Psychometrics Council

Bert Green, Ph.D. Johns Hopkins University, Emeritus
Huynh Huynh, Ph.D. University of South Carolina
Robert Lissitz, Ph.D. University of Maryland
William Schafer, Ed.D. University of Maryland, Emeritus
Steven Wise, Ph.D. Northwest Evaluation Association

Memorandum

To: The Maryland State Board of Education

From: National Psychometric Council

CC: Janet Bagsby

Date: August 1, 2008

Re: 2008 MSA linking

Given the recent media coverage of the MSA results and the suggestion by some that the 2008 MSA was less difficult than it was in 2007, we would like to make it clear that, in the opinion of Maryland's National Psychometric Council (NPC), the test scores on the 2008 MSA in reading and mathematics were properly linked to the score scales of the baseline year. Because of the quality of the linking, we are confident that the 2008 MSA gains in the percentages of students achieving at the proficient and advanced levels are not a consequence of the 2008 assessments being less difficult than in previous years. Below we outline the process we used and the rationale for our conclusions.

By way of background, the National Psychometrics Council provides advice and recommendations to MSDE and its assessment vendors on procedures to help ensure that Maryland's assessments are technically sound. We do so by reviewing test designs, protocols for the development of test items and test forms, administration protocols, and by monitoring the psychometric characteristics of the test forms, linking and equating procedures and quality control results, and the procedures used to compute individual student scores.

In the spring and summer of 2007, when the 2008 MSA was being designed, we discussed the potential impact of removing the NRT portion of the test and the necessary adjustments to the linking procedures. The statistical linking protocol that was used for the 2008 MSA is identical to the one used in previous years (including 2007) for MSA and is similar to those used for linking across years in many other states in their assessment programs. The protocol calls for use of

linking items that are reasonably stable from year to year and that meet industry standard statistical benchmarks. As it does every year, the NPC conducted multiple reviews of the 2008 linking results before making its final recommendations to the Department on the linking procedures for official scoring purposes.

Following the 2008 test administration, we reviewed the psychometric characteristics of the linking items, the procedures used to link the 2008 assessment to the baseline year, and the nature of the resulting score distributions. Based on our initial review of the technical nature of the linking items, we asked the vendor to conduct further analyses on the stability of the initial item linkings for all reading tests. We examined the consistency among various technical elements including item parameters and mean scores. Based on these analyses we concluded that the linking procedures were properly done, the final linking results were appropriate, and the 2008 scale was appropriately linked to the baseline year scale. We did not find anything in the data that suggested a lack of technical accuracy of the linking results. Subsequently, we examined the impact data (the percentage of students achieving at the proficient and advanced performance levels) and concluded that the increases were reasonably consistent with historical trends.

The statistical analysis of the MSA results offers no explanation for the observed improvement in student achievement. Of course, progress in student achievement is the goal of the current federal “No Child Left Behind” program. Speculation about the reasons for Maryland’s observed substantial improvement includes the reduced testing burden on the students, as well as changes in policy, funding, instructional focus, and instructional interventions. Sorting out the relative contributions of these and other factors would require extensive educational research, with little promise of definitive answers. Nevertheless, we would be willing to provide advice and support for such research.

Please contact us if you have questions or comments.