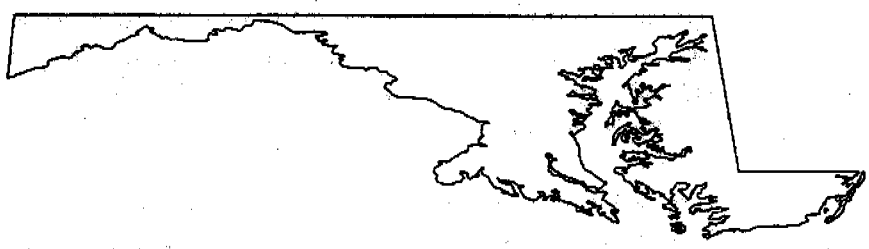
High School Assessment Design

A Report to the Maryland State Board of Education



Maryland State Department of Education

Wayne Camara
William Harris
Howard Everson
Robert Majoros
and colleagues

The College Board

Ernest Kimmel
Tim Ligget
Patricia McAllister
Elissa Greenwald
Walter Jimenez
Patricia Klag
Susan Krupka
and colleagues

Educational Testing Service

January 1997

In September 1996, the College Board was awarded the contract for the Design of the Maryland High School Assessment. Educational Testing Service (ETS) is serving as the subcontractor for this project. The work reported in this publication was supported by a contract from the Maryland State Department of Education.

The College Board is a national nonprofit association that champions educational excellence for all students through the ongoing collaboration of more than 3,000 member schools, colleges, universities, educational systems, and associations. The College Board promotes--by means of responsive forums, research, programs, and policy development--universal access to high standards of learning, equity of opportunity, and sufficient financial support so that every student is prepared for success in college and work.

Educational Testing Service (ETS) is a private, nonprofit corporation devoted to measurement and research, primarily in the field of education. ETS offers a wide range of products and services to improve education through the design, development, administration, and interpretation of high quality measurement tools. An extensive program of research on measurement theory, teaching and learning, and educational policy supports the programs and services provided to the educational community.

The College Entrance Examination Board is located at 45 Columbus Avenue, New York, NY 10023. Educational Testing Service is located on Rosedale Road, Princeton, NJ 08541

CONTENTS

ACKNOWLEDGMENTS

The purpose of this interim report is to present preliminary findings about the design and use of the Maryland High School Assessment to the Maryland State Board of Education and Maryland State Department of Education. Between September 1996 and January 1997, staff from the College Board and Educational Testing Service have worked collaboratively with the Maryland State Department of Education on this project.

We would like to thank staff in the Division of Planning, Results, and Information Management (PRIM), especially Mark Moody and Steve Ferrara who worked with us extensively throughout all phases of developing and evaluating various design options and their implications for the HSA. Gail Goldberg and Janet Bagsby of PRIM were also especially helpful in many aspects of this work. Dan Gadra, Ray Keech, and Ronald Peiffer coordinated the public engagement activities and were responsible for scheduling, over a very short period of time, and ensuring the success of more than 40 meetings with various key stakeholder groups throughout the state. They also provided important perspectives on many issues involved in the design of the HSA. Ted Schuder and Jessie Pollack from Research and Development assisted us with many of the most complex and important issues involved in the assessment design. Staff in Curriculum and Instruction were essential in providing us with the important perspectives of curriculum leaders and content teams in each of the subject areas. Trude Collier, Mary Jo Comer, Elaine Crawford, Gary Hedges, and Diane Johnson worked directly with the content area staff at ETS to ensure that the designs included elements important to the content teams. Equally important, they continued to work with content teams and teachers across the state to maintain an open dialogue on the potential designs and their related constraints on other aspects of the HSA. We are also extremely appreciative of the hundreds of Maryland citizens, including parents, students, and educators who shared their input, perspectives, and concerns at the various town meetings and local meetings held this past fall. Finally, we wish to thank Nancy S. Grasmick, State Superintendent of Schools, and the Maryland State Board of Education who provided explicit objectives for the HSA and a vision of the integration of this assessment within the Maryland School Performance Program. We will rely on the assistance and collaboration of these groups and individuals during the second phase of the test design contract.

AN OVERVIEW

Over the past several years, Maryland has created a comprehensive accountability system to support and encourage the systemic reform of K-12 education throughout the state. The High School Assessment (HSA) program has been proposed as the culmination of this system. More than two years of intensive discussion among stakeholders has led to identification of a lengthy set of desired features for the HSA. Many of these features were included in a Request for Proposal through which the College Board, along with its subcontractor, Educational Testing Service, were chosen to collaborate with the Maryland State Department of Education (MSDE) on the design of the HSA. A critical goal of this design process is determining whether and how the desired features can be incorporated in the HSA. The Maryland State Board of Education (MSBE) has proposed an extremely ambitious approach for the HSA -- an approach that is both innovative and unique among state assessments, as well as challenging because of the many proposed uses and desired features.

Over the past five months, the contractor and MSDE staff generated a number of design options and evaluated those options against several criteria. The contractor also worked with MSDE staff in conducting an extensive series of public engagement meetings with stakeholder groups throughout the state. The process and materials used, and the information gathered during these public engagement meetings are documented in Appendices B and C. During these meetings, a number of features above and beyond those originally enumerated were identified as highly desirable by one or more groups of stakeholders.

The total set of desired features identified by the stakeholders encompasses a very ambitious assessment program. Unfortunately, not all of the features identified are compatible. Thus, choosing particular features precludes the inclusion of others. This report does not attempt to provide detailed prescriptions about all features and elements of the High School Assessment. Rather, the goal of this report is to identify those key choices that must be made before the design of the assessment program can proceed and to reflect the diversity of thinking expressed by educators and other key stakeholders concerning these choices. We have attempted to present objective information on key aspects and assumptions related to the HSA. We have also attempted to provide several alternative options or choices for these key issues where possible. The Maryland State Board of Education (MSBE), as the representative of all the people of Maryland, is in the best position to indicate which features should be given greater weight. Therefore, the contractor and the MSDE staff are seeking the State Board's guidance as to which trade-offs are most in keeping with the goals and intentions of the HSA.

This report focuses attention on those choices that must be made before the second phase of the contract can be carried out, that is, the design of the HSA and the assessments that are part of it. These choices, which are central to the future shape of HSA, range from the conceptual and legal to the practical and political. People who care about the education of the children of Maryland can, and do, disagree about the relative importance of various design features. What is clear, however, is that all desired features cannot be incorporated.

There are other issues, identified in Appendices A and B, that must be resolved prior to implementation of HSA but are not integral for the design process to proceed. However, to the extent that the State Board can resolve some of these issues early on, it may ease public misgivings and concerns about HSA as well as provide for more productive discussions with key stakeholders.

Organization of the Report

In the report we have identified six major areas where critical issues have emerged that must be considered by the Maryland State Board of Education (MSBE) in the immediate future to permit the design and initial test-development work to proceed on schedule. A brief introduction is provided to each of these six issues, followed by options for addressing them. Finally, policy questions concerning these issues are posed. While virtually hundreds of decisions must be made before the HSA can be implemented, the questions highlighted in the report focus on the most critical elements to be considered by MSBE and MSDE in the short term.

Appendix A provides additional elaboration on or discussion of many of the issues. It identifies important constraints, strengths and weaknesses, and trade-offs associated with many of the options. The constraints and trade-offs listed do not represent barriers to the development and implementation of the HSA, but are important to consider in making critical decisions regarding policy and operational aspects of the HSA. Appendix B provides an overview of the public engagement process and a summary and classification of the comments received from key stakeholders across the state. Finally, Appendix C, which is separately bound, contains materials used in public engagement meetings and a summary of the comments made at each meeting.

STRATEGIC ISSUES IN THE DESIGN OF THE HIGH SCHOOL ASSESSMENT PROGRAM

The issues described below require decisions by the Maryland State Board of Education before the HSA design work can be completed. Each issue is described briefly here with further elaboration given in Appendix A. **Although the issues are presented serially, it is important to note that they are interrelated and that decisions about one issue will almost certainly affect others and will limit MSBE's range of choices.**

I. STANDARDS AND THE USES OF THE HSA

A. Proposed Uses of the HSA

Three uses were initially identified for the results of the HSA:

- Individual Student Accountability
- School Accountability/Program Improvement
- Higher Education Placement and/or Admission Decisions

The preponderance of discussion and comments during the public engagement meetings focused on the first of these. Proponents of using the HSA for individual student accountability argued that:

- Students need to become responsible partners in their own education.
- High stakes for individual students will motivate parents and teachers in all grades to give greater emphasis to the development of the knowledge and skills outlined in the Core Learning Goals.
- Holding each student accountable is the best way of enhancing the value of the high school diploma.
- It would allow for verifying that the state has met its constitutional responsibility for ensuring the quality of the education received by all students.
- Expectations need to be raised for all students; this is best done by holding each student accountable.

The opponents of using the HSA for individual student accountability have argued that:

- The quality of education can be better improved through a focus on school accountability/ program improvement, as has been demonstrated by the use of MSPAP.
- Individual student accountability creates more practical and policy problems than it does incentives for improved learning. Problems include how to handle transferees, absentees, ESL and special needs students, and the logistics of record-keeping.
- Holding all students to the same high standards will result in an unacceptable failure rate across the state because of student variation and individual differences that are found in all indicators of student abilities and performance.
- There will be a disproportionate rate of failure among minority and/or poor students, creating major legal and political challenges to the HSA system, and/or resulting in the lowering of standards over time.

- The individual school/district is in a better position to judge whether a student meets graduation requirements, including, but not limited to, the Core Learning Goals.
- Preparing for the HSA tests will become the de facto curriculum, narrowing the scope of what is dealt with in each of the 12 relevant courses and the curricular and assessment options for local districts.
- The practice would not benefit high school students who have not been held accountable for the first eight or nine years of their schooling. It would also be unfair unless a level playing field exists across districts in terms of the quality of education provided to students.

There are a number of concerns that MSBE should consider given the current assumptions about the multiple uses of the HSA. For example, Gearhart (1995) noted that single assessments, either norm-referenced multiple choice or performance based, do not well serve multiple, high-stakes needs. Often key proponents of large-scale assessments support multiple uses, but actually have very different priorities given these uses. Kirst and Mazzeo (1996) explain that when the California Learning Assessment System (CLAS) moved from a design concept to becoming an operational program it became clear that not all these proposed uses and priorities could be accommodated. When priorities of key stakeholders could not be accommodated, support for the program decreased. Lessons learned from California, other state, and national testing programs can be useful to MSBE as it moves to consider what appears to be the most ambitious plan for multiple high-stakes uses for a very innovative assessment system within a relatively short period of time. More discussion of these issues follows in the report and in Appendix A.

Does MSBE want the HSA to continue to be used as an absolute requirement for high school graduation of all students in spite of the challenges to implementation and the opposition from key constituency groups?

B. Differentiated Diplomas

During the public engagement meetings, there were repeated requests that the tests be used for awarding a "Regents" type diploma, not as a condition for graduation. At least two different models for such differentiated diplomas were suggested in public comments:

- A diploma that reflects the current graduation requirements and an additional diploma with an HSA endorsement certifying that the student had attained a satisfactory level of performance on the HSA assessments. This is advocated by those who want to be sure that students can receive something more meaningful than a Certificate of Attendance even if they cannot meet the HSA standards. The current New York State Regents Examination program is cited as the exemplar of this approach. Maintaining high standards would be facilitated by such a program because there would be no reason to dilute assessments to ensure that students pass.
- A standard diploma that reflects attainment of a minimally satisfactory level of performance on the HSA assessments as well as other graduation requirements. This would be augmented by one or two higher-level diploma(s) that reflect a higher standard of performance. That is, three diplomas might be offered: (1) the standard diploma for meeting graduation requirements along with satisfactory HSA performance; (2) a "meritorious" diploma that reflects higher grades and higher performance on HSA; and (3)

a "distinguished" diploma that reflects very high levels of accomplishment both in terms of grades and performance on the HSA. For example, some post-secondary institutions may use the different levels for placement decisions (e.g., exemption from remedial coursework; placement into a higher level course) or additional values may come to be associated with the higher level diplomas (e.g., work entry, state scholarships). It is argued that higher level diplomas would create more challenging expectations for stronger students.

The New York State Regents testing program has often been cited as a benchmark for the HSA. However, the *current* Regents program differs in use and design from the proposed HSA. In addition, for decisions made by post-secondary institutions to be directly tied to the HSA, the design must become more complex to give institutions the freedom and flexibility they need to meet their varying criteria for valid and informed admission and placement decisions. Another major distinction between the Regents program and the HSA is that the former provides more specification for curriculum and instruction, while Maryland seeks to permit local districts autonomy in fitting their curriculum to the assessments. See Appendix A (I, B) for additional discussion.

Does MSBE want the HSA design to provide for a differentiated diploma? If so, which model should be employed in designing the HSA assessments, and how can appropriate uses be ensured by Maryland's higher education institutions?

C. How High is High Enough?

There is no clear understanding of the intention of MSBE regarding the application of the test scores to graduation requirements. At least four possible uses have been identified. The choice depends on the relative value placed on expecting students to do well on all 10 examinations versus expecting them to do well overall. This latter position stems from the belief that individuals have different strengths and weaknesses, i.e., that all of us use our strongest skills to help us cope with those parts of school and life for which we don't have all of the skills we would like. The four models are:

- The Ten High Hurdle Race. A student must demonstrate a satisfactory level of competence on the ten tests, i.e., reach or exceed the passing score on each test. This model expects all students to demonstrate a satisfactory level of achievement in all parts of the Core Learning Goals. With this model, the lowest proportion of students will qualify for a diploma (See Appendix A, I C)
- The Low Hurdle Decathlon. A student must demonstrate a minimum level of competence on each of the ten tests and also achieve a satisfactory (higher) score on an overall composite such as the average of the ten scores. This presumes that all students will demonstrate some minimum level of achievement in all subjects but allows above-satisfactory scores on some tests to compensate for below-satisfactory scores on others.
- The Four Event Competition. A student must get a satisfactory composite score in each of the four subject areas. This model would allow an above-satisfactory score on one test *within* a subject to compensate for a below-satisfactory score on a test within the *same* subject area.

- The Decathlon. A student must achieve an overall satisfactory composite score across the ten tests. An above-satisfactory score on some tests would compensate for below-satisfactory scores on others, as long as the composite, e.g., average, of the scores was at the satisfactory level.

Each of the three compensatory models creates a substantial counseling and course scheduling problem for the student and the school: if a student receives a below-satisfactory score on a test, should he or she receive remediation and re-take the test or assume that he or she will receive a higher, compensating, score on another test? Nonetheless, during public engagement activities, many educators voiced support for some form of a compensatory model. As noted above, it is more likely that a student will receive a diploma with the compensatory models than it is with the multiple hurdle option described as the first possibility.

A related issue in setting high standards for graduation using any of these models is "How High is High Enough?" Given the level of variation of all aspects of student achievement reported from a wide range of national and international assessments, there are substantial challenges involved in setting a relatively high standard that all students are expected to meet within a relatively similar period of time. While some students may achieve many of the standards at roughly the same time during their 8th through 12th grade careers, others may achieve some of the standards at widely varying times. Yet using an end-of-course assessment presupposes that students will be equally prepared to achieve the standard at the same time.

Appendix A (I, C) provides illustrations of the impact of each of the four models on student passing. The "Ten High Hurdle" imposes a very difficult standard that would result in high failure rates because students must be successful on ten separate tests. The model selected by MSBE for the *decision rule* that will determine whether an individual student receives his or her diploma has substantial implications on the overall passing rate and difficulty of the HSA.

Which of these models best reflects MSBE's intention for the decision rule to be used for the HSA graduation requirement?

D. Implications of Ten Separate Tests.

The number of tests students must successfully complete is a significant issue in setting decision rules and determining the overall passing rate for students across the state and within each district. As noted above, requiring all students to pass 10 tests creates a substantial burden, but even compensatory models based on this number of tests (where students can combine scores across exams) will result in an assessment system that will be:

- overly complex and costly
- difficult to track and monitor student performance
- burdensome on local districts who must manage the administration and score reporting

Most high-stakes assessments require individuals to complete all assessments at one time. When scores are reported, the individual is aware of the outcome (pass or fail, or a specific score) and can make a decision about his or her future (e.g., the score is *good enough* and I will not retake the test; I will or must retake the test). Many of these tests do have separate subtests that are

scored independently, but they are administered in the same time period. As the number of separate tests required of students increases, the probability of failing one or more tests statistically increases as does the complexity and cost for managing the separate test administrations and the test score data and reports that must be issued following each administration. Because of the high consequences associated with each test, there are enormous pressures to maintain the quality control over all aspects of test development, administration, equating, scoring, validation, and reporting. The detailed, elaborate, and costly[1] procedures required to develop and maintain such an assessment system must be weighted against the advantages and disadvantages of requiring 10 separate tests.

Alternative designs that MSBE should consider (or reconsider) include:

- Developing end-of-program assessments in four areas to be completed at the end of appropriate coursework as designated by LEAs (this would reduce the 10 tests to 4 tests). Separate sections of each assessment could be devoted to specific content (or correspond to specific courses, e.g., Algebra, Geometry) areas.

- Introducing four end-of-course tests for 2001 (one per content area) and getting that system operational. After the four-assessment HSA is working satisfactorily, MSBE would approve additional assessments that would be phased in over time. The additional assessments may be required of all students for graduation, as currently proposed. Because all students would complete the same four assessments, school- and student-level accountability uses would be more appropriate. Another option would be to devote these additional assessments to advanced courses and offer these to students seeking a differentiate diploma or advanced certificate.

MSBE should consider the tremendous financial, operational, educational, and technical challenges and burdens that are associated with an assessment system comprised of ten separate end-of-course exams that are required of all students. The discussion in Appendices A and B attempts to raise the many complexities, risks, and concerns that must eventually be addressed if the current conceptualization of the HSA remains unchanged.

Does MSBE want to require ten separate end-of-course tests or consider alternative approaches for reducing the number of tests required for graduation?

E. Timing or Phase-in of Tests/Graduation Standards

A related issue concerns timing of the introduction of the requirement to pass 10 HSA assessments. Several issues were raised during public engagement about this particular issue. Many participants in public engagement felt that schools will not be equipped to adequately prepare the class of 2004 to master the Core Learning Goals. Several methods of overcoming this obstacle were proposed:

[1] Cost can be defined in terms of financial costs, instructional time, student burden (testing time), and local burden on schools which must maintain and manage an elaborate assessment system.

- Phase in the number of tests that a student must pass, that is, begin the process by requiring the Class of 2004 to pass a limited number, say 3, of the tests and then progressively increase the number required with each subsequent class.
- Phase in the passing standard, that is, make the initial "passing" score comparatively low, for example, a 2 on a 5-point scale, and then raise the level as schools and students become familiar with the demands of the HSA.
- Continue "no fault" administrations of the HSA beyond the planned pilot tests in 1999-2000, that is, delay the individual accountability provision of the HSA so that schools and teachers can become familiar with the demands of HSA and redesign the curriculum to better aid students in developing the knowledge and skills required by the Core Learning Goals.
- Employ multiple methods for making graduation decisions by combining performance on the HSA with other factors (e.g., course work, grades, GPA) to reduce complete reliance on one assessment system for high stakes decisions and to reduce potential misuses of the HSA.

Finally, Appendix A (I, E) provides a brief discussion of the advantages to extending the no-fault administration and other aspects of the timing of the HSA program that could minimize the negative impact on students. Section II E of the appendix discusses implications and risks of moving too quickly to implement a high-stakes performance assessment system.

Does MSBE want the HSA design to provide for a phase-in strategy of implementation?

II. THE DESIGN OPTIONS

A variety of alternative designs for the assessments were explored in discussions between the contractor and MSDE staff. Four options, ranging from one that is relatively unstandardized to one that is quite constrained by time and format, were developed and have been refined in subsequent discussions, including those with the content teams that developed the Core Learning Goals.

A. The Design Options

All of these options assume a maximum of three hours will be allotted for the administration of an assessment, although it may be possible to limit this to 120-150 minutes. The first three options all include a mixture of selected-response and constructed-response questions which will require scoring by people with knowledge of both the subject and the Core Learning Goals (CLGs). The distinguishing features of the four options are briefly described and the proposed components of each, stated in generic terms, are summarized in the chart below.

1. Portfolio Plus is based on the premise that a student's knowledge of and ability to do the things specified in the Core Learning Goals will be reflected in work done throughout the year. This option proposes that such information be collected in a portfolio that would be scored by the classroom teacher. The information derived from the portfolio would be combined with a student's score on a timed (2-3 hour) assessment given under standardized

conditions and scored centrally. This option provides for a wide sample of a student's work/performance to be used in evaluating achievement of the CLGs; it also allows for the classroom teacher's judgment to affect final pass/fail decisions.

2. Preparation Plus is based on the idea that students should share a common learning experience prior to assessment and that that learning experience can provide the substantive basis for a portion of the assessment. The common assignment or learning experience would model the kind of instructional strategies that are advocated for the particular subject. This common learning experience makes it possible for the timed (2-3 hour), standardized assessment to draw on more complex stimuli, e.g., a "messy" real-world mathematics application, than would otherwise be possible within the time limits of the assessment administration. The standardized assessment would be scored centrally.

3. Combination reflects a mature measurement technology that combines a number of constructed-response questions with a significant number of selected-response questions to create a reliable assessment instrument. Within the 2-3 hour time limit for administration of the assessment, it would include an extended constructed-response question that draws on a number of brief documents, laboratory results, or other authentic stimuli. The assessment would be scored centrally.

4. Limited Combination would be an entirely machine-scorable assessment administered in 2-3 hours in a standardized situation. Challenging selected-response questions would be the predominant type of question although in the sciences and in mathematics, a machine-scorable constructed-response format would also be used. The answer sheets would be scored by MSDE.

PORTFOLIO PLUS	Combines information from student's portfolio, scored over time by the teacher, with information from an "on-demand" assessment including: • Two Extended Constructed Responses • Several Brief Constructed Responses • Selected-Response section
PREPARATION PLUS	All students are assigned a common learning task prior to the "on demand" assessment. This preparatory work is not scored but provides a common learning experience as the basis for a more specific task on the assessment. The assessment would include: • One Extended Constructed Response based on preparatory task • One Extended Constructed Response on another topic • Several Brief Constructed Responses • Selected-Response section
COMBINATION	All work directly related to the assessment will occur in a timed, "on-demand" situation. The assessment will include a series of short texts, documents, laboratory reports, etc. which will provide the substantive basis for some of the questions. Students will be asked to respond to: • One Extended Constructed Response based on documents • One Extended Constructed Response on another topic • Several Brief Constructed Responses based on documents • Selected-Response section
LIMITED COMBINATION	The entire assessment would consist of Selected Response and, where available, Machine Scorable Constructed-Response questions that measure various content and application expectations.

Brief Constructed Response = (3-5 minutes)
Extended Constructed Response = (30+ minutes)

Costs The costs associated with each of the design options can be broken into several elements: start-up costs, annual costs per assessment title, and annual costs per assessment taken by a student. The estimated costs for the first three options are sufficiently similar that they are shown together as "Combination Models." The costs for the Limited Combination option are shown as "Machine Scorable." These are shown below in the Table 1.

A longer-term projection of costs for implementing and operating the High School Assessment program is provided in Table 2, "Ten Year Projection of Costs." This table is in two parts: the first half shows costs on the assumption that HSA will be used for individual student accountability and, therefore, that students will repeat assessments when they do not "pass" the first time; the second half shows costs on the assumption that HSA would be initially implements for only school and district accountability and, therefore, that students would not repeat assessments.

Table 1 - Estimated Start-Up & Unit Costs

OPTION	START-UP	ANNUAL UNIT COSTS		TOTAL ANNUAL COSTS*
		PER TITLE	PER ASSM'T	
MACHINE SCORABLE Limited Combination	$3,000,000	$182,000	$10	$11,934,000
COMBINATION MODELS Portfolio Plus Prep Plus Combination	$3,500,000	$242,000	$21	$23,379,000

* Assumes 12 assessment titles and 975,000 assessments administered - peak-year volume.

Table 2 - Ten Year Projection Of Costs (in thousands)

INDIVIDUAL STUDENT ACCOUNTABILITY

	'99	'00	'01	'02	'03	'04	'05	'06	'07	'08
MACHINE SCORABLE										
Start-Up & No-Fault	$ 1,293	$ 1,518	$ 1,018							
Operational			$ 2,678	$ 6,981	$ 10,959	$ 11,934	$ 11,934	$ 11,934	$ 11,319	$ 10,569
Total	$ 1,293	$ 1,518	$ 3,696	$ 6,981	$ 10,959	$ 11,934	$ 11,934	$ 11,934	$ 11,319	$ 10,569
COMBINATION MODELS										
Start-Up & No-Fault	$ 1,601	$ 2,076	$ 1,576							
Operational			$ 5,063	$ 13,539	$ 21,332	$ 23,379	$ 23,379	$ 23,379	$ 22,151	$ 20,511
Total	$ 1,601	$ 2,076	$ 6,639	$ 13,539	$ 21,332	$ 23,379	$ 23,379	$ 23,379	$ 22,151	$ 20,511

Assumes 50% repeat rate during the first five operational years and 20% thereafter

SCHOOL & DISTRICT ACCOUNTABILITY

	'99	'00	'01	'02	'03	'04	'05	'06	'07	'08
MACHINE SCORABLE										
Start-Up & No-Fault	$ 1,293	$ 1,518	$ 1,018							
Operational			$ 2,678	$ 6,491	$ 8,694	$ 8,694	$ 8,694	$ 8,694	$ 8,694	$ 8,694
Total	$ 1,293	$ 1,518	$ 3,696	$ 6,491	$ 8,694	$ 8,694	$ 8,694	$ 8,694	$ 8,694	$ 8,694
COMBINATION MODELS										
Start-Up & No-Fault	$ 1,601	$ 2,076	$ 1,576							
Operational			$ 5,063	$ 11,811	$ 16,554	$ 16,554	$ 16,554	$ 16,554	$ 16,554	$ 16,554
Total	$ 1,601	$ 2,076	$ 6,639	$ 11,811	$ 16,554	$ 16,554	$ 16,554	$ 16,554	$ 16,554	$ 16,554

Appendix A contains further elaboration of the designs and a discussion of scoring. Appendix A includes an extensive evaluation of each option and, for each subject area, provides examples of questions linked to particular Core Learning Goals, Expectations, and Indicators, broad cost implications, and criteria for evaluating the design options for each content area.

MSBE must select a design option to permit the Phase II effort to proceed on schedule. Postponing a decision would result in a delay in the work required for Phase II. A greater danger would be if the design preference changed after substantial work occurred in Phase II. We recommend that the following criteria be used for both selecting an appropriate design option(s) and making other critical decisions regarding the HSA (see Appendix A, II, F for additional definitions):

- academically rigorous
- professionally acceptable
- practically doable
- legally defensible
- economically affordable

Once a design option is selected, the contractors and the test specification teams must begin to determine the number and types of items required to measure the various Core Learning Goals. Each design option is distinct, having its own unique requirements, and work for one design will not generalize to a second design.

Which assessment design model does MSBE want used for the Phase II work?

B. Can Multiple Design Options Work?

A related issue is whether all assessments and/or content areas must conform to the same design model. That is, would it be permissible for two content areas to employ a Combination design with the remaining content areas using the Preparation Plus design? Would differences in design be permitted within the same content area -- e.g., English assessments 1 and 2 use Portfolio Plus and English assessment 3 would use the Combination design?

Educationally, it is appropriate to design the HSA to reflect the skills and abilities required of students (and that best reflect the Core Learning Goals) as well as the developmental level of students. Some educators believe that certain designs are more appropriate for students at different grades, ages, or levels of maturity. This would argue for multiple designs across content areas and/or tests within each area. Proponents would argue that imposing one design (or test format) on all content areas and assessments would be like fitting a square peg into a round whole.

However, employing the same design may increase student familiarity across all assessments and may create somewhat less anxiety for students than completing tests with varying designs. This issue may be of more concern if different design options were used in the same content area (e.g., English) than across content areas. Using the same design for all assessments may also help to build a general identity for the HSA more quickly than would occur if multiple designs (or test

formats) are used. While proponents of multiple design formats argue that flexibility is an important skill for today's learner and that students are already confronted with different evaluations in different classes, substantial variations across tests could be problematic for some students, educators and the public in attempting to get a sense of what the HSA looks like and what appropriate preparatory activities might be.

An approach used by several large-scale testing programs is to adopt one design, but to permit one common variation for one or more tests or content areas when appropriate. While a number of large-scale testing programs are comprised exclusively of multiple-choice items, the English or writing tests often include essays or other constructed response items. Applying a similar approach to the HSA would require choosing one design option for all tests, but permitting representatives from content areas to justify a second design option in their area. This approach assumes that no more than two design options are permitted, that these two options are fairly similar,[2] and that a second design option is permitted only if there is a compelling and educationally sound rationale established.

How much variation in design options will be permitted across content areas and within content areas? Will one design option be selected for all content area assessments, or will a second similar design option be considered for specific subject needs?

III. IMPLICATIONS OF END-OF-COURSE ASSESSMENTS

A. Turnaround Time for Scores

HSA was originally conceptualized as a series of assessments that would be administered in secondary schools at the end of courses, designated by each district, which correspond to each assessment. Four major and interrelated concerns about the timing of the assessments have emerged:

- What is the maximum turnaround time acceptable for reporting of test scores?
- When will the scores be reported to schools, students, and parents?
- At what point in the term will the tests be administered?
- Can the schedule ensure that the curriculum is more comprehensive than the Core Learning Goals?

These issues have been of concern to participants at every public engagement activity. School and district administrators, teachers, and parents have all reacted negatively to the time required to score existing state assessments, e.g., MSPAP (12 weeks), Functional Tests (8 weeks), and expressed a belief that the HSA should follow a different model so scores can be reported in significantly less time. Participants have emphasized that:

- Assessments must be administered as close to the end of the course as possible, and
- Results must be available to students and schools before the end of the term.

---

[2] For example, use of designs 1 and 2, 1 and 3, 2 and 3, or 3 and 4 would provide designs that are complementary. Use of designs 1 and 4 would be contradictory.

Unfortunately, these two conditions appear to be mutually exclusive and have not been successfully implemented in any large-volume, paper-based, high-stakes assessment program. Inclusion of performance assessments (associated with the first three designs) further extends the turnaround time and complexity for scoring the HSA[3].

It takes several weeks to score high-stakes tests. The design options that include one or more extended constructed-response items, several shorter constructed-response items, and a selected-response section, will require approximately 7-9 weeks for score turnaround, and probably more time if Maryland teachers are to serve as scorers. The fourth option, which is entirely machine scorable, would require approximately 4-5 weeks or longer. It should be noted that meeting these estimates for score turnaround will require substantially increased financial and human resources from the state and the schools, as well as the successful resolution of a number of very critical issues. These issues are addressed more specifically in Appendix A under "End of Course Assessments." Thus, there is a significant trade-off between minimizing score turnaround time and the possibility of including performance assessment components in each HSA instrument. And a more obvious trade-off between end-of-course administration, as proposed, and having scores available for counseling, placement, and remediation of students who fail the test.

Does MSBE want the HSA design to emphasize the importance of turnaround time for score reports at the expense of other design features?

B. Score Reporting and Scheduling Test Administration in the Schools

A related issue is the timing of reporting scores to schools, students, and parents, e.g., must the scores be returned before the end of the semester or course? In planning the schedule for an assessment program, one can work backward by first specifying when the test scores must be reported and using that information to determine the latest date for administration. The high-stakes impact on individual students is cited as a major reason that scores must be returned by the end of the term. However, there are other reasons:

- To enable schools to plan remediation that may be required during the summer or in the following semester for students who fail, as well as to inform course placement decisions.
- To provide scores to seniors before graduation.
- To provide scores to students and schools while they have meaning. Quick turnaround is viewed as essential for ensuring that the assessments are viewed as important levers of reform in the schools.

The time required for scoring and processing the assessments creates a strategic conflict for many constituency groups who both want scores before the end of the course but want the assessments administered as close to the end of the course as possible. It is clear that both of these demands cannot be met via a large-scale, high-stakes, paper-and-pencil testing program.

---

[3] Appendix A, section VII (c), contains a discussion of the implications of technology (e.g., computer-based testing) on this and other issues involving the HSA. We do not discuss this in the report because it does not appear feasible to administer the volume of tests proposed to students across all districts in Maryland via computer-based testing within the next few years. However, the proposed designs, and all future work conducted by MSDE, should not preclude the eventual transitioning of the HSA to a computer-based delivery and scoring platform.

However, some educators have argued that administering the assessments well before the end of the course would have a beneficial impact on the curriculum by creating a period of time when teachers must go beyond the Core Learning Goals. The perceived benefit of an earlier administration stems from a concern that the curriculum could become synonymous with the Core Learning Goals (CLGs) and not include the additional, but not necessarily common, topics and skills that were envisioned by the content teams that authored the CLGs.

Given the practical limits on turnaround time, there are two distinct alternatives[4]:

- Administer the HSA part way through the course and report the results just prior to the end of the course (relaxing the association between assessments and conclusion of the course).
- Administer the HSA at the end of the course and report the results several weeks after the end of the semester.

The first option provides the possibility of teachers using HSA results as part of a course grade as well as facilitating the placement of students into sequence courses or remediation. It may, as has been argued, encourage teachers and schools to implement a more comprehensive curriculum that extends the subject matter beyond the CLGs. Assessments might focus primarily on CLGs from a previous course (completed earlier) as opposed to the course students are currently enrolled in.

The second option provides the students with maximum time to master the CLGs covered in a particular assessment. This option would more easily accommodate the use of Maryland teachers to score the HSA.

These two alternatives are illustrated in Table 3.

[4] These are the only alternatives, if MSBE retains the requirement to have ten separate end-of-course tests. Earlier discussions about moving to an end-of-program assessment system or introducing four content tests (which might be administered midway through the next course level) would provide additional alternatives that could address some of the concerns about score reporting.

Table 3 - Illustrative schedules for administering and reporting HSA

	HSA Administered	*HSA Results Reported*
Option 1 - Scores Reported Before End of Course/Semester		
1A- Full-Year Program or Spring Semester Block		
Designs 1-3	4/5 - 4/15	June 7
Design 4	5/5 - 5/15	June 7
1B-Fall Semester Block		
Designs 1-3	11/1 - 11/10	Jan 10
Design 4	11/29 - 12/8	Jan 10
Option 2 - Scores Reported After End of Course/Semester		
1A- Full-Year Program or Spring Semester Block		
Designs 1-3	5/24 - 6/2	July 28
Design 4	6/5 - 6/15	July 13
1B-Fall Semester Block		
Designs 1-3	1/5 - 1/15	March 11
Design 4	1/12 - 1/22	Feb. 19

* In each model, there should be an opportunity for students who miss the test to retake it a few days later.

Which overall administration and reporting option does MSBE consider most viable and supportive of the school improvement goals that motivate the concept of HSA?

IV. FEASIBILITY OF ACCOMMODATING SPECIAL CIRCUMSTANCES

As with any large-scale testing program, there will be students who need to participate in HSA who do not fit into the usual administrative patterns developed for the vast majority of students. Policy guidance will be needed regarding the extent of effort that should be made in the HSA design to accommodate a variety of special circumstances. A number of policy and administrative issues are outlined in Appendix A, IV. These issues are not addressed here because decisions will not immediately impact the work required under Phase II. However, before additional elaboration can be provided on the costs, constraints, and operational and design implications of the range of accommodations and issues involved for several different groups of students, MSBE will need to address the policy questions delineated in Appendix A, IV.

There are two operational issues which must first be determined because they have implications on the number of test forms required, the administrative schedule for testing, and the types of items and designs that are most appropriate for the HSA. These issues involve students who are absent on the day of testing and the use of alternative evidence for demonstrating competency on the Core Learning Goals.

A. Student Absences and Make-Ups

Inevitably, when some 600,000 assessments are administered annually, there will be students who are absent. Some of these absences will be due to legitimate illness, accidents, or family

emergencies. Others will be strategic, i.e., students deliberately miss a day when they know they will be tested. While agreeing about providing a "make-up" opportunity for the absentees, educators and others have expressed concern about the number of opportunities students will have to take a High School Assessment in a particular subject if they fail to pass it initially. A number of individuals have argued that there needs to be a third test administration each year for students who have taken remedial courses during summer school.

Each additional administration introduces substantial additional costs and burdens on schools. The costs of additional administrations are largely determined by the following decisions:

- Does there need to be a new form of the assessment for the "make-up" and summer administrations to avoid the potential security risk of re-using the same form of the HSA?
- Do the papers from make-up or summer administrations need to be scored as quickly as possible, or can they be scored at the same time as those from one of the two main administrations, even though that means a delay of several months in students receiving their scores?

There are several options for the number of administrations to be offered:

- Administer assessments twice a year, once at the end of the fall semester for schools on a block schedule and once in the spring for full-year courses (and for spring semester courses). For each administration there would be a single make-up test date approximately one week later. All tests would be administered on the same one or two "make-up" dates across the state. These "make-ups" would use the same form and be scored along with the assessments from the major administration.
- Same as above, except that a third form would be developed for the make-up administrations in each year, but would be scored at the same time as the regular administration. This option reduces security concerns but increases developmental costs.
- Administer assessments three times a year, once at the end of the fall semester for schools on a block schedule and once in the spring for full-year courses (and spring semester courses), and once at the end of summer school. For each administration, there would be a single make-up test date approximately one week later. All tests would be administered on the same one or two dates. These "make-ups" would use the same form and be scored along with the assessments from the major administration. This option requires the development of three forms and three separate scoring sessions -- a substantial increase in costs of both development and scoring.

Other combinations, e.g., more "make-ups," can be imagined but all entail even more substantial cost implications.

Does MSBE want the HSA to provide for two annual administrations, each followed by one make-up administration?

V. FLEXIBILITY IN ACCOMMODATING VARIATIONS AMONG LOCAL DISTRICTS

The tradition in Maryland, as in most of the United States, of substantial local control of K-12 education makes the implementation of any centrally directed process difficult. HSA is no exception. There are persistent demands for choice and flexibility within the structure of the proposed HSA program.

The demands for flexibility relate to four central issues:

- Accommodating curricular patterns other than the twelve described by the Core Learning Goals reports.
- Accommodating half-credit courses.
- Accommodating accelerated courses.
- Accommodating districts and students by providing greater flexibility and choice through the use of modules.

Each of these issues raises a different set of considerations for the design of the High School Assessment program.

A. Alternative Curricular Patterns

Many education leaders and organizations both in mathematics and science have argued that the secondary curriculum in each of these areas needs to better integrate content and move away from the traditional disciplinary structure of the respective fields. A number of Maryland districts have responded to these suggestions by creating courses that do not correspond to the traditional organization reflected in the Core Learning Goals reports in these subjects. Instead, students are enrolled in courses that address the goals and expectations of the respective Core Learning Goals documents, but in sequences and combinations that differ from those in traditional courses in these fields.

A basic principle underlying the conceptualization of any assessment program is that the assessments should be aligned with the curriculum in order to demonstrate curricular validity and to reinforce and support the standards implicit in the Core Learning Goals.

Mathematics. According to a recent survey completed by MSDE staff in curriculum and instruction, approximately one-quarter of the districts offer or plan to offer integrated mathematics courses rather than or in addition to Algebra and Geometry courses. Although integrated courses cover all of the Algebra and Geometry Core Learning Goals in a two-year sequence, Algebra and Geometry are intertwined. At the end of each year of the integrated sequence, students would be prepared to be assessed on a portion of the Algebra goals and a portion of the Geometry goals, but neither of the planned assessments would be an appropriate end-of-course assessment.

While many of the questions written for the Algebra and Geometry assessments could be used in assessments entitled "Integrated Mathematics I" and "Integrated Mathematics II," some additional tasks may be needed. Test specifications would be required to guide the development of the

integrated mathematics assessments and to serve as the basis for descriptions of the assessments for students and teachers. These assessments would also need to be separately scaled and standards would have to be set for them. Thus, to accommodate students in integrated mathematics courses, the HSA would have to be expanded so that schools and students would have a choice from among:

Algebra
Geometry
OR
Integrated Mathematics I
Integrated Mathematics II

Providing this element of flexibility for districts would require a number of complex rules to ensure that all students do reach all Core Learning Goals. However, some number of students transferring within-state from districts offering these two different models (i.e., an Algebra/Geometry sequence vs. the Integrated Mathematics) would potentially not reach one or more of the CLGs (both in the curriculum and assessments) because of such differences across districts. Developing and operating four math assessments rather than the original two assessments would have financial implications. Each additional math tests would require approximately $250,000 annually even though the total number of tests administered would not increase. There may be some cost savings to the extent that the same items or tasks could be used in both assessment models.

Science. Based on a survey conducted by MSDE, more than one-third of districts offer one or more courses that integrate portions of two or more of the sciences identified in the Core Learning Goals. The most common are Environmental Science and a Chemistry/Physics course (sometimes called "Matter and Energy"). None of the four assessments currently planned would be an appropriate end-of-course assessment for these courses.

Selected questions written for the Chemistry and Physics assessments would provide some of the questions to be used in an assessment entitled "Integrated Chemistry/Physics." Although some of the questions written for the Chemistry and the Biology assessments would be suitable for an assessment entitled "Environmental Science," other questions will have to be created because biological and chemical principles are treated in a quite different context in environmental science than in a typical Biology or Chemistry course. For both of these integrated science assessments, test specifications would need to be developed to guide the development of the integrated assessments and to serve as the basis for descriptions of the assessments for students and teachers. These assessments would also need to be separately scaled and standards would have to be set for them. Thus, to accommodate students in these courses, the HSA would have to be expanded so that schools and students would have some degree of choice from among six science assessments[5]:

Biology

[5] Not all combinations would be acceptable under the CLGs. For example, a student could not complete the Chemistry assessment and later complete the Integrated Chemistry/Physics assessment. More complex decision rules would be required to ensure students complete non-overlapping assessments that do meet the intent of the requirements established by MSBE.

Chemistry
Earth/Space Science
Environmental Science
Integrated Chemistry/Physics
Physics

If MSBE wishes to accommodate districts employing or planning to introduce such integrated courses in mathematics and science, additional assessments are required. In mathematics, four separate assessments are required rather than the two already planned; in science, six separate assessments are required as opposed to the four originally proposed. Overall, HSA would have 16 tests (not 12), with students required to successfully complete ten tests. There will be substantial additional costs and time required to develop and maintain these four additional assessments. As noted above, each additional assessment will require approximately $250,000 annually to develop and maintain even though the total number of assessments administered and scored does not increase. The total costs of developing four additional assessments would result in an increase of approximately $1,000,000 for the HSA program each year. These additional assessments may also be phased in at a later date.

During Phase II additional test specification committees will be required. During the test development work significant additional expenses will be incurred with developing new items, piloting the tests, and scaling tests. Substantial additional expenses will also occur annually as MSDE must maintain and operate four additional tests.

Does MSBE want the HSA design to include specifications for integrated courses in mathematics and/or science (Integrated Mathematics I, Integrated Mathematics II, Integrated Chemistry/Physics, Environmental Science)?

B. Half Credit Courses

A few districts divide year-long courses in half, providing students with a half unit of credit for work completed in each semester, e.g., a school may give a half-unit of credit for U.S. History (1700-Civil War) and another half-unit for U.S. History (Civil War-Today). The students may be in different classes with different teachers and receive semester grades for each half of the course. Schools employing this model have asked if half of the U.S. History exam can be administered at the culmination of the first half course (January) with the second half administered at the end of the second half course (June).

This request raises severe logistical issues and psychometric concerns. Modules (or half tests) would not provide reliable scores for individual students. The combination of two modules, taken several months apart, into a total assessment score could not be interpreted in the same way as the total assessment taken at one time. The logistics of getting the appropriate half-tests to schools and students would add an expensive complication, while the need to create a system to match partial information from two separate half tests adds further expense. In addition, it is almost certain that schools that offer half-credit courses will not agree on how the content, skills, and processes from the entire course are distributed among their units. We believe all of these issues make administering and scoring separate modules for half-credit courses unfeasible.

Does MSBE want the HSA design effort to include the provision of modules to meet half-credit courses ?

C. Accelerated Courses

Exemption from assessments for students successfully completing Advanced Placement (AP) or International Baccalaureate (IB) exams has been suggested by many participants during various public engagement activities. It has been proposed that a waiver from the corresponding HSA be granted for students who successfully pass an AP or IB examination that MSDE has determined to adequately represent the subject area of an HSA (e.g., AP U.S. History as a substitute for the HSA U.S. History assessment)[6].

Nationally, AP and IB curriculum and assessments are perceived as among the most rigorous offered in secondary schools. Given that more than 2,000 universities and colleges consider successful performance on these exams as equivalent to freshmen-year work, a passing grade should meet or exceed the state's standards even if the course does not cover all elements of the Core Learning Goals. In Maryland, 71 percent of high schools participate in AP and more than 13,000 students annually take in excess of 21,000 AP examinations. Per 100 11th and 12th graders in Maryland, 18.8 AP exams are taken, a figure that is substantially higher than the national average of 13 exams per 100 11th and 12th graders. Thirty-three colleges and universities in Maryland grant sophomore standing for successful performance on AP exams[7].

Participants in public engagement activities were concerned that expectations for the most able students would be depressed if there were no incentive to complete AP or IB courses and examinations. They argued that requiring students to take the HSA in all subjects without exception would discourage college-bound students from taking AP or IB courses because students would be required to take the distric course (where CLGs are covered) in order to pass the Maryland course assessment. For example, if waivers are not integrated into the system, requiring students to complete the CLGs and state assessments in Chemistry or U.S. History (which may not all be covered in the AP courses) would likely discourage students from enrolling in the AP Chemistry and AP U.S. History courses. Alternatives have been suggested that would accommodate both the state's desire for accountability and the general objectives of encouraging students to take honors courses. These include:

- Award students who pass the AP and IB exams the maximum HSA scores when computing school accountability data so schools are not penalized when a large percentage of their student are exempted from the HSA by a waiver for AP or IB. Such a policy would avoid the incongruity of a school with many students in such honors

[6] Currently, because of the turnaround time for scoring, AP and IB results would not be available prior to the end of the course.

[7] If AP or IB examinations are used, successful performance would not ensure that students have mastered all the Core Learning Goals, but rather assume that these students would do well on the HSA because they have performed successfully on examinations that are very likely to be more rigorous than the HSA. Research could be conducted to determine the relationship between the tests and student performance levels once the HSA are in place.

programs being ranked near the bottom of all schools because their best students are not in the HSA program.

- Allow students to meet an HSA requirement via AP or IB, but compute school accountability data separately from the HSA data for this group of students.
- Provide incentives for students to take AP or IB exams by paying exam fees. The revenue to do this would come from savings realized from lower numbers of students taking the HSAs.
- Determine whether there are additional mechanisms that would allow local districts to petition for exemptions for other honors courses that require students to demonstrate proficiency on standardized assessments (locally or nationally developed) that meet or exceed HSA standards and are psychometrically acceptable. Districts may then develop or adopt standardized tests for use in evaluating students in local honors courses.

Does MSBE want the HSA design to provide for a waiver for students who demonstrate a satisfactory performance on either Advanced Placement or International Baccalaureate exams that correspond to specified HSA assessments?

D. Modules

During public engagement activities, representatives from several districts advocated the use of modules for a number of distinct purposes. We briefly describe each desired use of modules and the feasibility of the approach given the current proposed high-stakes uses of the HSA. The initial two proposed uses have been described above in this section:

- To respond to districts employing integrated mathematics and science courses. However, as noted above (see section V, A above) we believe that separate and distinct tests would be required. While some item overlap is likely to reduce costs, modules will be not adequate for this purpose and additional tests (with their unique specifications) will be required.
- To facilitate testing students enrolled in half credit courses. In this instance we do not believe modules are feasible (see section V, B above).

However, several additional uses of modules have also been proposed to permit:

- Districts to more generally select a subset of modules from a larger pool of modules to ensure the assessments most closely reflect the local curriculum. Proponents argue that modules would permit schools to select and use assessments which are best aligned with their curriculum. If the assessments do not reflect the curriculum offered in the school, they will lack curricular validity, and it will be difficult for schools to demonstrate that students had an opportunity to learn the requisite Core Learning Goals.
- Districts to combine modules from the state assessments with modules from their local assessments to meet the state graduation requirement.
- Students with disabilities to complete tasks proximate to instruction. Modules would provide an additional accommodation to students with disabilities who will have more difficulty performing on-demand weeks or months after instruction.

Permitting districts to select among several modules, as proposed in the first two scenarios, is most problematic because not all students would complete the same tasks, and there would be no way to compare individual student level performance with adequate precision to justify high-stakes uses for students. That is, if students in different districts are completing different modules, there is no way to ensure comparability (e.g., in terms of content coverage, difficulty of tasks, familiarity). Variations among students' performance across districts may be as attributable to the modules selected as it is to students' competency on the CLGs.

The final proposed use of modules for students with disabilities also raises an issue that complicates attempts to compare scores of students. The combination of student performances from several modules, taken several months apart, into a total assessment score could not be interpreted in the same way as the total assessment taken at one time.

There are additional difficulties raised with modules and other aspects of choice in a large-scale high-stakes assessment program such as the HSA. Each variation within a program, such as modules:

- introduces non-standardized conditions which make it difficult to make valid and fair comparisons of student and school performance. When different tasks or forms of assessments are completed, and/or assessments are administered at different times to different populations of students, the complexities of equating and scaling assessments become more difficult and may make it impossible to meet psychometric and technical standards needed.
- creates substantial logistical burdens for schools which must determine which variations of assessments are to be completed by which students and for the state (or its contractor) who must produce multiple iterations of each assessment and coordinate printing, shipping, and scoring across districts (the probability of errors which impact the quality of the results increases dramatically).
- significantly increases costs for development, implementation, psychometric studies, scaling, and scoring.
- complicates score reporting at the district and state level -- producing district or school reports that cannot be compared to each other.
- hinders students and teachers from gaining familiarity and comfort with the assessments and increases the difficulty of developing a common understanding of HSA by parents and the public.
- creates an impression of unfairness among stakeholders (even if variations can be linked in some way). This impression is exacerbated by the difficulties of empirically demonstrating the equivalence of the variant tests. The general public may not understand how you can compare performance of students who take different forms of tests (e.g., modules) at different times when tasks differ substantially.

These issues are discussed further in Appendix A, V. For these and other reasons we believe that modules are generally not feasible given the proposed high-stakes uses of the HSA with individual students.

VI. SUPPORTING STUDENTS WHO DO NOT DEMONSTRATE COMPETENCE ON THE HSA

One of the most complex and controversial issues involving the HSA is determining what additional opportunities will be provided to students who do not pass[8] one or more of the assessments. MSBE must first consider whether alternative forms of evidence or alternative assessment practices will be permitted for students who do not initially succeed on the HSA.

A. Alternative Evidence of Competency

A continuing issue is whether there should or can be alternative demonstrations of competence for students who do not pass one or more of the HSAs. Opinion during public engagement activities was sharply divided between those who believe that there should be local alternatives to passing the HSAs and those who argue that any local procedure is susceptible to manipulation and that the credibility of the HSA program will be undercut by permitting local alternatives to be used. The latter argue that such alternatives will be perceived as unfair.

Consultation with psychometric experts has indicated that there is no practical means of empirically demonstrating that alternative activities will provide "*equivalent evidence of competence*." That is, you can not align several different assessments to the Maryland state assessment and speak with any level of confidence about the equivalence of student performance across these different assessments (i.e., you can not statistically determine that students with a specific score on the local options would perform comparably to students with the same score on the state assessment in such an operational testing program). Equivalence is required by professional standards and would appear to be essential both from a legal and ethical perspective if assessments are to be used for high-stakes decisions as currently proposed.

If Maryland strongly desires to approve local alternatives for demonstrating competency on the Core Learning Goals, one possibility is to accept a lower standard for student comparability between these assessments. The state or local districts could develop one or more alternative options (or certify those options developed by LEAs) that can be used to demonstrate that standards have been met (by all students or just those who initially fail the assessment) as acceptable (but not equivalent) documentation of competence on the CLGs. This would give students alternate routes for demonstrating competence, yet it would not be possible to compare students because of basic differences between the alternatives. This model would required MSBE to determine the standards for these equivalence of alternative forms of evidence. Because this would be a judgmental process and performance on alternatives would not be statistically comparable, this model would not be legally or professionally defensible if results were used for high-stakes decisions. Before MSBE embarks on the use of any alternative options or choice among high-stakes assessments, we strongly recommend that a legal review be commissioned to evaluate the potential uses and risks.

---

[8] While MSDE documents do not specifically talk about "passing or failing" the tests, students who must reach a specified score or proficiency level (e.g., 3 of 5) to demonstrate competency in a content area in order to avoid retaking a test are, in fact, required to "pass" the test. The proficiency score is used as a cut score, and students who do not reach that score are required to retake the test or, possibly, demonstrate competence in the area through another means.

Is MSBE willing to eliminate local options for consideration at this time? If not, what additional work must MSBE commission in the next few months to ensure professional standards are maintained and that legal risks are minimized?

B. Additional Issues

The general topic of how to support students who do not demonstrate competence on the assessments engendered a variety of interrelated issues during public engagement activities:

- Will students have to retake an assessment a minimum number of times?
- If options other than HSA are available, how will student performance on these alternatives be judged in view of the fact that the results will not be strictly equivalent with the proficiencies measured by HSA?
- Will the state develop additional options for these students or support LEAs as they seek to develop their own options?
- What types of remediation will be available to students and what state and local support will be required?
- How soon following the initial testing will students be permitted to retake a test?

While Phase II of the design work can proceed without any final decisions regarding these issues, we believe that MSBE must resolve these issues by the spring or early summer of 1997. There is very substantial concern among the various key stakeholder groups about these issues. Failure to resolve these issues, or at least to provide firmer statements defining the options for students who fail a test, may result in opposition to the entire program and may not provide MSDE and local districts the time to build support (e.g., fiscal, staff, logistics) and plan for the extensive remediation that is likely to be required. Few people assume that students who initially fail an end-of-course assessment will subsequently achieve the same proficiency level without intervention. Consequently, the education system in Maryland must decide on the types of remediation that will be available to students who initially fail the exams and the types of resources and support required from the state and local systems. These issues should not be deferred much longer.

Section VI of Appendix A provides a discussion of the major issues that MSBE must confront within the next six months concerning the support of students who do not demonstrate competency on the assessments. We recommend MSBE review these issues as soon as possible in order to reach some general resolutions in the next few months.