

13

**MARYLAND HIGH SCHOOL ASSESSMENT  
SUMMARY OF DISCUSSION ON ITEM/TASK BANK PROPOSALS  
WITHIN AN ACCOUNTABILITY TESTING PROGRAM**

**CHIEF STATE SCHOOL OFFICERS' CONFERENCE  
ON LARGE SCALE ASSESSMENT  
PHOENIX, ARIZONA  
JUNE 19, 1995**

***Present:*** Wendy Yen (NCME and CTB), Nadir Atash (WESTAT), David Bayless (BAYLESS AND ASSOCIATES), Philip Daro (NEW STANDARDS), Steven Ferrara (MSDE), Hannah Kruglanski (EDUNETICS), Leslie Walker (HOWARD COUNTY), Bob Gabrys (MSDE and CONVENER)

***Discussion:*** Discussion focused on a reading of the Assistant Superintendents for Instruction and Charles County proposals related to the feasibility of an item/task bank with test construction by local school systems occurring under state guidelines.

***Conclusion:*** While it is conceivable such a system could exist, the amount of resources in dollars, personnel, expertise, and time for twenty-four school systems to implement such a system makes the proposals non-feasible. Additionally, comparability of scores and the potential for litigation in a high-stakes graduation requirement assessment system raise significant psychometric concerns regarding validity and reliability.

The group also had difficulty understanding why state instruments would be viewed as undesirable given the accountability purpose of the program, since local school districts would be responsible for the legal issues of test validity. The Core Learning Goals are what would drive the assessment, and test specifications could be written defining percentages of selected and constructed responses.

The group did suggest, however, that the state might investigate modularizing the testing program. Such a structure would provide more flexibility to local school systems in the delivery of instruction to different levels of students based on their needs. They felt this option should be explored for its technical merits, but cautioned that even this option may have psychometric barriers in the context of a testing program being used as an individual student graduation requirement assessment.



**MEMORANDUM**

**TO:** Steven Ferrara/Jessie Pollack

**FROM:** Bob Gabrys *Bob*  
Assistant State Superintendent for Research and Development

**DATE:** June 7, 1995

**SUBJECT:** Request for Assessment Opinion on High School Assessment Model

---

Steve/Jessie:

As you know, the local Assistant Superintendents for Instruction and some of the Superintendents recently endorsed the concept of the high school assessment being developed through an item/task bank concept rather than the use of a set of state test instruments.

At the same time, at the May State Board meeting, the Board members were quite clear in both the discussion of the Howard County approach to assessment and my High School Assessment Task Force update that the purpose of accountability overrides that of school improvement when it comes to the high school assessment. It is also assumed that accountability will strongly rest with the individual student.

I have received consistent information that the psychometric issues related to operating an item bank assessment system with local school systems determining (within state guidelines) the time of testing and whether it will be end-of-course or end-of-course-of-study might be overwhelming and, even if possible, quite a drain on resources. To attempt to resolve this issue with the Task Force, I will be convening a conference call and meeting with measurement people at the large-scale conference in order to discuss the feasibility of such an effort. You, of course, will also be invited to be part of that group.

Apart from the general discussion, however, I request that you review the two enclosed proposals (the Assistants and Charles County) we have received that are most germane to the issue of the bank concept. Would you then provide me with your professional views on the feasibility of conducting such an approach to the assessment. I ask that you consider the issue of feasibility at this time rather than all of the steps and issues that would be addressed in a more comprehensive review since it is important that the High School Assessment Task Force have the benefit of the

best thinking of possible approaches as it prepares to report to the State Board in July. Actually, I'm sorry that I didn't know the Psychometric Council was in town this week, because they might have been able to assist us firsthand. The bottom line is that I would not want the Task Force to recommend to the State Board a course of action that is not feasible in the context of a large-scale accountability program.

I want to thank you on behalf of the Task Force for any help you can offer on this issue. I think we will be reaching closure at the June 22 meeting. Presentations on the two designs will be part of the June 9 Task Force meeting, so you will hear additional information at that time.

Thanks again.

REG/bef

c: Nancy Grasmick  
Joan Palmer  
Mark Moody  
Ted Schuder

psychom.hsa



**MEMORANDUM**

**TO:** Bob Gabrys  
**FROM:** Jessie Pollack *Jessie*  
**DATE:** July 1, 1995  
**SUBJECT:** Psychometric Evaluation of HSA Item Bank Proposals

---

In response to your written request for a psychometric evaluation of the proposals submitted by the Assistant Superintendents of Instruction and Charles County Public Schools, I offer the following comments.

From a psychometric perspective, the proposal from the Assistant Superintendents of Instruction is not feasible. By contrast, the proposal submitted by Charles County Public Schools is possible, but not practical within the proposed development timeline and with the staff and financial resources that would be required. My concerns are related to test reliability and validity, test development including standardized administration and scoring, test security, quality control, and data management.

1. **Validity.** The principal reason to reject the proposal from the Assistant Superintendents of Instruction is because the scores that will result from end-of-course or end-of-discipline administrations will not be comparable. I have assumed that "... the option to determine the time of test administration" may mean testing at end-of-course or end-of-discipline, or possibly "on demand" within a specific course. In addition, the present draft of the Maryland Core Learning Goals was written by the Content Teams assuming the State-mandated assessments would be end-of-course assessments. If end-of-discipline assessments were to be implemented, these Core Learning Goals would have to be revised. Scores from end-of-course and end-of-discipline measures are not comparable, because the depth and breadth of content coverage is not equivalent. Hence, the content validity of these two types of assessments would be substantively different. The scores could be equated, but should not be equated unless the assumption is tenable that the trait or construct measured by the various instruments is the same. In this case, that assumption is NOT true. Tests developed to measure the domain of an end-of-course assessment would be more detailed; items would be sampled from a narrower content domain. Tests developed to measure the domain of an end-of-discipline assessment would be more general; items would be sampled from a broader content domain.

The following concerns are applicable to both proposals.

2. **Test Development.** Even if end-of-course tests were mandated, the item bank size required in order to have the number of items minimally necessary to develop 24 different but reliable and valid assessments would be vast. Reliability of high-stakes tests used for decisions about individual students should be higher than the reliability levels of measures such as MSPAP which are used for decisions about school improvement. And reliability is a function of test length. Reliability is also a necessary but not sufficient condition for validity. It is also important to note that item bank development includes concurrent scoring tool development and test administration manual development. Therefore, the test development demands, even with the aid of a test development contractor, would place an unacceptable burden on resources (MSDE assessment and content specialists, funds for field testing, scoring, etc.). MSDE staff are stressed presently producing an annual edition of MSPAP 3, 5, and 8 which contains typically less than 30 tasks per edition and grade. Additionally, if performance tasks were a desired format for at least a portion of the tests, the test specifications to ensure validity and generalizability of the results would need to sample from both content domains and context domains, further increasing the number of tasks required in the item banks.
3. **Test Documentation/Manuals.** Unambiguous secure test and item/task specifications would have to be developed. Specific guidelines to ensure standardization of test administrations, scoring tools, and procedures would have to be developed. These could be seen by only those in the local school systems with a need to know, adding yet another layer to the test security and quality control monitoring requirements.
4. **Test Security.** High-stakes testing mandates test security. Item banks of performance assessment tasks that satisfy Maryland's criteria would have easily remembered themes, another reason to have either a larger task pool or less flexibility in administration. This flexibility desired by the locals would then have to be compromised.
5. **State Quality Control Monitoring.** The best efforts at producing test specifications and procedures for implementation will be wasted if local educational systems differentially implement the regulations or guidelines. Therefore, the workload associated with monitoring State procedures for local test development, accuracy of pre- and post-equating, and test security would create another layer of oversight beyond the rescoring of a sample of papers to document scoring reliability.

Therefore, because the State would have to place so many restrictions on the development, administration, and scoring of the district-developed measure, the resulting locally-developed instruments would not warrant the expense or actually provide the flexibility desired.

JP/bef

Propeval7/95

## MEMORANDUM

TO: Bob Gabrys

FROM: Steve Ferrara *Steve*

DATE: July 14, 1995

RE: Your Request for Evaluation of the Practical and Psychometric Feasibility of Two Proposals for the High School Assessment (HSA)

C: M. Moody, J. Pollack, F. Albert

---

I am writing in response to your request to provide my professional views on the proposals from the Charles County Public Schools and the local Assistant Superintendents for Instruction for item/task banks for use by local school systems for the HSA. These proposals seek maximum flexibility in the construction of HSA tests and in when the tests can be administered so that school systems can organize curriculum, instruction, courses, and high school schedules without restraints caused by state testing schedules. I recognize the importance of providing for (a) this flexibility, and (b) psychometrically sound HSA tests that can be used in a high school graduation testing program.

Although I found these two proposals helpful to us -- because they forced us at MSDE to think hard about possibilities, not just needed constraints -- I consider these two proposals practically and technically unfeasible. However, the proposals led us to the notion of HSA "modules" -- which you and I discussed in preparation for the June meeting in Phoenix. And the modules idea actually reflects what would have been likely to result, I believe, from implementing the Charles County proposal "using the blueprint for [flexible local school system test] construction provided by MSDE" (p. 2 of the June 9 plan from Charles County).

Jessie Pollack has detailed practical and technical problems that make the two proposals unfeasible. I agree in general with her analysis and with her conclusions about both proposals. I consider the two proposals unfeasible primarily because:

- Some local school systems are not likely in the near future to have the necessary staff time and technical and print production capabilities to produce HSA test materials that would meet MSDE quality requirements.
- The flexibility in test construction that both proposals seek would make it unworkable for the various HSA test forms around the state to produce equivalent scores and standards for success. Equating test scores under more ideal conditions is difficult enough as it is.

As you know, the measurement experts we discussed these proposals with in Phoenix were in agreement about the unfeasibility of the two proposals and about the promise of the modules approach. These experts included Phil Daro of the New Standards Project, Leslie Walker-

Bartnick of the Howard County Public Schools, David Bayless and Nadir Atash of Westat, and Wendy Yen of CTB McGraw-Hill.

Test modules can be thought of as akin to subtests that cover medium size chunks of related material, like the subtests in the SAT, ACT, Functional Tests, CTBS/4, and so forth. The sizes of modules that we might produce in each content area for the HSA could be smaller or larger than in these examples; that would have to be determined in the process of designing the HSA. Test modules would not provide the ultimate flexibility and control the two proposals seek, but they could provide for a great deal of flexibility. This flexibility should allow for local school systems to organize courses and high schools as they deem appropriate within some limits in order to meet state HSA module administration schedules. This flexibility should also allow MSDE to produce equated and technically sound alternate forms of modules for administration two or more times per school year.

I am eager to develop further the notion of modules for the HSA by working with other MSDE and local staff in curriculum and assessment. I hope this memo responds adequately to your request.

st2921