

Consistency in the Decision-Making of HSA and MSA:
Identifying Students for Remediation for HSA¹

Robert W. Lissitz

And

Hua Wei

MARCES in EDMS

University of Maryland

The Maryland High School Assessments (HSA) have been developed and administered as end-of-course exams to assess student's knowledge of the Core Learning Goals in English, Government, Algebra/Data Analysis, and Biology. Starting from fall 2001, students who take courses in those content areas have been required to take the tests, but in a relatively low-stakes environment. Beginning with the graduating class of 2009 (students entering grade 9 in fall 2005), students are required to earn a satisfactory score on the HSA in order to obtain a Maryland High School Diploma.

The Maryland School Assessment (MSA) is an assessment program developed to comply with the requirements of the federal No Child Left Behind (NCLB) Act. It is a measure of students' achievement in reading and mathematics (and science, starting in 2008), as specified in the Maryland Content Standards. The test is administered annually to students in grades 3 through 8 as well as to those students who have taken 10th grade English and Geometry (up through 2005) or Algebra I (beginning in 2006).

Like the high-school exit exams in other states, the HSA has the goal of improving student achievement in high school and ensuring that all students who graduate from high schools demonstrate skills that are essential for life after high school. The MSA, like other state-wide assessment programs, is intended to provide educators, parents, and the public valuable information about student, school, school system, and state performance. Both tests contain a mixture of multiple-choice questions and questions requiring written responses. As indicated by the Maryland State Department of Education, "The content and format of the MSA will help prepare students for success on the HSA. Together, the two testing programs provide schools,

teachers, and parents a good picture of student performance, and help identify students' strengths and weaknesses” (High School Assessment Fact Sheet, retrieved March 22, 2006 from http://www.mdk12.org/mspp/high_school/what_is/factsheet.html).

Once passing the HSA becomes a graduation requirement, the pass-fail decisions regarding students' performance on the HSA have tremendous consequences to individual students. Evidence needs to be collected to prove that the standards by which students are categorized as having passed or failed the exams are valid. One type of validity check is to compare the results of the HSA to those of the MSA, since the two tests overlap a lot in terms of the content being assessed. This report summarizes the comparison in a succinct way and the degree of consistency in the decision-making of HSA and MSA has strong implications for the standard-setting practice in the HSA.

DATA and ANALYSIS The data examined in this report consist of the HSA and MSA scores of 14,972 students in four counties in Maryland. The HSA scores were obtained from the 2004 administration in the subject area of English and the MSA scores came from the 2003 administration in reading. The English test in the HSA and the reading test in the MSA are closely related. Thus it is plausible to compare the decisions made on one test with those on the other. The students were categorized into three categories (basic, proficient, and advanced) based on their MSA scores, and they were categorized as passing or failing based on their HSA scores. The following tables display the distribution of students in terms of their “joint” status in the two assessments in each of the four counties from which the data were obtained. A table summarizing all four counties is also presented. Note that the percents are calculated as a percent of all the students so that summing to 100% is over the whole table (all 4 or 6 cells). This is quite different from calculating the percents so that they sum to 100% across a single row. The reader can use the data we have provided to calculate any of these tables for themselves.

RESULTS

Charles County

		HSA (English)		Total
		0 (fail)	1 (pass)	
MSA (Reading)	0 (below proficient)	110 (10.3%)	5 (0.5%)	115 (10.8%)
	1 (proficient)	202 (18.9%)	65 (6.1%)	267 (25.0%)
	2 (advanced)	131 (12.3%)	553 (51.9%)	684 (64.2%)
Total		443 (41.6%)	623 (58.4%)	1066 (100.0%)

Harford County

		HSA (English)		Total
		0 (fail)	1 (pass)	
MSA (Reading)	0 (below proficient)	213 (8.8%)	14 (0.6%)	227 (9.4%)
	1 (proficient)	376 (15.5%)	144 (5.9%)	520 (21.5%)
	2 (advanced)	282 (11.6%)	1395 (57.5%)	1677 (69.2%)
Total		871 (35.9%)	1553 (64.1%)	2424 (100.0%)

Howard County

		HSA (English)		Total
		0 (fail)	1 (pass)	
MSA (Reading)	0 (below proficient)	180 (5.5%)	19 (0.6%)	199 (6.1%)
	1 (proficient)	293 (9.0%)	135 (4.1%)	428 (13.1%)
	2 (advanced)	274 (8.4%)	2355 (72.3%)	2629 (80.7%)
Total		747 (22.9%)	2509 (77.1%)	3256 (100.0%)

Prince George's County

		HSA (English)		Total
		0 (fail)	1 (pass)	
MSA (Reading)	0 (below proficient)	2119 (25.8%)	112 (1.4%)	2231 (27.1%)
	1 (proficient)	1924 (23.4%)	741 (9.0%)	2665 (32.4%)
	2 (advanced)	751 (9.1%)	2579 (31.4%)	3330 (40.5%)
Total		4794 (58.3%)	3432 (41.7%)	8226 (100.0%)

The four counties as a whole

		HSA (English)		Total
		0 (fail)	1 (pass)	
MSA (Reading)	0 (below proficient)	2622 (17.5%)	150 (1.0%)	2772 (18.5%)
	1 (proficient)	2795 (18.7%)	1085 (7.2%)	3880 (25.9%)
	2 (advanced)	1438 (9.6%)	6882 (46.0%)	8320 (55.6%)
Total		6855 (45.8%)	8117 (54.2%)	14972 (100.0%)

As mentioned above, two achievement categories result from the HSA and three from the MSA. Before we complete our examination of the degree of consistency between the decision-making in the HSA and that of the MSA, we can ask a simple question regarding the MSA scale: How should we best dichotomize the MSA scale so that it functions like a pass/fail scale? There are two natural ways of achieving this goal, although a cut point can certainly be placed in other locations along the MSA scale. The cut score for “proficient” can be used to divide the students as below proficient or at or above proficient. Or the cut score for “advanced” can be used to categorize the students as below advanced or advanced. The following tables summarize the results of decision consistency for each county as well as for the four counties aggregated as a whole in both these ways.

Charles County

		HSA (English)		Total
		0 (fail)	1 (pass)	
MSA (Reading)	0 (below proficient)	110 (10.3%)	5 (.5%)	115 (10.8%)
	1 (at or above proficient)	333 (31.2%)	618 (58.0%)	951 (89.2%)
Total		443 (41.6%)	623 (58.4%)	1066 (100.0%)

Note: Percentage of decision consistency: **68.3%**

		HSA (English)		Total
		0 (fail)	1 (pass)	
MSA (Reading)	0 (below advanced)	312 (29.3%)	70 (6.6%)	382 (35.8%)
	1 (advanced)	131 (12.3%)	553 (51.9%)	684 (64.2%)
Total		443 (41.6%)	623 (58.4%)	1066 (100.0%)

Note: Percentage of decision consistency: **81.1%**

Harford County

		HSA (English)		Total
		0 (fail)	1 (pass)	
MSA (Reading)	0 (below proficient)	213 (8.8%)	14 (.6%)	227 (9.4%)
	1 (at or above proficient)	658 (27.1%)	1539 (63.5%)	2197 (90.6%)
Total		871 (35.9%)	1553 (64.1%)	2424(100.0%)

Note: Percentage of decision consistency: **72.3%**

		HSA (English)		Total
		0 (fail)	1 (pass)	
MSA (Reading)	0 (below advanced)	589 (24.3%)	158 (6.5%)	747 (30.8%)
	1 (advanced)	282 (11.6%)	1395 (57.5%)	1677 (69.2%)
Total		871 (35.9%)	1553 (64.1%)	2424(100.0%)

Note: Percentage of decision consistency: **81.8%**

Howard County

		HSA (English)		Total
		0 (fail)	1 (pass)	
MSA (Reading)	0 (below proficient)	180 (5.5%)	19 (.6%)	199 (6.1%)
	1 (at or above proficient)	567 (17.4%)	2490 (76.5%)	3057 (93.9%)
Total		747 (22.9%)	2509 (77.1%)	3256 (100.0%)

Note: Percentage of decision consistency: **82.0%**

		HSA (English)		Total
		0 (fail)	1 (pass)	
MSA (Reading)	0 (below advanced)	473 (14.5%)	154 (4.7%)	627 (19.3%)
	1 (advanced)	274 (8.4%)	2355 (72.3%)	2629 (80.7%)
Total		747 (22.9%)	2509 (77.1%)	3256 (100.0%)

Note: Percentage of decision consistency: **86.9%**

Prince George's County

		HSA (English)		Total
		0 (fail)	1 (pass)	
MSA (Reading)	0 (below proficient)	2119 (25.8%)	112 (1.4%)	2231 (27.1%)
	1 (at or above proficient)	2675 (32.5%)	3320 (40.4%)	5995 (72.9%)
Total		4794 (58.3%)	3432 (41.7%)	8226 (100.0%)

Note: Percentage of decision consistency: **66.1%**

		HSA (English)		Total
		0 (fail)	1 (pass)	
MSA (Reading)	0 (below advanced)	4043 (49.1%)	853 (10.4%)	4896 (59.5%)
	1 (advanced)	751 (9.1%)	2579 (31.4%)	3330 (40.5%)
Total		4794 (58.3%)	3432 (41.7%)	8226 (100.0%)

Note: Percentage of decision consistency: **80.5%**

The four counties as a whole

		HSA (English)		Total
		0 (fail)	1 (pass)	
MSA (Reading)	0 (below proficient)	2622 (17.5%)	150 (1.0%)	2772 (18.5%)
	1 (at or above proficient)	4233 (28.3%)	7967 (53.2%)	12200 (81.5%)
Total		6855 (45.8%)	8117 (54.2%)	14972 (100.0%)

Note: Percentage of decision consistency: **70.7%**

		HSA (English)		Total
		0 (fail)	1 (pass)	
MSA (Reading)	0 (below advanced)	5417 (36.2%)	1235 (8.2%)	6652 (44.4%)
	1 (advanced)	1438 (9.6%)	6882 (46.0%)	8320 (55.6%)
Total		6855 (45.8%)	8117 (54.2%)	14972 (100.0%)

Note: Percentage of decision consistency: **82.1%**

CONCLUSIONS and RECOMMENDATIONS

We approached the issue of predicting the passing or failing of students who took the HSA in two ways, although more than two clearly exist by using other cut-score values. This problem could be approached as one in which we want to optimize the prediction of the HSA pass/fail determination by dichotomizing the MSA performance. The problem with this way, is that schools are familiar with the existing cut-offs on the MSA and changing that perception or adding another approach to thinking about the MSA scores might be confusing. Besides, the existing cut scores on the MSA have other important educational implications for NCLB, for example.

Setting the MSA cut-score at the proficient level: In the first approach, the MSA cut score (the predictor) was set to define the “proficient” category in the MSA as the single cut score, and students were re-categorized as at or above proficient or below proficient. The MSA and HSA results for the same group of students were compared in order to evaluate the predictability of passing the HSA (by using the cut score for “proficient” in the MSA as an external criterion). Ideally, students who were categorized as proficient in the MSA would pass the HSA, and students who were below proficient in the MSA would fail the HSA. From the above tables, we can see that the percentage of decision consistency for each county is greater than 65%, which indicates a moderate degree of consistency. Inconsistency is attributed largely to the fact that many students who were proficient in the MSA still failed the HSA (28.3% for the four counties as a whole) versus the opposite error (1.0% for the four counties as a whole). This implies that the cut score for passing the HSA was more demanding than the cut score for the “proficient” category in the MSA.

Setting the MSA cut-score at the advanced level: Alternatively, we could use the cut score for the “advanced” category in the MSA to divide students into advanced or below advanced, and compare the results with those of the HSA. The percentage of decision consistency increased, and the two cut scores displayed a moderate to high degree of consistency (the rate of correct classification is above 80% for each county). An examination of the sources of the inconsistency shows that, with the exception of

P.G. County, it comes from the fact that somewhat more students who were classified as advanced on the MSA actually failed the HSA. When aggregated over the four counties, slightly more students who were advanced on the MSA also failed the HSA (9.6%) versus the opposite classification (8.2% who were below advanced on the MSA passed the HSA). This implies that the cut score for the “advanced” category in the MSA was set slightly too low, if the purpose is to identify students who are likely to fail the HSA at the same percent as those below the MSA advanced category and passing the HSA. In other words, equating the two misaligned conditions.

Recommendation: The study by Lissitz, et. al. (2006) should be referred to for a more complex system (the use of MSA and other school and student level variables) to identify students at risk, but if a school system is going to use just the existing categories of MSA performance to identify students for interventions aimed at reducing failure on the HSA, they have a choice as to which cut-off to use. If they are equally interested in identifying students who are going to pass the HSA when they are predicted not to, and fail the HSA when they are predicted to pass, they should initially select all the students who were below the MSA advanced category for possible remediation. This is estimated to give them an overall decision accuracy of 82.1% correctly classified.

In this decision situation, the errors are usually based on the conditional selection of the students, and must be recalculated using percents that add to 100% for the row, since everyone in that row is selected for remediation or not selected for remediation and that is the group we are concerned with. The two types of classification error are of about the same magnitude: 17.3% ($\frac{1438}{1438 + 6882} = .173$) of the advanced students who are projected to be able to pass the HSA without aid from the remediation program will actually fail the HSA, and 18.6% ($\frac{1235}{1235 + 5417} = .186$) of the non-advanced students selected for remediation will actually pass the HSA even if they are not enrolled in the remediation program. If the County or the State initially only selects all the students who are below proficient for possible remediation only

5.4% ($\frac{150}{2622 + 150} = .054$) of the students selected this way would actually pass the

HSA even if they did not take the remediation and hence waste the resources associated with the remediation. However, more importantly, 34.7%

($\frac{4233}{4233 + 7697} = .347$) of the at-or-above proficient students who are not identified

as in need of remediation are estimated to fail the HSA. This is too large an error rate to tolerate, we believe.

These data suggest that the State and County should initially select all the students below the advanced level on the MSA for potential remediation since the potential failures from not providing remediation are estimated to be lower in this case than in the case of selecting only those below proficient. Even when providing remediation to all those students who are below advanced there is an error rate that might not be acceptable (17.3%). If this is the case, additional data might be collected that could involve, for example, interviews with teachers as well as other performance indicators that are available in a school such as grades or other test information to try to select the MSA advanced students who are still at risk.

It is not clear that this additional information can be used productively to identify students for remediation, although the paper by Lissitz, et. al. (2006) provides some evidence on this matter. Further work needs to be done prior to determining the confidence one might place in more sophisticated models of selection and the level of success of the various remediation strategies that might be suggested. For example, it might be that only students at the higher levels of MSA (e.g., proficient) can be remediated so they will be able to pass the HSA. Perhaps students at the lower levels (e.g., basic) cannot be brought up to the point where they can pass the HSA even with intense relatively short-term remediation. We do not yet know the answers to these matters. There are a number of interesting and important empirical questions that need to be answered before the state or any county would want to invest heavily in any remediation program offered for students identified by their MSA scores.

Finally, as noted above, this study is based on low-stakes testing of the HSA. The study should be repeated using the HSA in its high stakes testing environment.

References:

Lissitz, R. W., Wei Hua Fan, Alban, T., Hislop, B., Strader, D., Wood, C., and Perakis, S. (2006) The prediction of Performance on the Maryland High School Graduation Exam: Magnitude, Modeling and Reliability of Results. NCME, San Francisco.

Footnotes:

1. Funding for this project was provided by the Maryland State Department of Education to MARCES.