

Local Assessment Validity Study Report
February 2009
Executive Summary

Rayne A. Sperling, PhD
Associate Professor of Education

Jonna M. Kulikowich, PhD
Professor of Education

The Penn State Research Team

Introduction

Pennsylvania's State Board of Education regulations require every school district to have a local assessment system that can serve as an alternative high school graduation requirement for students who do not pass the 11th grade administration (or 12th grade retest) of the Pennsylvania System of School Assessment (PSSA). In 2007, more than 56,000 students statewide graduated based on these district-level determinations of proficiency.¹

This study documents considerable variance in the type and form of these local assessments as well as the manner in which these assessments are used as graduation requirements. In addition, the study reports ratings by panels of Pennsylvania educators that raise questions about the quality of local assessments used for graduation purposes statewide. Based on criteria established by the panels, evidence of alignment to standards and practices that could result in valid measures of proficiency was present from 5 percent of those school districts that submitted local assessments.

While the Penn State Research Team conducted this study at the request of the Secretary of Education, his staff had no role in reviewing or rating the assessments.

Survey of Local Assessments

Approximately 85 percent of school districts statewide responded to the Secretary of Education's July 28 and August 12, 2008 written requests for copies of local assessments used as graduation requirements in Mathematics and Reading; 418 districts submitted assessments by the deadline. The rate of response was roughly equal across rural, suburban, and urban districts.

In cataloguing the materials, considerable variance in the type and format of assessments used to inform local graduation decisions was noted. This variance included, for example, that some districts submitted a letter and others submitted various materials, in some cases multiple binders or boxes of materials, to represent their local assessment. Some districts reported that they did not have a local assessment. Others noted that proficiency on their local assessment was not a requirement for graduation purposes. Many districts reported that their curriculum was aligned to standards, and, as such, if students passed their coursework their performance in Mathematics and Reading was deemed proficient. Other districts reported use of attendance or citizenship as measures of proficiency.

¹ Pennsylvania Partnerships for Children, 2008

Many districts submitted commercially-developed assessments or assessments developed by Intermediate Units or other states. In all, districts submitted more than 60 different such assessments. Some examples of these types of assessments include the 4Sight Benchmark Assessments, tests developed to assess vocational skills, college entrance exams, and test preparation materials such as PSSA Coach.

Further, some districts reported the use of district-developed tests. In addition, as noted, many districts use course grades, portfolios, completion of a culminating project, or remedial programs—or some combination of these sometimes coupled with commercially available materials—as the local assessment. Whatever each district sent was used to create a folder that represented the local assessment for that district. Therefore, the local assessment folder might not have been an individual test, or any test at all, but might include a collection of tests from several courses or some combination of tests and other materials.

Ratings by Panels & Penn State Team

For purposes of this study, validity of a district’s assessment system was evaluated based on two factors: **Alignment** (do the assessment materials adequately measure 11th grade proficiency standards for Mathematics and Reading?) and **Practice** (how the district reports that the assessment is administered and how the results are used to determine proficiency). Given these rating systems, a local assessment could be considered as aligned but the practice reported might not be considered a valid means for determining proficiency. As an example, released PSSA items may be aligned to state academic standards; however, panelists concluded that the administration of previously-used test items may not be a valid practice to ensure proficiency.

Step 1 – Alignment

Eight panels comprised of public educators from across Pennsylvania reviewed all local assessment folders of materials submitted by the deadline in a process facilitated by the Penn State Research Team and national content experts. Educators reviewed assessments in their area of specialty (e.g., Mathematics teachers or curriculum coordinators only reviewed Mathematics assessments) and assigned ratings based on the rubric below.

<i>0 = No content areas represented; no alignment of outcomes to standards.</i>
<i>1 = Some content areas represented; some outcomes are aligned.</i>
<i>2 = Many to most content areas represented; most outcomes are aligned.</i>
<i>3 = All content areas represented; all outcomes are aligned.</i>

Panelists assigned each district folder a rating on the above 0-3 scale based on the degree of connection between the assessment and 11th grade state proficiency standards in Mathematics and Reading. For this judgment of alignment, a district's folder could earn a rating of 3 as long as every standard area was represented by at least one item across the range of submitted materials. Given the rating systems, a district's folder of assessment materials could be rated a score of '3' even if there were no practices reported that would indicate an individual student completed all items of the assessment. While more than 300 Mathematics assessments received a rating of 3, just 100 Reading assessments received scores of 3 by the Pennsylvania expert panelists.

Step 2 – Practice

The panels also established criteria for rating the manner in which local assessments are administered and how results are used in graduation decisions. To establish these criteria, the panels employed the use of a four-point scale, as found below, based on key factors such as test security, retake procedures and exposure (e.g., could students be assessed more than once based on the same items?), and the weight assigned to the local assessment for graduation decisions.

<i>0 = The practice cannot ensure proficiency level in Mathematics and/or Reading is met.</i>
<i>1 = The information about practices provided is insufficient to determine whether proficiency in Mathematics and/or Reading is met.</i>
<i>2 = There are some good practices reported to determine proficiency level; however, more information is needed.</i>
<i>3 = The practices reported represent a valid system of procedures to determine proficiency in Mathematics and/or Reading.</i>

Once again, diversity in documents provided and verified by school districts was evident. Local assessment practices vary considerably from district to district, and sometimes within district for individual students. As examples of this variance in practice, some districts monitor attendance in a remediation class and others administer retakes of released PSSA items until students meet proficiency levels. Some districts administer a secure locally-developed assessment. Another practice is the use of students' report card marks to measure proficiency in Mathematics and Reading.

Based on information submitted by districts and determinations made by the expert panels, fewer than 9 percent of Mathematics practices and 6 percent of Reading practices represent a valid system of procedures to determine proficiency (a rating of 3). Nearly half of all district practices (180) received ratings of 0 for both Mathematics and Reading. A 0 rating represents a practice as reported that cannot ensure students have met proficiency. Examples of practices that would have received ratings of 0 include summer course attendance after graduation, remedial course attendance without explanation of how grades are assigned, courses without an objective measure or portfolios with no standard rubric system, and tests not aligned to proficiency standards for performance in Mathematics or in Reading.

Based on this two-step process for examining Alignment and Practice, 5 percent of school districts that submitted assessment materials and reported practices were rated as having a valid local assessment system for measuring proficiency.

Table 1. Relations Between Alignment and Reported Practices in Mathematics and Reading.*

Alignment: Mathematics by Reading Alignment Overall (n = 378, 90.4% of reporting districts)

Reading→ Mathematics ↓	No Alignment	Some Alignment	Most Aligned	All Aligned
No Alignment	6 (.016)	0 (.000)	0 (.000)	0 (.000)
Some Alignment	2 (.005)	6 (.016)	3 (.008)	0 (.000)
Most Aligned	6 (.016)	22 (.058)	17 (.045)	1 (.002)
All Aligned	24 (.063)	74 (.195)	118 (.312)	98 (.259)

Practices: Mathematics by Reading Alignment Overall (n = 407, 97.1% of reporting districts)

Reading→ Mathematics ↓	Cannot Ensure Proficiency	Insufficient Evidence	Some Good Practices	Valid Practice
Cannot Ensure Proficiency	180 (.442)	3 (.007)	1(.002)	0 (.000)
Insufficient Evidence	41 (.101)	73 (.179)	1(.002)	0 (.000)
Some Good Practices	5 (.012)	17 (.042)	54 (.133)	1 (.002)
Valid Practice	4 (.010)	3 (.007)	5 (.012)	18 (.044)

*Table 1 in this executive summary corresponds to Table 14 in the report.

Conclusion

This study represents the most comprehensive examination to date of Pennsylvania district-level local assessments that serve as an alternative high school graduation requirement. In its review of nearly 1,000 assessments provided by 418 school districts statewide, the research team noted diversity in the type, format, and quality of materials that districts submitted and reported that they use to determine students' proficiency levels in Mathematics and Reading to meet graduation requirements.

Review of the materials by expert panels found variance in the degree of connection between the assessments and 11th grade proficiency standards. While Mathematics assessments were more often rated as aligned, ratings were considerably lower for Reading assessments.

Because alignment is a necessary but insufficient criterion for determining validity of an assessment, the research team also evaluated the administration and use of local assessment practices based on information provided by district staff. The researchers noted considerable variance among districts.

Finally, the research team examined ratings for alignment and practice and reported that for Mathematics, thirty-one school districts (8.1%) had ratings of '3' for both, while nineteen (5.1%) districts received ratings of '3' for both alignment and practice in Reading. Eighteen respondent districts received '3/3' ratings on both the Mathematics and Reading components of their local assessment system.