



A PACER Policy Brief:

Creating a Comprehensive Picture of School Performance

November 2015

Mark Duffy • Della Jenkins

I. Introduction

For more than two decades, states have been required to report publicly on the academic performance of schools and districts—first under the 1994 Improving America’s School Act, and more recently, under the No Child Left Behind Act (NCLB) and related state-level waivers. These reporting efforts typically build on results of large-scale standardized tests. For example, more than 90 percent of Pennsylvania’s current rating system—the School Performance Profile or SPP—is derived from standardized test results.

School rating systems have received increased public attention amid growing concerns about the prevalence and cost of standardized testing in schools. In a recent Gallup Poll, 64 percent of the public overall, and 67 percent of public school parents, said there is too much emphasis on standardized testing in education.¹ Another part of the same poll asked respondents to rate the importance of five approaches to measuring school effectiveness; standardized testing came in last.

These perceptions have been acknowledged by policymakers at both the state and federal levels. Earlier this year, Governor Tom Wolf expressed concern about the SPP, saying: “Education is a full and holistic process. We’ve reduced it to a bunch of high-stakes tests that don’t seem to me to be tied to the specific, comprehensive skills that we want students to have.”² And, just last month, the U.S. Department of Education (USDE) released a Testing Action Plan to help states in administering “fewer and smarter assessments.”³ In an open letter to parents and teachers, President Obama argued that “classroom work, surveys, and other factors” can provide “an all-around look at how our students and schools are doing.” The USDE guidance calls on Congress to ensure that any reauthorization of NCLB/the Elementary and Secondary Education Act (ESEA) allows states to use indicators beyond standardized test scores in holding educators and schools accountable for student success.

In this brief, Research for Action (RFA) provides background on the existing (though limited) research related to best practices in reporting on school performance. We also offer examples of reporting systems from neighboring states, and states considered to be education leaders based on results of the National Assessment of Educational Progress (NAEP). This brief is intended to support education policymakers and stakeholders in their deliberations on the future of Pennsylvania’s approach to school ratings.

¹ <http://pdkpoll2015.pdkintl.org/236>

² <http://www.newsworks.org/index.php/local/education/79942-pa-gov-wolf-says-school-ratings-should-be-less-tied-to-tests>

³ <http://www.ed.gov/news/press-releases/fact-sheet-testing-action-plan>

II. Pennsylvania’s School Performance Profile

Under NCLB, each state education agency that receives Title I, Part A funds is required to prepare and disseminate an annual report card with information about public school performance. Pennsylvania’s current reporting system centers on 100-point SPP scores awarded at the school building level every academic year.⁴ SPP scores are primarily derived from two types of data elements—student proficiency and growth—both of which are calculated based on standardized test results. Overall, 90 percent of schools’ base scores rely on test scores (see Table 1).

Table 1. Pennsylvania’s School Performance Profile

DATA ELEMENTS	PERCENTAGE OF TOTAL SPP SCORE	ELEMENT DETAILS
1. Indicators of Academic Achievement	40 percent	Percent scoring proficient or advanced on PSSAs or Keystone Exams in tested subjects, performance on industry standards-based competency assessments, grade 3 reading proficiency, and SAT/ACT benchmarks
2. Indicators of Academic Growth/PVAAS	40 percent	Meeting state-identified annual academic growth expectations on PSSAs or Keystones in tested subjects for grades 4-8 and 11
3. Indicators of Closing the Achievement Gap - All Students	5 percent	Percent of annual achievement gap closures met in math, reading, science, and writing among all students and historically underperforming students
4. Indicators of Closing the Achievement Gap: Historically Underperforming Students	5 percent	
5. Other Academic Indicators	10 percent	Cohort graduation rate, promotion rate, attendance rate, Advanced Placement/International Baccalaureate or college credit, PSAT/Plan participation
6. Extra Credit for Advanced Achievement	Up to 7 additional points	Percent scoring advanced on PSSA in math, reading, science, and writing; on industry standards-based competency assessments; percent scoring 3 or higher on AP exams

Source: Pennsylvania Department of Education

How the SPP is Calculated

Each school’s SPP is calculated by taking points earned in the five main categories as a percentage of points available, then adding points earned in the sixth extra credit category, as applicable. Although the weight for each main category is constant across schools (with a partial exception for Career and Technical Centers), the weight of some factors within categories varies between elementary and secondary schools. For example, the “Indicator of Academic Achievement” category includes a measure of 3rd grade reading proficiency, which does not apply to high schools. “Other Academic Indicators” includes availability of Advanced Placement courses, which does not apply to elementary schools.⁵

⁴ <http://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/parenwalreq2015.pdf>, p. 2

⁵ <http://www.researchforaction.org/wp-content/uploads/2015/03/RFA-PACER-SPP-Brief-March-2015-Final.pdf>

In September 2015, Governor Wolf announced a moratorium on the use of the Pennsylvania System of School Assessment (PSSA) data in the SPP following the release of scores that dropped between the 2013-14 and 2014-15 school years. This drop resulted from the more demanding nature of the test, now fully-aligned to the Pennsylvania Core Standards.⁶

III. Best Practices for Measuring and Reporting School Performance

Overreliance on standardized assessments as the sole indicator of school or student performance is widely discouraged, both in policy and academic circles. For example, researchers from the University of Southern California’s Rossier School of Education argue that such systems “expect the same performance from all schools regardless of their student inputs” and therefore may penalize schools for factors they cannot control.⁷ School accountability systems which rely heavily on test scores “assume that aggregate math and English scores closely proxy important unmeasured goals such as citizenship, ethics, and critical thinking.”⁸

Similar concerns inherently arise around the design and implementation of large-scale systems of reporting on school performance; however, limited research exists on the validity and utility of current school ratings systems and ways they measure school performance.

Joan Hermann from the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) at UCLA argues that sound school performance indicators are:

1. “Aligned with and provide accurate data on essential education system elements, processes and outcomes;
2. fair, valid, and reliable; comparable, credible, and meaningful;
3. comprehensible and understandable to intended users; and
4. actionable and feasible.”⁹

How do rating systems—especially those that, like SPP, employ a summary index score or grade letter—fare against these standards?

The CRESST principles—along with more general guidance from the American Psychological Association (APA), American Education Research Association (AERA), and National Council on Measurement in Education (NCME) Standards for Educational and Psychological Testing (2014) concerning assessment practices—highlight several points for consideration. First, in order for measures to be valid, they need to address a more comprehensive set of outcomes for schools than simply test results. Second, a reliable school performance report should include multiple measures of performance and/or the results of several administrations of the same measure to ensure consistency. Third, to report fairly, a school profile should account for factors outside of measures included—comparing schools with similar student demographics as opposed to a statewide comparison is one way to address this issue.

The RAND Corporation has inventoried school ratings measures beyond those required by NCLB. RAND’s survey found that such measures commonly include student performance in additional subjects not normally covered by state-wide assessment systems (e.g., social studies); measures of growth in student

⁶ <http://www.mcall.com/news/local/mc-pa-pssas-waiver-school-performance-profile-20150908-story.html>

⁷ <http://edr.sagepub.com/content/43/1/45.full.pdf+html>

⁸ <http://edr.sagepub.com/content/43/1/45.full.pdf+html>

⁹ https://www.cse.ucla.edu/products/states_schools/ISSQ_v3.pdf

performance over time as compared with levels of proficiency in a given year; and college-readiness measures, such as ACT scores or AP course taking and test results.¹⁰

Similarly, a work group of scholars and policy organizations from across the country developed a common set of recommendations around school accountability. The lead authors, researchers from the Center on Reinventing Public Education (CRPE) and the Stanford Center for Opportunity Policy in Education (SCOPE), concluded that longer term measures of student achievement, such as progress toward graduation and career readiness, “provide needed counterweights to standardized testing.”¹¹

The National Education Policy Center (NEPC) reviewed “A to F” grading schemes in school reporting and argued that such an approach, while designed as an intuitive signal of school performance, “hides valuable information while invalidly combining disparate and unrelated objects.”¹² While SPP rests on a 0 to 100 scale—rather than a letter grade—it similarly uses a singular value to quantify a school’s performance. In addition, the use of a single index score to label a school as successful or struggling requires the use of arbitrary cut points which may not offer a meaningful distinction of performance.¹³ As an alternative, NEPC suggests using multiple indicators that reflect different components of school quality.¹⁴ NEPC also recommends that states balance data on desired outcomes with data on the inputs that may affect these outcomes when reporting on school performance, such as student characteristics and funding.¹⁵

IV. Reporting on School Performance in Neighboring and Leading States

To provide context for Pennsylvania’s deliberations on SPP, we conducted a scan of school reporting practices in neighboring states, as well as in three states that consistently perform well across grades and subjects on the NAEP: Massachusetts, Minnesota, and New Hampshire.¹⁶ Table 2 indicates whether the state:

- Sums indicators into a single letter grade or score;
- Reports on each measure individually with no scoring or comparison; or
- Compares the performance of individual schools to other (peer) schools with similar student populations.

¹⁰ http://www.rand.org/content/dam/rand/pubs/technical_reports/2011/RAND_TR968.pdf

¹¹ https://edpolicy.stanford.edu/sites/default/files/publications/accountability-and-federal-role-third-way-esea_0.pdf

¹² http://greatlakescenter.org/docs/Policy_Briefs/Mathis-RBOPM.pdf

¹³ <http://edr.sagepub.com/content/43/1/45.full.pdf+html>

¹⁴ <http://nepc.colorado.edu/publication/why-school-report-cards-fail>

¹⁵ http://greatlakescenter.org/docs/Policy_Briefs/Mathis-RBOPM.pdf

¹⁶ Consistent high scorers on Math and Reading across both 4th and 8th grades.

Table 2. Comparing Pennsylvania’s school performance reporting with neighboring and leading states

STATE	RATING		INDIVIDUAL MEASURES ONLY	PEER COMPARISON
	A-F LETTER GRADE	NUMERIC INDEX SCORE		
Pennsylvania		●		
Neighboring States				
Delaware		●		
Maryland		●		
New Jersey				●
New York			●	
Ohio	●			
West Virginia		●		
Leading States based on NAEP Results				
Massachusetts			●	
Minnesota		●		
New Hampshire			● ¹⁷	

Looking across our sample, many states use a summative rating in the form of either a single letter grade or numeric index encompassing multiple measures. While Pennsylvania’s SPP is technically a composite index, its current form is so heavily reliant on standardized test data that it functions more like a singular grade. Other states present data on each indicator or measure individually, rather than aggregating to a grade or score. In Massachusetts and New Hampshire, schools are not scored at all; instead, both states provide data on a number of input and outcome measures, which, taken together, form a snapshot of each school.

New Jersey’s rating system includes a comparative approach: School level data is reported across performance areas including academic achievement, college and career readiness, student growth (elementary and middle schools), and graduation and postsecondary (high schools). For each performance category, the public can compare a given school to other schools with similar characteristics—allowing school performance to be considered in the context of a meaningful peer group, and acknowledging the impacts of factors such as student poverty, special education enrollment, and English Language Learners.

¹⁷ Comparisons provided only on enrollment and only at district level.

V. Examining Measures of School Performance in Neighboring and Leading States

Using Outcomes to Measure Student Achievement and Advancement

How states describe school performance is one question; the specific indicators that inform that description is another. Table 3 highlights common or notable outcomes measures used in neighboring states and those used in three states that consistently perform well on NAEP.

Table 3: Outcomes measured by school performance systems

		STATES BORDERING PENNSYLVANIA						HIGH-PERFORMING NAEP			PA
		DE	MD	NJ	OH	WV	NY	NH	MA	MN	
Achievement & Progress	State assessments	●	●	●	●	●	●	●	●	●	●
	SAT/ACT scores/participation	●	●	●	●	● ¹⁸					●
	AP/IB involvement		●	●	●						●
	Achievement gap reduction		●	●	●	●		●	●	●	●
	Student growth	●	●	●				●	●	●	●
	NAEP proficiency			●			●	●	●		●
Advancement	Attendance	●	●	●		●		● ¹⁹	●	●	●
	Promotion rate by grade	●			● ²⁰						●
	Dropout rate		●	●		● ²¹	●	●	●	●	●
	Graduation rate	●	●	●	●	●	●	●	●	●	●
	Postsecondary planning or enrollment		●	●			●		●	●	●

Our review indicates that the most common measures of school performance in neighboring and high performing states are consistent with those identified in the RAND study, and focus on student achievement, student progress, and college and career readiness. Outcomes in these areas are measured almost exclusively with some combination of standardized test scores, including state-specific assessments, NAEP, SAT, ACT, and AP tests. Pennsylvania utilizes all but NAEP among these common outcome measures of student achievement.

¹⁸ ACT only

¹⁹ Elementary and middle schools only

²⁰ Based on percent of students meeting 3rd grade reading requirement to advance to 4th grade.

²¹ Only per county, not per school.

Many states also report on measures of student advancement, including attendance, grade promotion, dropout rate, graduation rate, and postsecondary enrollment. Pennsylvania’s SPP utilizes four out of five of these non-assessment based indicators of student advancement; however, they cumulatively account for only 10 percent of a school’s overall score.

Capturing Input Measures

Some states also report on inputs: Characteristics of the student body, staff, and school which may impact outcomes as well as the resourcing, staffing, and other supports available to schools. Table 4 summarizes notable characteristics by state and type. Importantly, in Pennsylvania and other states, these elements are often reported as descriptive data, and do not factor in to the grades or scores used for rating purposes.

Table 4. Input measures: Neighboring and high-performing states

		STATES BORDERING PENNSYLVANIA					HIGH-PERFORMING NAEP				
		DE	MD	NJ	OH	WV	NY	NH	MA	MN	PA
Teachers	Percent of teachers by ethnicity	●									
	Percent highly qualified teachers	●	● ²²			●	● ²³	●	●	●	●
	Student-staff ratio or class size	●		●		●	●		●	●	
Student Demographics	Enrollment by subgroup	●		●		●	●		●	●	●
	Language diversity			●							
School Climate	School environment survey	● ²⁴								●	
	Suspensions or expulsion rate	●		●			●		●		
Curriculum & Programs	Instructional time			●							
	Participation in visual & performing arts			●							
	Career & technical ed. participation		●	●							
School Funding	District expenditure per pupil	●									

²² Reported as % of classes taught by highly qualified teachers.

²³ Reported as % of classes taught by highly qualified teachers.

²⁴ Beginning in SY 2015/16

Most states in our sample—Pennsylvania included—report on the percentage of teachers who are considered highly qualified and provide student demographic data by subgroup. Measures of school climate and curriculum are less common. Ohio is the notable exception from our sample, in that they do not currently report on any input measures as part of their school performance reporting. Delaware focuses heavily on teachers, reporting on the percentage of teachers who are highly qualified, experience level, degree type, and ethnicity of teachers, as well as student-staff ratios. New Jersey emphasizes curriculum and programs, and is the only state to report on instructional time.

Models from Large Urban Districts

In addition to required, state-level reporting systems, several major districts have pursued their own reporting initiatives. We highlight two below.

New York City

New York City’s school performance system is innovative in two major ways: 1) It reports on a large set of student demographic and academic data, including percentage of students in temporary housing and average incoming proficiency; and 2) It supplements the standard data on student achievement, student growth, and college readiness with qualitative measures of instruction and school climate. Schools are observed and evaluated on curriculum, teaching and learning, and staff communication by an experienced educator. Student, teacher, and parent satisfaction with curriculum, climate, and safety are also measured annually by the NYC School Survey. The end result is a “Quality Snap Shot” for each school, which offers both quantitative and qualitative data to provide a more nuanced picture of the building.

Philadelphia

Philadelphia’s School Progress Reports include four domains: 1) Achievement, 2) Progress, 3) Climate, and 4) College and Career. The Climate domain is largely quantitative and focuses on attendance, retention, and suspensions, as well as survey results measuring student and parent perceptions of school climate and parent engagement. For each domain, information is reported as an index score, along with a “performance tier” (intervene, watch, reinforce, and model), city rank, and peer rank. Peer ranks compare a school’s overall and domain scores to those of a peer group of schools with the same grade configuration and similar student demographics, including poverty, ethnicity, special education status, and limited English proficiency.

VI. The Future of Pennsylvania’s School Performance Profile

As Pennsylvania prepares to re-evaluate its approach to measuring and reporting on school performance, policymakers may want to consider both audience and purpose. Who are the consumers of information on school performance, and what are the principal concerns of these constituencies? How is the information depicted and communicated, and for what purposes? And how can the state balance calls for accessibility and utility of school performance measures, with commitment to validity and accuracy?

In considering these questions, Pennsylvania may wish to look to alternative models in neighboring and leading states or large urban districts which:

- Report on a broader set of school performance measures;
- Include more input measures about school staff, student body, climate, and curriculum; and
- Forgo a single grade, score, or other school rating in favor of peer comparisons between schools with similar traits.

Together, these approaches may allow for more nuanced and valid measures of school performance.

RFA is grateful to The Heinz Endowments and The Philadelphia Foundation for their generous support of this initiative.

THE HEINZ ENDOWMENTS
HOWARD HEINZ ENDOWMENT • VIRA I. HEINZ ENDOWMENT

The Heinz Endowments supports efforts to make southwestern Pennsylvania a premier place to live and work, a center for learning and educational excellence, and a region that embraces diversity and inclusion.

**THE
PHILADELPHIA
FOUNDATION**

The Power of Endless Possibilities.™

The Philadelphia Foundation, a public charity community foundation, has been committed to improving the quality of life in Bucks, Chester, Delaware, Montgomery, and Philadelphia counties for 90 years,

The authors would also like to thank Policy Intern Garrett Baker for providing invaluable assistance with data collection and analysis for this policy brief.