

# CENTER FOR EVALUATION AND EDUCATION POLICY ANALYSIS (CEEPA)

DEPARTMENT OF EDUCATION POLICY STUDIES  
COLLEGE OF EDUCATION, THE PENNSYLVANIA STATE UNIVERSITY

---

ED FULLER, PHD  
EXECUTIVE DIRECTOR

## Policy Brief 2014-1: An Examination of Pennsylvania School Performance Profile Scores

### Executive Summary

The purpose of school accountability systems is to accurately assess school performance and effectiveness and communicate the judgments about schools to educators, policymakers, and the public. The Commonwealth of Pennsylvania has created School Performance Profile (SPP) scores as a measure of school performance and effectiveness. Moreover, the Commonwealth has included these school-level scores as a component of teacher and principal evaluations.

Researchers have consistently argued that accountability measures such as SPP scores must be adjusted for factors *outside the control of educators* in order to accurately identify school effectiveness. This policy brief examines the relationship between SPP scores and student- and school- characteristics that are outside the control of educators.

The analyses contained in this brief strongly suggest that the Commonwealth's SPP scores are strongly associated with student- and school-characteristics, thus are inaccurate measures of school effectiveness. SPP scores, in fact, are more accurate indicators of the percentage of economically disadvantaged students in a school than of the effectiveness of a school.

Thus, SPP scores should not be used to make judgments about school effectiveness unless the scores from one school are compared to only the SPP scores from schools with similar student- and school- characteristics. There are a number of options that the Commonwealth could employ to calculate SPP scores that are more accurate measures of school effectiveness. In doing so, the Commonwealth would be assisting educators to improve their practice while providing valid information to the public and policymakers about the effectiveness of their local schools.

### Introduction

There is near universal agreement that schools should be held accountable for meeting high expectations.<sup>1</sup> Indeed, every state has adopted some form of a school accountability system. There are, however, serious questions about what these accountability systems actually measure and whether the systems accurately identify school effectiveness.

This policy brief examines the components of the School Performance Profile (SPP) scores, the stated purposes of the scores, and whether the current calculation of the scores meet the stated purpose of the scores as defined by the Pennsylvania Department of Education (PDE). Specifically, this brief focuses on if SPP scores accurately identify school effectiveness.

### Components of the School Performance Profile Scores

According to PDE, there are multiple components of the SPP scores.<sup>2</sup> These components, and their associated weight in calculating the overall SPP score, include:

- Indicators of Academic Achievement (40%)
- Indicators of Closing the Achievement Gap – All Students (5%)
- Indicators of Closing the Achievement Gap – Historically Underperforming Students (5%)
- Indicators of Academic Growth / PVAAS (40%)
- Other Academic Indicators (10%)

The individual indicators within each component are used to calculate the percentage of each component met by a school. The component percentages are then added together to arrive at a School Profile Performance score that ranges from 0 to 100. In addition, schools can earn up to seven bonus points through meeting additional indicators such that the maximum score is 107.<sup>i</sup>

### **Purpose of School Performance Profile Scores**

The research literature on school accountability systems suggests there are two primary purposes of school accountability systems:

- (1) to accurately measure school performance and effectiveness as a means of providing information to the public, policymakers, and educators; and,
- (2) to identify effective and ineffective schools as a means to signal to those working in ineffective schools to model their behavior after those working in effective schools.<sup>3</sup>

Evaluations such as school accountability programs should be based on defensible criteria<sup>4</sup> that lead to “ethical, fair, useful, feasible, and accurate” conclusions.<sup>5</sup> This recommendation is also referred to as construct validity—the ability of the evaluation effort to provide accurate information that can lead educators and policymakers to make appropriate conclusions from the information. The Joint Committee on the Standards for Educational Evaluation recommends that only evaluations that can provide evidence about construct validity should be used.

In order for the signaling effects of the SPP scores to have their intended effect on educators, educators must perceive the system as accurate, fair, and equitable.<sup>6</sup> In other words, the SPP scores must have “face validity”<sup>7</sup> from the perspective of educators. If, in fact, educators *do not* perceive the SPP scores to have face validity, then they are likely to ignore, subvert, or “game” the entire evaluation system.<sup>8</sup>

If an evaluation such as a school accountability system lacks either construct or face validity, then the system will clearly not have the intended effects upon educators. Thus, it is critical that the Commonwealth provide evidence about the construct validity of the SPP effort in order to ensure that (1) educators view the SPP scores as credible and (2) use them in ways that effectively improve schools.

### **Examination of the Construct Validity of SPP Scores**

Perhaps the biggest threat to the construct and face validity of the SPP scores is the inability of the scores to actually capture the true effectiveness of schools. To capture true school effectiveness, the SPP scores must be adjusted for the effects of student- and school-characteristics *outside the control of educators* such as the percentage of economically disadvantaged students<sup>ii</sup>, percentage of English-Language Learner students, percentage of minority students, parental level of education as well as school factors such as school size, school geographic location, and per pupil expenditures.<sup>9</sup> In short, if the scores accurately capture the true effectiveness of schools and the educators within schools, there should be only a weak or non-existent relationship between the SPP scores and the percentage of economically disadvantaged students.<sup>10</sup>

In the graphs below, we examine the relationship between school SPP scores and the percentage of economically disadvantaged students in the schools. Because the measures used to create the scores and the percentage of economically disadvantaged students differ across school levels, the analysis is presented for three school levels: elementary schools (those serving grades PK-6), middle schools (those serving grades 6-8), and high schools (those serving grades 9-12). *Again, if the scores truly capture school effectiveness, there should be only a weak or non-existent relationship between the SPP scores and the percentage of economically disadvantaged students.*

---

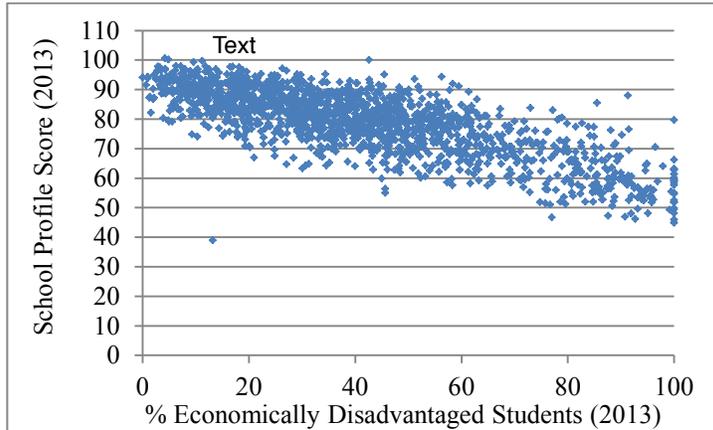
<sup>i</sup> See [http://www.portal.state.pa.us/portal/server.pt/community/pennsylvania\\_department\\_of\\_education/7237/p/1604316](http://www.portal.state.pa.us/portal/server.pt/community/pennsylvania_department_of_education/7237/p/1604316)

<sup>ii</sup> This is based on student participation in the federal free-/reduced-price lunch program

## Elementary Schools

As Figure 1 shows, there is a ***strong relationship*** between the percent of economically disadvantaged students and the SPP scores. Indeed, the correlation coefficient—a measure of the strength of a relationship<sup>iii</sup>—is greater than 0.600.

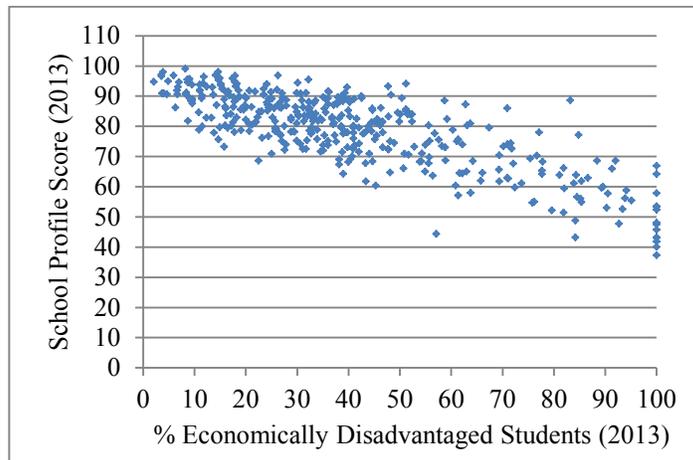
Figure 1: Relationship between the Percent of Economically Disadvantaged Students and SPP Scores for Elementary Schools



## Middle Schools

As Figure 2 documents, there is a ***strong relationship*** between the percent of economically disadvantaged students and the SPP scores at the middle school level. Specifically, the correlation coefficient is 0.649. Thus, at the middle school level, the SPP scores appear to be a better measure of the percentage of economically disadvantaged students than as a measure of school effectiveness.

Figure 2: Relationship between the Percent of Economically Disadvantaged Students and SPP Scores for Middle Schools

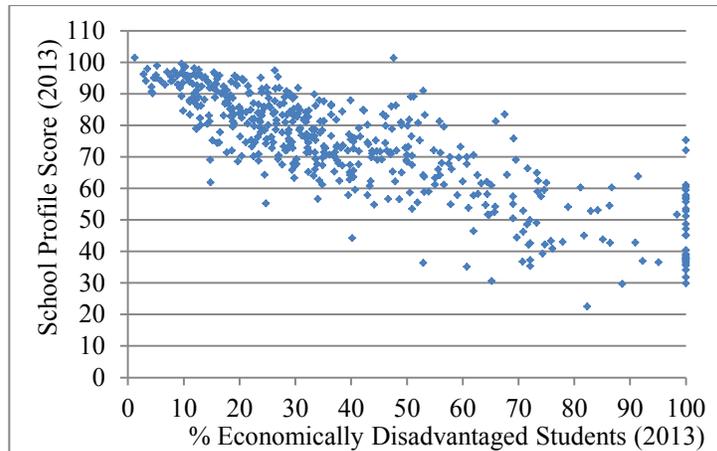


<sup>iii</sup> The correlation coefficient ranges from 0.000 to 1.000, with coefficients near 0.000 indicating a very weak relationship and coefficients approaching 1.000 indicating an extremely strong relationship.

## High Schools

Figure 3 demonstrates there is also a *very strong relationship* between the percentage of economically disadvantaged students and the SPP scores. In this case, the correlation coefficient is 0.675 when all high schools were included in the analysis and 0.686 when career/technical and special education schools were excluded from the analysis. Both relationships are very strong and indicate that the percentage of economically disadvantaged students is strongly related to the SPP scores.

Figure 3: Relationship between the Percent of Economically Disadvantaged Students and SPP Scores for High Schools



On the right hand side of Figure 3 are the schools with 100% of students identified as economically disadvantaged. Some of the schools have SPP scores similar to the SPP scores for schools with much lower percentages of economically disadvantaged students. Examination of these schools identified almost all of them as magnet schools. Such schools typically do not enroll all students, but admit only those students meeting certain entrance criteria set by the school. When magnet schools are removed from the analysis, the correlation coefficient increased to 0.723. Thus, when considering only open-enrollment high schools, the percentage of economically disadvantaged students explained 72.3% of the variation in SPP scores across schools.

### Adjusting for Multiple School Factors through Regression Analysis

While the percentage of economically disadvantaged students is strongly associated with school profile scores, other school factors *outside the control of educators* also influence student outcomes that are components of the total school profile score. While correlational analyses examine the relationship between two variables—in the above cases, the percentage of economically disadvantaged students and the SPP scores—regression analysis can concomitantly examine the relationship between multiple school factors and SPP scores.

To examine the relationship between multiple school factors outside the control of educators and school profile scores, the following school characteristics were included in the regression analyses: percentage of economically disadvantaged students, percentage of White students, percentage of Asian American students, percentage of students on an individualized education plan (special education students), school size, geographic location (rural, suburban, urban), and measures of district wealth.

When all of these school characteristics are included in the regression analysis, they explain 61%, 66%, and 71% of the variation in SPP scores for elementary-, middle-, and high-schools, respectively (See Table 1). At the high school level, when special education-

career/technical-, and magnet- schools are removed from the analysis, the set of school characteristics explains 78% of the variation in SPP scores.

Table 1: Percent of the Variation in PA School Profile Scores Explained by School Characteristics Outside the Control of Educators

School School Level	% of Variation in SPP Scores Explained
Elementary Schools	61%
Middle Schools	66%
High Schools	71%-78%

Why is this important? Because these results indicate that the vast majority of the differences in SPP scores across schools are explained by student- and school-characteristics that are *not under the control of educators*. In fact, as currently calculated, ***the SPP scores are more accurate at identifying the percentage of economically disadvantaged students in a school than at identifying the effectiveness of a school.***

### Conclusions

There are a number of important conclusions from this analysis that are relevant to both educators and policymakers.

- (1) Both existing research on school accountability measures and the analyses contained in this report strongly suggest that the SPP scores provide little information about the effectiveness of schools, principals, or teachers. Indeed, the scores are very strongly correlated with factors outside the control of educators. ***Thus, as currently calculated, SPP scores should not be used as an indication of either school effectiveness or as a component of educator evaluations.***
- (2) ***Support for the current system will quickly dissipate once educators, policymakers, and the public understand in that SPP scores are not accurate indicators of school effectiveness.*** The increasing lack of support for the system will severely undercut the purposes of the SPP scores.
- (3) Because SPP scores are inaccurate assessments of school effectiveness, ***the use of the SPP scores in teacher and principal evaluations will lead to inaccurate judgments about teacher and principal effectiveness.*** Because the SPP scores are so strongly correlated with student characteristics, teachers and principals in schools serving high percentages of economically disadvantaged students will be identified as less effective than they really are while those serving in schools with low percentages of economically disadvantaged students will be identified as more effective than in actuality.
- (4) Because of the strong relationship between school- and student-characteristics and SPP scores, the use of SPP scores in teacher and principal evaluations will create an additional incentive for the most qualified and effective educators to seek employment in schools with high SPP scores. ***Thus, the use of SPP scores on educator evaluations will simply exacerbate the existing inequities in the distribution of educator quality across schools based on the characteristics of the students enrolled in the schools.***

## Recommendations

The US Department of Education (USDoE) prohibits states from adjusting student growth scores based on student characteristics when identifying the effectiveness of schools with high percentages of economically disadvantaged students (Title I schools). Unfortunately, this prohibition would likely be true for attempts to adjust the specific indicators of the SPP or the SPP scores based on student characteristics. This does not mean, however, that there are not options that would improve the SPP scores as measures of school effectiveness.

- (1) ***Review the percentage weights assigned to the various SPP components.*** Specifically, the Commonwealth should carefully assess the weights assigned to the individual indicators and components and discuss increasing the weights of the indicators and components with the weakest relationships with student- and school- characteristics.
- (2) ***Create an online tool that identifies comparison schools for each school in the Commonwealth.*** The identification of comparison schools would be based on high-quality statistical efforts that accurately identify schools with similar student- and school- characteristics. The set of comparison schools would provide educators with an appropriate set of schools against which they could compare their own school effectiveness score. Such a system would also give local educators and policymakers a far more accurate view of local school effectiveness.
- (3) ***Construct an alternative rating system outside the system required by the USDoE.*** This alternative system would adjust the SPP scores for student- and school- characteristics outside the control of educators so these alternative SPP scores would more accurately capture school effectiveness. This would be beneficial in two ways. First, the public and policymakers would have more accurate information about schools, thus could make far more informed judgments and choices about the schools. Second, educators in lower performing schools could accurately identify high-performing comparison schools from which they could learn.
- (4) ***Recognize the flaws in the current system and work collaboratively to build a more accurate system.*** The Commonwealth should recognize the strengths and weaknesses of the current SPP effort and engage educators, policymakers, and the public in a discussion about how to more accurately capture school effectiveness. Importantly, the Commonwealth should provide data to researchers so that those with experience in evaluating such systems could provide unbiased and useful information about creating more effective systems.

## Further Study

The next policy brief will review the research concerning the specific components of the SPP and the degree to which the components are correlated with student characteristics. The brief will also examine the correlation between some of the specific components of the SPP and student characteristics—in particular, the percentage of economically disadvantaged students.

In addition, CEEPA will create a new index that adjusts the existing scores based on available data related to student characteristics and other school contextual factors. This will assist educators in making more accurate judgments about their own effectiveness and in selecting appropriate comparison schools.

---

The views expressed in this policy brief are solely those of the author. The expressed views do not necessarily represent those of the Educational Leadership Program, Department of Education Policy Studies, the College of Education, or The Pennsylvania State University or any employees within these organizations.

- 
- <sup>1</sup> Polikoff, M. S., McEachin, A. J., Wrabel, S. L., & Duque, M. (2014). The waive of the future? School accountability in the waiver era. *Educational Researcher*, 43(1), 45-54.
- <sup>2</sup> School Performance Profile Understanding Academic Performance Score, PDE (n.d.) Retrieved at [http://www.portal.state.pa.us/portal/server.pt/community/pennsylvania\\_department\\_of\\_education/7237/p/1604316](http://www.portal.state.pa.us/portal/server.pt/community/pennsylvania_department_of_education/7237/p/1604316). on September 19th, 2014.
- <sup>3</sup> Polikoff, M. S., McEachin, A. J., Wrabel, S. L., & Duque, M. (2014). The waive of the future? School accountability in the waiver era. *Educational Researcher*, 43(1), 45-54.
- <sup>4</sup> Fitzpatrick, J., Sanders, J. R., & Worthen, B. R. (2011). *Program Evaluation: Alternative Approaches and Practical Guidelines* (4th ed.). New Jersey: Pearson Education.
- <sup>5</sup> Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks, CA: Sage.
- <sup>6</sup> Grissom, J., Kalgories, D., & Loeb, S. (2012, November). "Using student test scores to measure principal performance." Working paper No. 18568. Cambridge, MA: National Bureau of Economic Research.
- Kane, T.J. & Staiger, D.O. (2002). "The promise and pitfalls of using imprecise school accountability measures," *The Journal of Economic Perspectives*, 16, 91-114.
- Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks, CA: Sage.
- <sup>7</sup> Fitzpatrick, J., Sanders, J. R., & Worthen, B. R. (2011). *Program Evaluation: Alternative Approaches and Practical Guidelines* (4th ed.). New Jersey: Pearson Education.
- Koretz, D. M. (2008). *Measuring up*. Harvard University Press.
- <sup>8</sup> Erdogan, B. (2002). "Antecedents and consequences of justice perceptions in performance appraisals." *Human Resource Management Review*, 12: 555-578.
- Kane, T.J. & Staiger, D.O. (2002). "The promise and pitfalls of using imprecise school accountability measures," *The Journal of Economic Perspectives*, 16, 91-114.
- <sup>9</sup> Ehler, M., Koedel, C., Parsons, E., & Podgursky, M. (2013). *Selecting growth measures for school and teacher evaluations: Should proportionality matter?* (CALDER Working Paper No. 80). Washington, DC: National Center for Analysis of Longitudinal Data in Education Research.
- <sup>10</sup> Fuller, E. J., & Hollingworth, L. (2014). A Bridge Too Far? Challenges in Evaluating Principal Effectiveness. *Educational Administration Quarterly*, 50(3): 466-499.
- Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks, CA: Sage.

This paper was subjected to peer-review that included four professors and two graduate students. Two of the professors and the graduate students are at Penn State. The other two professors are at other institutions outside of the Commonwealth of Pennsylvania.

#### **ABOUT THE CENTER FOR EVALUATION AND EDUCATION POLICY ANALYSIS (CEEPA)**

*The mission of the CEEPA is to provide unbiased, high-quality evaluation and policy analysis services to education and other organizations in the Commonwealth of Pennsylvania and across the nation. CEEPA provides evaluation and research services to schools, school districts, universities, governmental entities, and other organizations.*

*Dr. Fuller is an associate professor in the Educational Leadership Program within the Education Policy Studies Department in Penn State's College of Education. His areas of expertise are evaluation, educator quality, educator turnover, and the evaluations of school leaders and leadership preparation programs. He has extensive experience in conducting evaluations and working with state legislators in Texas and has presented and published numerous papers on a wide-variety of education topics. He can be reached at [ejf20@psu.edu](mailto:ejf20@psu.edu) or 814-865-2233(w)/512-971-5715(c).*