



Testimony on Keystone Exams, Teacher and Principal Evaluation System, and School Accountability

Hearing of the Senate Education Committee June 5, 2017

Chairman Eichelberger, Chairman Dinniman, and members of the Senate Education Committee, my name is Michael Leichliter and I serve as Superintendent of the Penn Manor School District in Lancaster, PA. I am here with our Executive Director, Mark DiRocco, on behalf of more than 800 superintendents, assistant superintendents, and school leaders of the Pennsylvania Association of School Administrators to provide comments regarding the Keystone Exams and their relation to the teacher and principal evaluation system, and the school accountability system.

As an individual superintendent I was an early advocate six years ago for House Bill 1980, later Act 82 of 2012. The Penn Manor School District participated as a pilot district in the teacher effectiveness project, which formed the basis for working with then Rep. Aument as he refined and moved the bill through the General Assembly. Improving Educator Effectiveness is critical to providing our children with a quality education and we are pleased to offer our insights from the field.

We were asked to address four questions. First, **“Is the Keystone Exam graduation requirement a good tool for evaluating teachers and principals?”** Keystone Exams are state mandated end-of-course tests that were designed for two purposes: 1) to determine high schools’ Adequate Yearly Progress (AYP) status as part of the former No

Child Left Behind (NCLB) accountability requirements, and 2) for students in the Class of 2017 (now the Class of 2019) and beyond to demonstrate proficiency in core subjects as a graduation requirement. The Keystone Exams were not designed to be used for educator evaluation or as a high school graduation requirement. Using a single standardized test score such as the Keystone Exam as a component of an educator evaluation process fails to take into account a variety of variables that could affect the test scores and does not reflect the overall academic performance of students or an educator's impact on student learning.

The Keystone Exams were designed to measure student proficiency in relation to the Pennsylvania Standards, which is a requirement by the former No Child Left Behind Law and by the new Every Student Succeeds Act. They were designed to be part of a school evaluation system and serve as a point of reference for all schools across the state to ensure that students were being taught to the same level of rigor no matter where they attended school in Pennsylvania. Results of these exams were to be used by educators to review their curriculum and instructional practices to enhance and improve student achievement. These exams were not designed to be a component of an educator evaluation system. The current weighting given to Keystone and PSSA Exams through the School Performance Profile (SPP) on teacher and principal evaluations is disproportionately high and can significantly skew the performance rating of an individual.

The American Educational Research Association and the American Statistical Association have warned against using test scores to rate individual teachers. There are too many uncontrolled variables, as well as individual differences among students to make these ratings valid. While quality teachers can make a significant impact on student learning, a student's family income and home environment can have just as great an impact on student achievement as the teacher. It is also problematic to use an SPP score as part of an overall evaluation of a teacher that did not teach the three subjects measured by the Keystones. Although 43 states required some use of student test scores as part of their teacher evaluation system systems in 2016, Alaska, North Carolina,

Mississippi, Oklahoma, and Connecticut have all recently moved away from including student test scores as part of their teacher evaluation system with other states such as Massachusetts considering the same.

Penn Manor High School has experienced great success in recent years by focusing on improving academic rigor and the positive results are reflected through a number of measures including our students' performance on the Keystone Exams. However, the Keystones only assess students in Algebra, Literature, and Biology, which leads to evaluative data in that category for educators who teach those subjects. In our high school, only a quarter of approximately 120 teachers teach the subjects assessed by the Keystone Exams. This illustrates one of the primary concerns with using Keystone Exams for evaluating teachers.

The second question asks, “**Does our teacher evaluation system enable schools to recognize and reward high-performing teachers?**” Unfortunately, the current teacher evaluation system makes it difficult to reward high-performing teachers, especially when they work at a school with a low or moderate SPP Score, which is based primarily on the Keystone or PSSA Exams.

Pennsylvania's current framework for educator evaluations was established by the legislature in Act 82 of 2012. Regulations and rating forms were developed by the Pennsylvania Department of Education under PA Code, Title 22, Chapter 19. The current system uses a structure based on the domains and components of the *Danielson Framework for Teaching*, a model developed by education consultant Charlotte Danielson and marketed by the Danielson Group of Princeton, New Jersey. The Danielson framework is used in many other states and is well respected.

The current educator evaluation system makes it difficult to recognize outstanding teachers on one end of the spectrum and also makes it difficult to issue an unsatisfactory rating on the other end of the spectrum. If an outstanding teacher has the unfortunate

circumstance to work in a school with a low SPP Score, it is very difficult for that person to receive a Distinguished Rating, even if they have very high marks on the other components of the evaluation instrument. The SPP Score is weighted at 15% of a teachers overall performance. Even if they receive the highest rating in the other three components of the evaluation instrument known as the PDE 82-1 Form, a low SPP Score can reduce the overall rating from Distinguished down to Proficient. Many building principals and superintendents have expressed frustration over the manner in which the metrics of this instrument use the SPP Score and the unintended impact it has on an outstanding teacher's evaluation.

For example, the Penn Manor School District is the second largest district geographically in Lancaster County encompassing 110 square miles. Our district maintains seven elementary schools that look very different from each other despite teaching the identical curriculum as they serve unique student populations in a variety of areas that can be categorized as urban fringe, suburban, and rural. Letort Elementary School with a low free/reduced population sees very little population change between students who enter in kindergarten and complete the 6th grade while Hambright Elementary School with a free/reduced population of nearly 60% has only a small minority of its 6th graders who have been entirely educated within that building. This directly impacts the PSSA results, and ultimately the School Performance Profile of both schools. I can find an example in which a transfer of an outstanding educator with an overall proficient rating as measured by the current state evaluation system in one Penn Manor school would see her final rating move from proficient to distinguished by teaching students with the same Penn Manor curriculum in another one of the district's elementary schools. This illustrates the challenges we face as superintendents.

Another example I can share is that Penn Manor is seeing an increasing number of parents opt their children out of testing for sincere motives. We are finding that an increasing number of these opt outs are high performing academic students who would generally be assessed as proficient or advanced. The removal of these students has a direct and negative impact on our School Performance Profile as well as negatively

impacting the evaluative results of their teacher from the removal of these students from the testing pool of results.

We know this was not the intent of the legislation. However, these flaws underscore the concerns of using standardized tests results in the evaluation process. PASA is not opposed to using student academic data as part of a teacher evaluation system. However, the data must be derived from multiple sources that reflect the actual work of the teacher as opposed to one score for an entire building that many teachers had no influence in creating. Such data could include, but not be limited to:

- Student learning objectives
- PVASS rolling averages
- IEP achievement data/goals progress
- Other nationally recognized standardized tests
- Industry certification examinations
- Local assessments & other locally-chosen electives

In regard to the third question, “**Are administrators able to use teacher evaluations to identify, improve and remove poor-performing teachers?**” In part yes, and in part no. The Danielson Framework of the Teacher Evaluation System provides ample opportunities for an administrator to identify poor-performing teachers. The professional discussions that are part of the observation process have created more professional dialogue between school administrators and teachers about the teaching and learning process than occurred in the past. This has allowed for quality improvement plans to be implemented for struggling teachers based on research and best practices providing teachers with an opportunity to improve their professional performance.

Unfortunately, the use of the SPP score on the PDE 82-1 Evaluation Form makes it very difficult for a teacher to receive a Needs Improvement rating in a high performing school, and it is almost mathematically impossible in very high-performing schools to rate a teacher as Failing, the only rating which is considered Unsatisfactory when given the first time. Although the SPP Score only accounts for 15% of the overall evaluation, a high SPP score will skew the averaging process to the point where a poor-performing teacher will receive a Needs Improvement rating despite receiving failing scores in actual

classroom teaching. Again, we know this was not the intent of the legislation, but it has become the unfortunate reality. Prior to Act 82 of 2012, approximately 98% of teachers were rated proficient each school year. That percentage has remained unchanged with this new evaluation system due, in large part, to this situation.

One of the most frustrating aspects of the teacher evaluation system is that the “Gross Deficiency” option was eliminated from the previous version. The current mathematical scoring process of the teacher evaluation form practically prevents unsatisfactory ratings from occurring. Adding a “Gross Deficiency” option would allow an administrator to issue an overall unsatisfactory rating if a score of 0 is issued in just one area of the teacher observation component of the evaluation form. A gross deficiency in just one area of a teacher’s performance can be significant enough to warrant immediate action and should not be minimized by the averaging of scores as required in the current evaluation process. This change will enable evaluators to more clearly and easily identify employees with marginal or incompetent performance and immediately set an improvement plan in place or begin dismissal proceedings.

I have recent firsthand experience working with a teacher who was on an intensive instructional support plan due to performance deficiencies in the classroom. In my professional judgment, this teacher should have received a failing rating based on consistent poor performance that directly impacted students through both Domain 1 which is Planning & Preparation and Domain 3 which is instruction. However, due to other evaluative factors, this teacher received ratings of proficient and needs improvement. The teacher eventually resigned but not due to a rating. What prompted the teacher’s departure from the profession was ultimately the result of a labor intensive process of countless hours of observations and meetings involving the professional time of three administrators, the teacher, and a teacher’s union representative. Our district was able to make this process a priority but some school districts do not have the time or resources to make this happen due to other unique demands on administrators.

PASA has worked in conjunction with the Pennsylvania Principals' Association and the Pennsylvania School Board Association to develop a set of recommendations to improve the current educator evaluation instrument. We have attached a copy for your review.

The final question posed was, “**Is the Keystone Exam a quality measure of school accountability?**” The newly authorized Every Student Succeeds Act provides states with flexibility in how it evaluates schools and provides many options for states to use in their accountability plan. Standardized test scores are required for each state to administer to all students in grades 3-8 and once in high school as defined in the new ESSA law. They can serve as a piece of the overall performance picture of a school, but should not be used as the sole indicator of a school's effectiveness. The Keystone Exams can be used as one piece of school performance as they have been designed to measure student acquisition the state standards and can serve as an analysis tool for teachers and administrators.

One standardized exam should not be the primary measure used to determine the effectiveness of a school. Currently, the data from the Keystone Exams accounts for 90% of a high school's SPP Score. We believe this is an inappropriate use of these exams to determine the overall effectiveness of a school. There are a variety of other data sets that are readily available to incorporate into a school accountability system that should be given consideration at the high school level. These measures include standardized test data from multiple sources such as the SAT, ACT, NOCTI, NIMS, IB, and AP Exams. Parents and the public need to see a more complete picture of a school's performance as a result of viewing a variety of data as opposed to only viewing the results of one state mandated test.

We understand the Senate is reviewing the possibility of using the SAT Suite of Exams as the ESSA accountability test for high school students. This would present a shift in emphasis from measuring the attainment of academic standards to assessing college and career readiness for students. As presented by the College Board at the June 2 Senate Education Hearing at West Chester University, the SAT Suite of Exams has many positive applications that could prove to be very useful for students, parents, and

educators. Moving to the SAT Exam Suite would highlight the need to have an accountability system that uses multiple data sources to evaluate a school as each standardized exam provides a unique measure of learning based on its design. No matter which exam is used for federal accountability, PASA strongly recommends that these exams not be used to determine high school graduation.

Thank you for the opportunity to share our thoughts with you this morning. We would be happy to answer any questions that you have.