

Recommendations for Revising the Educator Evaluation System *(Act 82 of 2012)*

**Presented by the Pennsylvania Association of School Administrators (PASA),
Pennsylvania Principals Association, and the Pennsylvania School Boards Association (PSBA)**

Why modify? A revised system can provide enhanced opportunity for educators and administrators to have constructive dialogue about areas of success as well as areas where improvement is needed to ensure that every child has an effective teacher in the classroom. Changes to Act 82 can provide clarity in properly identifying high-performing and low-performing teachers, and reduce delays in taking action on a personnel matter.

With the enactment of the federal Every Student Succeeds Act (ESSA), Pennsylvania has an opportunity to revisit the structure and administration of its current method for evaluating educator effectiveness. The ESSA removes federal oversight of how educators are evaluated, allowing states to consider different approaches and weightings to their rating requirements and tools.

Pennsylvania's current framework for educator evaluations is established in legislation under Act 82 of 2012, and regulations and rating forms 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.

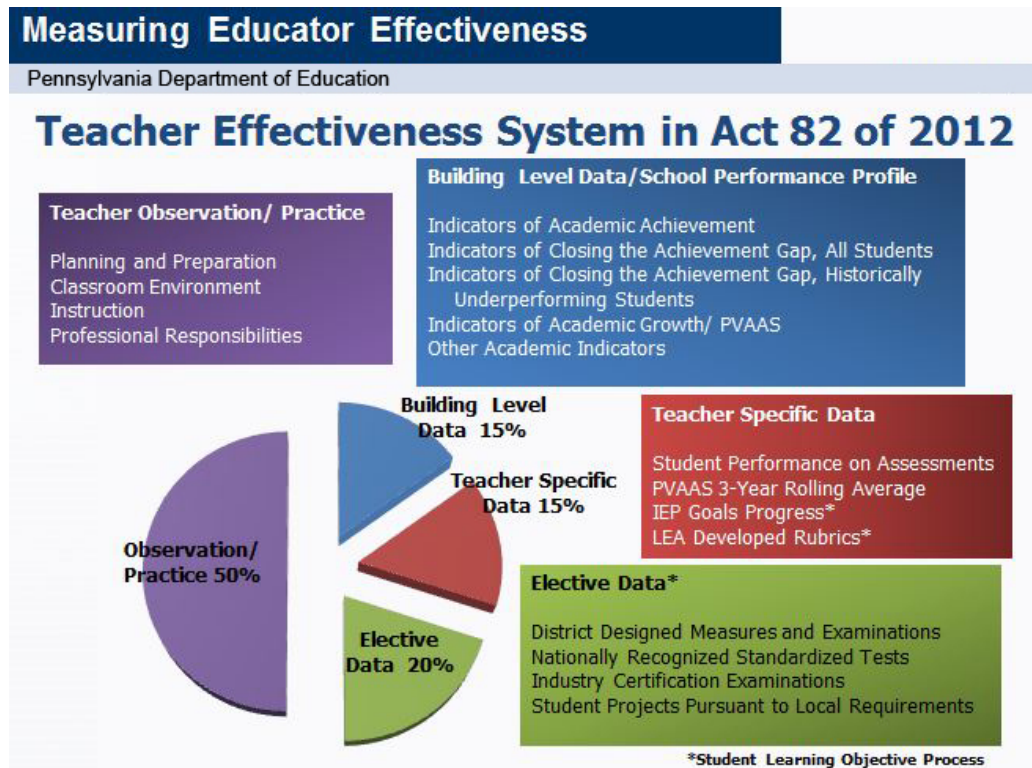
In conjunction with this effort, PDE also developed a system for rating the academic achievement data of public schools that is used in the evaluation system. The PA School Performance Profile (SPP) includes performance data for school districts, school buildings, comprehensive career and technology centers, and charter and cyber charter schools. The profiles on the SPP website (<http://paschoolperformance.org>) contain the building-level data that is a component of evaluation structure.

Positive outcomes under Act 82

Act 82 replaced an outdated structure that was criticized for not being deeply effective and did little to provide meaningful feedback to assist educators. Under the new system created by Act 82, critical conversations are taking place between teachers and administrators on planning and preparation, classroom environment, instruction and professional responsibilities. Teachers create student learning objectives. Discussions are targeted to both experienced and new teachers encourage actionable feedback. Struggling teachers receive guidance from administrators through the use of jointly-developed performance improvement plans. And from these critical conversations, school districts can create professional development activities that better meet the needs of all their teachers. Every student deserves an effective teacher, and every teacher deserves adequate support from school leaders in helping him or her promote improved student learning.

Understanding the current system

The chart explains the state's rating current system for educator effectiveness created under Act 82 and the Chapter 19 regulations that describes the 50%-50% split of evaluation measures.



As shown by the left side of the pie chart, at least 50% of the annual evaluation is based on observation and practice. The rating tool includes descriptions of the four domains set forth in Act 82 for teacher observation and practice. The four domains are as follows: planning and preparation (20% of the overall score); classroom environment (30%); instruction (30%); and professional responsibilities (20%). The Chapter 19 regulations provide descriptions of educator performance or behavior in the four domains.

The right side of the chart shows that the other 50% of the evaluation is based on three measures of student performance. The measures are: building-level rating using the SPP score (15%); teacher specific data that consists of measures based on student performance on assessments and value-added assessment system data (15%); and elective data which may include measures of student performance selected from a list provided annually by PDE. LEAs must select and develop measures using a Student Learning Objective process (20%).

For nonteaching professional employees, the system requires 80% of the overall rating to be based on observation and practice, and 20% on student performance. For principals, the system requires 50% of the overall rating to be based on leadership observation and practice (50%); building-level rating (15%); correlation rating (15%) and elective rating (20%).

The system requires evaluators to use specific forms to calculate the performance rating of the individual educator. The form is used to record the results of the data collection process which provides for a potential overall rating of Failing, Needs Improvement, Proficient or Distinguished. The rating form sets numeric values for these four rating levels on a 0-3 point scale, and the regulations provide descriptions of performance or behavior for each category. Based on this

performance rating, the employee then receives a final rating of Satisfactory or Unsatisfactory. When an employee receives a Needs Improvement rating twice within a 10-year period, the overall rating of the employee will be considered Unsatisfactory. An employee cannot receive a Failing rating based solely on test scores, and an employee who receives a Needs Improvement or Failing rating must participate in a performance improvement plan.

Concerns and unintended consequences

Implementation of the system has presented some challenges and unintended consequences for administrators and educators. Some modifications can improve the validity and purpose of the rating methods used. Changes can also provide clarity and reduce delays in taking action on a personnel matter, including dismissal for incompetency or unsatisfactory performance. Here are the concerns and unintended consequences that can be addressed:

- **Despite changes made to address criticism of the previous system, the vast majority of teachers remain rated Satisfactory.**

The previous system was criticized because it used only observation and provided only Satisfactory or Unsatisfactory rating choices, resulting in about 98% of teachers rated Satisfactory. It did not provide the depth needed to appropriately rate the effectiveness of an employee or take action to terminate an employee, if necessary. The new system under Act 82 was intended to create a more comprehensive and consistent evaluation that would provide better feedback to educators and school leaders using various measures to reflect performance. And while the new process uses four performance categories and is more complex in its mathematical calculations, the result is that the system still uses a final rating of either Satisfactory or Unsatisfactory, with about 98% of teachers still rated Satisfactory.

- **The mathematical calculations in the four domains for overall performance contain statistical weaknesses that make it harder to address meritorious or underperforming performance in a fair and timely manner. Also, the heavy weight of building-level data (SPP scores) can lead to unfair or inaccurate evaluation ratings.**

Use of building-level data in a low-performing school makes it difficult to differentiate and recognize excellent professional employees, and make it impossible for an excellent teacher to receive a rating of Distinguished. In fact, it could cause the teacher to receive a low score that could result in a Needs Improvement rating. On the other side, the use of SPP scores makes it too difficult for a teacher to receive a Needs Improvement rating in a high performing school. Further, it is mathematically impossible in very high-performing buildings to rate a teacher as failing, the only rating which is considered Unsatisfactory when given the first time.

For example, if you just look at the observation piece combined with SPP, a teacher receiving a 0 in planning and preparation, a 1 in classroom environment, a 0 in instruction, and a 0 in professional practice combined with an SPP score of 86 or above still does not render this evaluation as Unsatisfactory but a Needs Improvement rating. On the flip side, a teacher who receives all three ratings on the observation side and a perfect 100 SPP score, still cannot receive a Distinguished rating.

- **It is difficult to terminate an employee for Unsatisfactory performance.**

The current system has resulted in significant delay in terminating professional employees when the only basis for this is Unsatisfactory performance. Under the School Code, only a Failing rating is Unsatisfactory for purposes of termination (24 P.S. §11-1122). A tenured employee needs two consecutive Unsatisfactory ratings with a performance improvement plan of at least four months in place in order to be discharged. Those temporary professional employees (TPEs) receiving a Needs Improvement rating cannot be separated for Unsatisfactory performance unless a second Needs

Improvement rating is issued before tenure is granted. In addition, the current system limits evaluation to instructional practices and does not take into account other employee behaviors, such as absenteeism, conduct with peers, professional image, etc.

- **There are concerns with using a heavy weighting to student test scores and value-added data as a factor in evaluations.**

Test scores are often influenced by the socio-economic conditions of the community and the school district, with poverty and other factors leading to lower test scores. Inequities in school funding translate into varying resources for schools. Lower test scores translate to lower school building-level scores on the SPP ranking. Student, family and community demographics also impact assessment outcomes. Low building-level data and SPP scores are indicators beyond the control or influence of the teacher and are not really an accurate indicator of his or her performance in the classroom.

In addition, cautions have been raised regarding the use of value-added measures (VAMs) for high-stakes decisions related to teachers. The American Statistical Association issued a statement in 2014 noting that because VAMs are generally based on test scores, they do not directly measure potential teacher contributions toward other student outcomes. VAM scores and rankings can change substantially when a different test or model is used. The statement also notes that using multiple years of data does not help problems caused when a model systematically undervalues teachers who work in specific contexts or with specific types of students, since that systematic undervaluation would be present in every year of data. A VAM score may provide teachers and administrators with information on their students' performance and identify areas where improvement is needed, but it does not provide information on how to improve the teaching.

- **Use of SPP scores delays the evaluation process.**

School districts only get the necessary building-level data in October, which result in evaluations being completed in the school year following the evaluation period.

- **Use of data, SPP scores raises questions regarding temporary professional employees (TPEs) or other employees who need midyear review.**

The use of data and SPP scores for the evaluations of temporary professional employees, teachers with one to three years of experience who do not yet have tenure, raises questions because they are required to be evaluated twice a year. Teachers will challenge ratings that are based on older data. How does this impact the midyear evaluation for TPEs? What is the impact of being unable to do an end of year evaluation for TPEs who complete a third year of teaching with no evaluation given until October of their fourth year of teaching if that October evaluation is an Unsatisfactory rating? Will it be necessary to focus more intensely on the TPE's second year of service again to avoid the automatic transition to tenure? What if a tenured employee is put on a performance improvement plan as a result of a needs improvement evaluation? May the district evaluate this person again after a four-month period even if new data is not available? This, too, will be challenged.

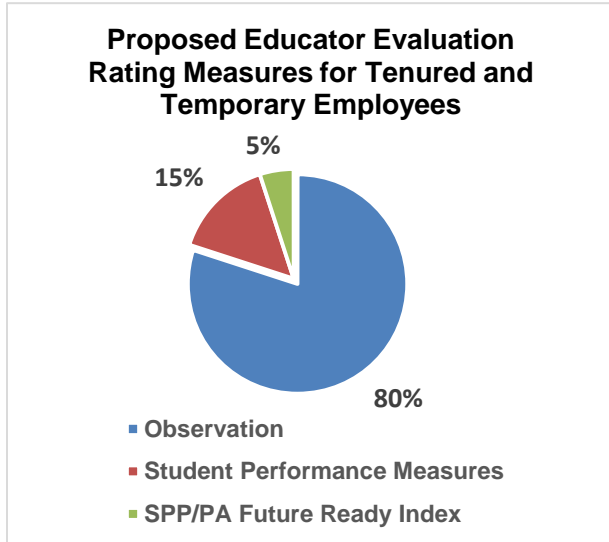
Recommendations for change

PASA, the PA Principals Association and PSBA are in favor of retaining a model using the Danielson or similar type of framework and believe that the observation piece of the evaluation is the most important. The observation piece provides opportunities for evaluation in four key areas, and educators report rich conversations and professional growth based on these components.

The organizations jointly offer these six recommendations for change:

Recommendation 1: Revise the weighting so that the observation/practice component counts for 80% of the evaluation for all professional employees (tenured and temporary), 15% of the rating uses student performance measures, and 5% uses SPP/Future Ready Index.

- **For the 80%:** Maintain the four existing domains for the teacher observation and practice that comprises the 80% of the rating.



- **For the 15%:** Revise the requirements for calculating the 20% by allowing districts to select measures from a list of options approved by PDE. The list should be flexible, and PDE should be permitted to modify/update these options as needed or desired. Finally, districts should be allowed to determine the weighting applied to each option. Electives could include, but be limited to, these types of measures:

- 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

- **For the 5%:** Use the SPP/Future Ready Index that contains building level indicators of student achievement and school performance.

Recommendation 2: Temporary professional employees should also be rated using the 80%-15%-5% split using the same rubric as professional employees. In addition:

- The mandated use of a Performance Improvement Plan currently required for an employee who receives an overall performance rating of Needs Improvement or Failing should be maintained for professional employees but not be mandatory for TPEs.
- The Performance Improvement Plan should be allowed to be used at the discretion of the employer for TPEs, since educators new to the profession need time to develop their skills. This allows supervisors to guide the development of TPEs without creating anxiety about potential dismissal issues.

Recommendation 3: Add “Gross Deficiency” to the observation rating scale and define it as a “0” in any category. The current mathematical scoring prevents Unsatisfactory ratings and this change will enable evaluators to more clearly and easily identify employees with marginal or incompetent performance. Clarify that a “Gross Deficiency” or a “0” in the observation rating may result in an overall rating of Unsatisfactory.

Recommendation 4: Reduce the current 10-year time period in considering when two Needs Improvement ratings convert to an overall rating of Unsatisfactory. Ten years is an extended time to keep and track ratings, and could be considered too punitive. A reduction to no less than five years is reasonable.

Recommendation 5: The state’s evaluation system should be applied equally to educators and principals in all public school entities, including those in charter and cyber charter schools.

Recommendation 6: Any system that is implemented in our schools needs to be constantly evaluated. Act 82 does not allow for changes to the system to be made and it is only under ESSA that we have a window of opportunity to make needed changes. The state's educator evaluator system should be under the control of the State Board of Education so that further changes, if needed, can be made without legislative approval.