PENNSYLVANIA STATE EDUCATION ASSOCIATION

Testimony of the

Pennsylvania State Education Association (PSEA)

Public Hearing Regarding

House Bill 1980

Presented to the

House Education Committee

November 3, 2011

By

Linda J. Cook
PSEA Board of Directors

Good afternoon, Chairman Clymer and Chairman Roebuck. My name is Linda Cook. I am a teacher in the Penn Delco School District and a member of the Pennsylvania State Education Association (PSEA) Board of Directors. PSEA represents 193,000 teachers, educational support personnel, and other individuals charged with the responsibility of educating Pennsylvania's students. On behalf of our members, thank you for inviting PSEA to speak today about House Bill 1980 and the importance of valid, fair, and effective teacher evaluation.

PSEA understands the importance of an evaluation system that defines high standards of professional practice, removes persistently ineffective teachers, and supports the professional growth of all teachers. As a matter of fact, we have advocated for several years for improvements to the current system; PSEA has participated in the Gates Momentum Statewide Stakeholder Committee, and we continue to advise Pennsylvania's Department of Education (PDE) on the new system as part of the Committee of Practitioners. We also encouraged our members to participate in both teacher evaluation pilots. Our members want a truly effective evaluation system – one that will help them fully demonstrate their professional practice and find ways to improve upon it.

House Bill 1980 calls for a new framework for Pennsylvania's teacher and principal evaluation system and an ambitious schedule for designing and implementing a new system that fits that framework. This is problematic for several reasons:

- It risks requiring the use of evidence of teacher and principal effectiveness that is neither valid nor reliable;
- It creates a framework that does not clearly connect an individual's performance evaluation to their own professional practice;
- And it calls for the development of a new system before results of the pilot designed to inform that development can be considered.

At its core, an effective evaluation system needs to be based on evidence that is strong, fair, and leads to valid conclusions about job performance that will help individuals improve their professional practice. There are many reasons why House Bill 1980, in its use of standardized tests to calculate at least 50 percent of a classroom teacher's overall evaluation, will not produce valid or fair conclusions about teachers' job performance or help teachers improve their professional practice. In the interest of time, I will mention four.

First, research makes clear that several factors besides the individual teacher influences student achievement and growth, including past teachers, tutors, curriculum quality, class size, student attendance, out-of-school learning (including summer learning), family resources, birth weight, medical care, food insecurity, student motivation, and peer influence. As a matter of fact, recent research concludes that student factors explain more variation in student achievement than

teacher factors.[1] It should come as no surprise, then, that teachers receive lower "effectiveness" scores on standardized tests when their students are disproportionately English learners, low-income, or special education students – even when statistical methods try to adjust for differences in students and teachers.[2] Students are never assigned randomly across teachers, and so student characteristics benefit some teachers and penalize others in a test-based evaluation system. Not only do student factors reduce the validity of standardized test scores as a measure of teacher effectiveness, but because standardized test scores primarily reflect variation in student factors rather than teachers, most teachers cannot use these results in any meaningful way to improve their own professional practice.

Second, an individual teacher's influence on student achievement and growth may span many classrooms, schools, and even districts. For example, many elementary school reading programs assign students to different reading teachers every eight weeks. In a special education program, one elementary student may work with four or five different specialists that each help the student in varying degrees of intensity. Some teachers are assigned to multiple schools within a district, and in the case of educators who work at intermediate units, assignments may span multiple districts. There are no models to accurately reflect the complex ways in which teachers are assigned to help different students, in different schools or districts, for different amounts of time. Because there is no way to calculate the actual "dosage" that one student has of one teacher, House Bill 1980 will, to an incalculable extent, evaluate teachers on others' work. This cannot lead to valid conclusions about individual teachers' job performance and cannot give teachers the information they need to improve their own professional practice.

Third, students grow every day and in many ways – they develop social skills, emotional maturity, behavioral self-control, speech and language abilities, fine and gross motor skills, and, of course, cognition and knowledge across the complete curriculum. Any comprehensive system to measure teachers' impact on student growth needs to take account of each of the ways students develop. Unfortunately, the bill relies on as little as one annual standardized test of specific academic content to calculate 50 percent or more of a teacher's overall evaluation.[3] State assessments are designed to measure a subset of grade-level content standards, not to measure all of the ways teachers help students grow. Teachers need real-time measures of all the ways their students are growing in order to analyze their own professional practice.

Finally, most standardized tests are not valid assessments of learning for students with special needs, including English Language Learners, some students with disabilities, and gifted students. The reasons for this vary, but in general, one-size-fits-all tests are not designed to measure achievement for students who are working significantly above or below grade level or students or who are not able to fully understand a test in English. This reduces the validity of the assessment and distracts teachers from the important work of improving their professional practice to meet the needs of all students.

Overall, using standardized tests as 50 percent or more of a teacher's evaluation will not produce evidence of teacher effectiveness that is strong or fair.[4] Nor will it lead to valid conclusions about a teacher's job performance or improvements in professional practice.[5]

Of equal concern to PSEA, however, is a growing body of evidence demonstrating that the high-stakes use of test scores in teacher evaluations is likely to have negative consequences, not only for individual teachers, but for the education system as a whole. Donald Campbell's[6] assertion that relying on a quantitative measure for high stakes decisions will distort the measure and corrupt the system has become so accepted in social science that it is now simply referred to as "Campbell's Law." Applied to education,[7] Campbell concluded, "When test scores become the goal of the teaching process, they both lose their value as indicators of educational status and distort the educational process in undesirable ways (p. 57)." [8]

Ultimately, an effective evaluation system needs to hold teachers accountable for their professional practice. Holding teachers accountable for outcomes that are out of their control is likely to exacerbate bad professional practices, not reduce them. According to Campbell's Law, we will distort the system we are trying to measure if we base half of a teacher's evaluation on standardized test results.

While there are many parts of the education system that teachers do not control, they do control their professional practice, and a high-quality teacher evaluation system should help all teachers know how to change their practice for the better. The most direct way to help an individual improve their practice is to focus on that practice in the evaluation system. This does not mean that student outcomes do not have a role to play in teacher evaluation, but it means that student outcomes are only meaningful to the extent that we can relate specific observable professional practices to specific student outcomes. Standardized tests simply are not able to provide this kind of data.

The good news is that there ***are*** ways to measure student outcomes that can also help teachers improve their professional practice. Teachers measure student growth all the times, through curriculum-based pre- and post-tests; student projects and performances; specially designed assessments relating to IEP goals; inventories of behavioral, social, and other kinds of skills; formative assessments, and other measures. These kinds of measures happen at the point of instruction and are based on instruction, so a teacher can more directly see a link between his or her teaching and these kinds of student results. Even more importantly, teachers can change their practice immediately and differentiate instruction based upon the results to achieve different outcomes for students. Teachers appreciate the value of these measures, and it is how they know that their students are learning.

It is, we think, possible to grow an evaluation system that relies upon these kinds of measures of student outcomes – that is one of the reasons we have participated in the Gates Momentum Study and have encouraged our members to participate in PDE's two teacher and principal evaluation

pilots. But, the Gates contractors are still analyzing the data from the first pilot of only five local education agencies, and the Commonwealth is just beginning to roll out the second pilot involving more than 100 LEAs. Results from it likely will not be available until after the deadline in House Bill 1980 for adopting the new system.

It is also possible to grow a system that adheres to all of the components of effective evaluation. PSEA consistently advocates for systems that reflect current research. We know what high-quality evaluation looks like. It should: (1) Be based upon a challenging definition of good practice;[9] (2) Be sufficiently flexible to allow managers to relate broad standards to local organizational goals;[10] (3) Set different expectations for professionals at different career stages;[11] (4) Provide ongoing training for evaluators and those being evaluated;[12] (5) Include an understanding of organizational supports and barriers to effective job performance;[13] (6) Maximize employee engagement, self-appraisal, and feedback;[14] (7) Be based on strong, diverse evidence (including self-produced evidence) that leads to valid conclusions;[15] (8) Link to professional development;[16] and (9) Hold individuals responsible for their own professional practice.

No one benefits when an evaluation system does not do its job, and our concern is that current proposals simply replace one ineffective system with another. But we can build a teacher evaluation system that removes persistently ineffective teachers and supports the professional growth of all teachers. To do so, we need to take advantage of decades of research about effective evaluation and the appropriate use of standardized assessments. We also need to apply research that defines elements of effective evaluation.

PSEA envisions a system that adheres to principles of effective evaluation, produces valid results, incorporates meaningful measures of student growth, and uses measures of professional practice as the basis of an evaluation result. In that regard, we believe that prudent policymaking requires that the legislature allow the Gates Momentum Project to continue its important work and report its findings ***before*** the Legislature acts. We stand ready to help the Commonwealth in producing a system that will apply the best research in evaluation and lessons learned from the pilot project to set and maintain high standards for all teachers.

Thank you for this opportunity, and I look forward to your questions.

[1] See, for example, Berliner, David C. (2009). *Poverty and Potential: Out-of-School Factors and School Success.* Boulder and Tempe: Education and the Public Interest Center & Education Policy Research Unit. Retrieved October 31, 2011 from http://epicpolicy.org/publication/poverty-and-potential, See also Rothstein, R. (2004). *Class and schools: Using social, economic, and educational reform to close the black-white achievement gap.* Washington, D.C.: Economic Policy Institute; Schochet, P. Z and Chiang, H.S. (2010). "Error rates in measuring teacher and school performance based on student test score gains." (NCEE 2010-4004) Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

[2] Baker, E.L. et al. (2010). Problems with the Use of Student Test Scores to Evaluate Teachers. Briefing Paper. Washington, DC: Economic Policy Institute; Newton, X.A., Darling-Hammond, L., Haertel, E., and Thomas, E. (2010). Value-Added Modeling of Teacher Effectiveness: An exploration of stability across models and context. Education Policy Analysis Archives, 18(23). Retrieved on October 31, 2011 from http://epaa.asu.edu/ojs/article/view/810/858.

[3] PSEA would note that the PSSA tests have never been validated as a measure of teachers' performance.

[4] Researchers have found that different standardized tests of the same subject will produce substantially different value-added scores and rankings for the same group of teachers (Papay, J. (2011). "Different tests, Different Answers: The stability of teacher value-added estimates across outcome measures." *American Educational Research Journal.* Vol. 48, No.1, pp.163-193). According to a review of the Gates Foundation-funded Measuring Effective Teaching (MET) study, more than 40% of teachers who placed in the bottom quarter using one test (the state test) placed in the top half when measured by their students' performance on an alternative standardized test in the same subject (Rothstein, J. (2011). *Review of "Learning About Teaching: Initial Findings from the Measures of Effective Teaching Project."* Boulder, CO: National Education Policy Center. Retrieved [date] from http://nepc.colorado.edu/thinktank/review-learning-about-teaching.). Rothstein concluded "teacher ratings based on the state assessment were only slightly better than a coin toss for identifying which teachers did well using the alternative assessment."

[5] In 2010, the Board on Testing and Assessment of the National Research Council of the National Academy of Sciences concluded: V[alue] A[dded] M[easurement] estimates of teacher effectiveness should not be used as the ***sole or primary*** basis for making operational decisions because the extent to which the measures reflect the contribution of teachers themselves, rather than other factors, is not understood. (National Research Council and National Academy of Education. (2010). *Getting Value Out of Value-Added: Report of a Workshop.* Committee on Value-Added Methodology for Instructional Improvement, Program Evaluation, and Educational Accountability, Henry Braun, Naomi Chudowsky, and Judith Koenig, Editors. Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.)

[6] Campbell, D.T. (1976). Assessing the Impact of Planned Social Change. Occasional Paper Series #8. Kalamazoo, MI: Western Michigan University, p. 54. Retrieved on October 31, 2011, http://www.eric.ed.gov/PDFS/ED303512.pdf.

[7] Over the last ten years, researchers have documented a number of negative consequences associated with assigning high-stakes to standardized test scores, including narrowing the curriculum, lowering standards, and increasing dropout rates. Researchers have also documented that when a standardized assessment is high-stakes, student achievement rises on that specific test, but not on other similar measures of student knowledge; in other words, instruction narrows and learning stagnates, even though test scores rise.

[8] Campbell, D.T. (1976). Assessing the Impact of Planned Social Change. Occasional Paper Series #8. Kalamazoo, MI: Western Michigan University. Retrieved on October 31, 2011, http://www.eric.ed.gov/PDFS/ED303512.pdf.

[9] Hassel, B. C. (2002). Better pay for better teaching: Making teacher pay pay off in the age of accountability. PPI Policy Report. Washington, DC: Progressive Policy Institute; Milanowski, A.T., Kimball, S.M., and White, B. (2004). The Relationship Between Standards-Based Teacher Evaluation Scores and Student Achievement: Replication and Extensions at Three Sites. CPRE-UW Working Paper Series. TC-04-01. Madison, WI: Center for Policy Research in Education. http://cpre.wceruw.org/papers/3site_long_TE_SA_AERA04TE.pdf

[10] Rynes, S., Brown, K., & Colbert, A. 2002. Seven common misconceptions about human resource practices: Research findings versus practitioner beliefs. *Academy of Management Executive, 16*(3): 92–102.

[11] Cron, W. L. (1984). Industrial Salesperson Development : A Career Stage Perspective. *Journal of Marketing*, 48, Fall, pp.41-52;Dalton, G., Thompson, P., and Price (1977). The four stages of professional careers: a new look at performance by professionals. *Organizational Dynamics.*, 6, Number 1, 23; Dalton, G., and Thompson, P. (1986). *Novations: Strategies for Career Management.* Glenview, IL: Scott Foresman and Company.

[12] Longenecker, C.O. and Fink, L.S. (1999). Creating Effective Performance Appraisals. Industrial Management, September-October, pp. 18-23.

[13] Marx, G.E. (2007). Research Brief: Teacher Evaluation. Eastern Michigan University: Principals' Partnership. Available online: http://www.principalspartnership.com/Teacherevaluation.pdf Schein, E. H. (1980). Organizational Psychology (3rd ed.). Englewood Cliffs, New Jersey: Prentice-Hall, Inc.; Super, D. E. (1984). Career and Life Development. In D, Brown, L. Brooks (Eds.), Career Choice and Development (2nd ed., pp. 422-454). San Francisco: Jossey-Bass; Joint Committee on Standards for Educational Evaluation (2009). The Personnel Evaluation Standards. Thousand Oaks, CA: Corwin Press.

[14] Roberts, Gary E. 2002. Employee Performance Appraisal System Participation: A technique that works. Public Personnel. Management, Fall. Available online: http://www.entrepreneur.com/tradejournals/article/160542351_1.html; Cotton, J. L. (1993). Employee Involvement: Methods for improving performance and work attitudes. Thousand Oaks, CA: Sage Publications; Latham G. P. and K. N. Wexley (1981). Increasing Productivity Through Performance Appraisal. Reading, Mass.: Addison-Wesley; Jordan, J. L. (1990). "Performance Appraisal Satisfaction and Supervisors' Traits," Psychological Reports, volume 66, 1337-1338; Daley, D. (1992). Performance Appraisal in the Public Sector. Westport, Connecticut: Quorum Books; Tjosvold, D. and J. A. Halco (1992). "Performance Appraisal of Managers: Goal Interdependence, Ratings and Outcomes," Journal of Social Psychology, volume 132, 629-639; Roberts, G. E. (1992). "Linkages Between Performance Appraisal System Effectiveness and Rater and Ratee Acceptance: Evidence from a Survey of Municipal Personnel Administrators," Review of Public Personnel Administration, volume 12, 19-41.

[15] Joint Committee on Standards for Educational Evaluation (2009). The Personnel Evaluation Standards. Thousand Oaks, CA: Corwin Press.

[16] Wenglinsky, H. (2002). The link between teacher classroom practices and student academic performance. Education Policy Analysis Archives, 10(12). Available online: http://epaa.asu.edu/epaa/v10n12/; Killion, J. (2002). "What Works in the High School:

Results-based staff development." Oxford, OH: National Staff Development Council; Alliance for Excellent Education (2008). Measuring and Improving the Effectiveness of High School Teachers. Washington, DC: Alliance for Excellent Education. Available online: http://www.all4ed.org/files/TeacherEffectiveness.pdf; Supovitz, J.A., and J. B. Christman. (2003). Developing communities of instructional practice. Philadelphia: Consortium for Policy Research in Education.