

“Value-Added Accountability”

Testimony Submitted to the Pennsylvania Senate Education Committee

March 16, 2004

Theodore Hershberg
Professor, Public Policy and History and Director, Operation Public Education
University of Pennsylvania

I am honored by the invitation to present this testimony to the Senate Education Committee. I would like to take this opportunity to make two basic points:

First, value-added assessment can serve as the foundation for a system of accountability that can be used at the level of classrooms, school and districts. This accountability system was developed by Operation Public Education (OPE), a foundation- and corporate-funded school reform effort based at the University of Pennsylvania. Originally introduced in the Pennsylvania House of Representatives in 2002 by Rep. Nicholas Micozzie, it was recently revised to incorporate an expanded role for teachers and teachers’ unions (see Appendix 1). With the assistance of a distinguished group of educators who have joined our national advisory board (Appendix 2), OPE is now introducing this accountability system to key education stakeholders nationwide.

Second, because neither of Pennsylvania’s teachers’ unions (PSEA or PaFT) nor the Rendell administration wishes to support this approach to accountability at this time, we believe that for now it makes sense to pursue the course set out by Education Secretary Vicki Phillips. Among the key characteristics of this approach are (1) deploying value-added assessment solely as a diagnostic tool; (2) introducing a career ladder for teachers based on skill and knowledge; and (3) using the ladder to differentiate how teachers are compensated.

Before moving on to discuss the key elements in this accountability system, a few other important points about value-added assessment need to be made.

Value-added Assessment

Value-added assessment is a statistical method of analyzing test data that isolates the impact of instruction on student learning. It is often confused with a gain score — the growth that a student makes in a subject from one year to the next. But it is actually far more powerful than a simple gain score because it separates the annual growth into two parts: that which can be attributed to the teacher and that which can be attributed to the student. It is this quality that makes it such a powerful diagnostic tool and permits it to serve as the basis for individual-level accountability.

Where value-added is being used

These qualities plus the annual testing requirement of students in No Child Left Behind have catalyzed the rapid adoption of value-added models across the nation. Although there are several different value-added models available, the only one being used on a statewide basis to date is the EVAAS method developed for Tennessee by Dr. William Sanders in 1992.

In September 2002, the Pennsylvania Board of Education adopted it (PVAAS) for diagnostic use in all school districts, and the State Department of Education is now engaged in a rollout that will bring value-added to all schools by 2005-06. Ohio passed legislation last June that requires value-added to be part of their school performance index no later than 2007; currently over 60 Ohio school districts are participating in value-added pilot being run by the Battelle for Kids Foundation. Most recently the Ohio Board of Regents, the State Department of Education and 51 colleges of education joined together in the “Ohio Partnership for Accountability” — a five-year, \$10 million study to use value-added to measure the instructional impact of new teachers in their classrooms as a means of improving the quality of teacher-preparation programs in higher education.

Minnesota recently passed a law calling for a value-added pilot by 2007. The Iowa School Boards Association has funded a value-added pilot, and the New York State School Boards Association is now preparing to launch a similar pilot. The New York City schools are considering a value-added pilot of their own, and last week at the invitation of Randi Weingarten, president of the United Federation of Teachers in New York City, I briefed the senior leadership of the UFT on how value-added works. Arkansas recently passed legislation that mandated value-added assessment for use in its schools, but they have not yet decided on which model they will adopt.

Recent reports on value-added assessment

My colleagues and I have recently published an essay that describes value-added assessment, included here as Appendix 3: Ted Hershberg, Virginia Adams-Simon and Barbara Lea-Kruger, “Measuring What Matters: How value-added assessment can be used to drive learning gains,” *American School Boards Journal* (February, 2004).

The Education Trust recently published a report on value-added: “The Real Value of Teachers: Using New Information about Teacher Effectiveness to Close the Achievement Gap,” (Winter, 2004). The report provides compelling examples of why value-added should be used in our schools and lists its many benefits for states, higher education, districts, administrators and teachers.

A careful statistical review of value-added models (VAM) was recently published by RAND: Daniel F. McCaffrey, J. R. Lockwood, Daniel M. Koretz, Laura S. Hamilton, *Evaluating Value-Added Models for Teacher Accountability* (2003). “VAM,” the authors explain, “refers to a diverse class of models with many possible paths to estimates.” As in any complex statistical system, RAND identifies the various ways in which error can be introduced in the use of VAM — missing information, testing error, longitudinal data, the estimation of effects — but their key conclusions are summarized by Dan Fallon, Chief of the Carnegie Corporation’s Education Division that funded the study: “(1) there is a teacher effect; (2) it could be large; and (3) it persists for years beyond the year in which it is first evident.”

Several other summary observations in the RAND report are worthy of your attention:

First, “VAM might actually provide less-biased and more precise assessments of teacher effects” than current approaches (p. 120). Second, “If test-based accountability remains an instrument of education policy, we recommend that, as policymakers evaluate alternative models for school or teacher accountability, VAM should be given serious consideration even in light of its limitations” (p. 120).

Finally, when value-added is used as part of an accountability system, the key consideration is how a particular system uses a specific form of value-added. The RAND report raises serious concerns about using value-added to rank individual teachers on the basis of their value-added scores and evaluate or reward them on a proportional basis. However, the report states “that estimates might be sufficiently precise for other inferences, such as identifying some teachers as distinct from the mean. In one small example, we estimated that about one-third of teachers had a very small probability of being equal to the average teacher, which suggests that, for some applications, sampling error in the estimates will not preclude identifying a fraction of teachers as above or below average” (see Summary, xviii-xix).

That is why the OPE approach assigns instruction to three categories: *effective* (students receive a year’s worth of growth in a year), *highly effective* (student growth significantly exceeds a year’s worth of growth in a year) and *ineffective* (student growth significantly falls below a year’s worth of growth in a year).

While acknowledging that “the application of VAM to educational achievement holds considerable promise” (p. xi), the RAND authors conclude that “the research base is currently insufficient to support the use of VAM for high-stakes decisions” (p. xx). We have always contended that no educator or student should ever be evaluated solely on the basis of a single measure, not even one as powerful as value-added. That is why the OPE approach to accountability uses *multiple measures* in the evaluation of educators. Half of a teacher’s evaluation in the OPE system, for example, is based on the expertise they demonstrate in the four domains developed by the Association for Supervision and Curriculum Development (ASCD): Planning and preparation; Classroom environment; Instruction; and Professional responsibilities.

Value-added and NCLB: complementary not contradictory

Is value-added incompatible with NCLB? Despite the contention of some observers, this is simply not true. Adequate Yearly Progress (AYP) measures absolute performance while value-added measures progress in performance. When value-added is included in the tool kit of educators, they are in a vastly improved position to understand the reasons that underlie differing levels of absolute performance (See Appendix 4).

Value-added enjoys other advantages over AYP measures. First, it traces individual students over time while AYPs use cohort comparisons — this year’s fourth grade is compared to last year’s fourth grade. Cohorts do not allow you to see if schools are having an impact on the *same students over time*. A school could be on the warning list one year and successful the next, largely because a different group of children is being tested. Second, because value-added measures student growth, schools that may have missed their absolute targets but are clearly making some progress, can be distinguished from those making no progress. Third, value-added maintains a focus on *all* children. Under NCLB, the inevitable focus of schools is on the subset of students whose improvement will satisfy their AYP goals. Thus students performing above proficiency and those far below it are more likely to be ignored in the process. Remember, all children deserve a year’s worth of growth in a year, not simply those who are most likely to help a school meet its AYP objectives.

The point of noting these value-added advantages is that all operate to help educators meet their AYP goals.

Value-added Accountability in Pennsylvania

Use value-added for diagnostics until educators become familiar with its operation

Since value-added assessment cannot be used without at least three years of data, and since close to half of all the state's school districts will not begin using value-added until the 2005-06 school year, we agree with Secretary Phillips that value-added be confined now solely to diagnostic uses. Value-added assessment plays catalytic roles in four areas: enabling "job-embedded" professional development; ending the isolation of teachers and teaching; building learning communities where educators are actively engaged in discussing what is happening in their classrooms; and enriching the process of data-driven decision making. In time, as educators become more familiar with value-added and the many advantages it enjoys over current measures of instructional quality, we believe they will grow comfortable with it and support the incorporation of value-added into an accountability system.

Collect value-added data at the classroom level

The Pennsylvania Department of Education should be urged to collect classroom-level data instead of limiting data to grade, subject and school levels as PVAAS currently does. We don't need to collect these data, the argument goes, because value-added enables us to track the progress of individual students, and in so doing get them the help they need to improve regardless of the specific classrooms in which they are found. The unspoken assumption in this practice is that struggling students are randomly distributed across classrooms. While such students are indeed found everywhere, the Tennessee data make absolutely clear that struggling students are found disproportionately in classrooms characterized by ineffective instruction. If we always provide help to the struggling students and never to their struggling teachers (we can't know who they are without collecting classroom-level data), we are doomed to treat the symptoms and never the underlying causes. This is a very poor use of the state's limited resources.

A "career ladder" for teachers based on skill and knowledge

Helping students master the learning required to be good citizens and productive workers in the highly competitive global economy of the 21st Century requires extremely well trained teachers. When they were in college preparing to enter their classrooms, precious few were exposed to the key components that are now thought necessary for success: knowing the difference between an "ability-based" and an "effort-based" theory of learning (you can get smart through hard work and high-quality instruction); understanding the difference between teaching to a bell-shaped curve and to a standards-driven curriculum (all children are expected to and can achieve at high levels); replacing a paradigm based on memorization with one based on problem-solving (critical thinking replaces "drill and kill"); dropping the "one-size-fits-all" approach in favor of "differentiated instruction" (students have varied learning styles); shifting from "teacher-centered" classrooms characterized by lecturing to "student-centered" classrooms (in which students carry a greater share of the responsibility for learning); and using data rather than anecdote to drive decisions (how facts inform decisions).

A highly effective way to provide teachers with incentives to improve their classroom instruction is to build a career ladder whose rungs mark the degree to which they succeed in developing the desired skills and knowledge. Pennsylvania has already adopted the ASCD protocols for teacher evaluations.

Using the “career ladder” to differentiate how teachers are compensated

The pedagogical challenges facing today’s teachers are highly demanding. Empirical evidence now makes clear that teachers, just like doctors, lawyers and athletes, differ widely in their performance. The teachers of the 21st Century should no longer be compensated as if they were industrial workers doing precisely the same thing.

It is time to compensate teachers based on their skill and knowledge. Career ladders, with compensation built into their rungs, make it possible for educators to climb while competing only with themselves, not with each other, and to rise as fast as their talents permit as is the case in other professions.

As noted earlier, when teachers are sufficiently comfortable with the workings of value-added assessment, we can incorporate empirical student-learning results into their evaluation and into their compensation through the career ladder.

Students Learn More from National Board Certified Teachers

At the top of the career ladder would be teachers who achieve certification from the National Board for Professional Teaching Standards. A new study of these teachers reported the following results: *Utilizing a sophisticated value-added model built from 600,000 North Carolina elementary student test scores during a 3-year period, an independent research team has found that National Board Certified Teachers (NBCTs) are far more likely to improve student achievement as measured by the state’s highly touted standardized testing system. The research team, led by labor economist Dan Goldhaber of the University of Washington and the Urban Institute, has noted that these findings “provide direct evidence that the National Board for Professional Teaching Standards (NBPTS) is identifying and certifying teachers who will raise student achievement” and they “could put to rest some of the controversy in education circles surrounding the national certification.” In particular, the study concluded that National Board Certified Teachers (NBCTs)*

- *Are more effective at raising student achievement than teachers who pursue, but fail to obtain, NBPTS certification.*
- *Are more effective at raising student achievement – outside of the year in which they apply – than teachers who do not pursue NBPTS certification.*
- *Have a greater impact with younger students.*
- *Have a greater impact with low-income students.*

Students of NBCTs improved an average of seven percent more on their year-end math and reading tests than students whose teachers attempted but did not earn certification. The influence of NBCTs was most pronounced for younger and lower-income students whose gains were as high as 15 percent more when taught by NBCTs. This is the first large-scale study using standardized tests to link NBCTs and student achievement, and study results confirm that the NBPTS assessment process identifies teachers who systematically produce larger achievement gains. (See http://www.teachingquality.org/resources/html/NBPTS_Goldhaber.htm)

Other key elements in an accountability system that uses value-added assessment

The key quid-pro-quo: a 50/50 sharing of power between unions and school boards

If teachers accept that their evaluation is based on what happens in their classrooms — half value-added and half observation using multiple measures — they must have an equal say in the key decisions over what affects their classrooms: curriculum and professional development.

Peer Review

Teacher evaluation is based half on value-added and half on observation (the four domains developed by Danielson for ASCD). Observation is done through a process of peer review.

Compensation

A “grandfather” clause could give those already teaching the option of remaining under the existing compensation system, but require that all new teachers would enter the profession under the new rules.

Mandatory Remediation

When the quality of their instruction places teachers in the "ineffective" category, they undergo mandatory remediation following the guidelines established by the Columbus Education Association over 25 years ago. Once again, teachers play a key role in this process. Four teachers and three administrators develop a remediation program for each teacher, guide the process and document progress over the year. No progress ends in termination, although a second year is available if the panel believes the teacher is making sufficient progress.

Multi-Year Mentoring

All new teachers participate in a multi-year mentoring program in which an advanced or distinguished teacher works with the new teacher to assist them in making the adjustment to full-time teaching go as effectively as possible.

Teacher Coaches

Create a new category of teachers — drawn from the ranks of advanced and distinguished teachers — whose responsibilities are to work with their colleagues to improve craft. Career teachers should be resourced on the basis of one for every 300 students — that is, roughly one teacher coach would be found in a typical K-4 school.

Expanded Professional Development

Increase the number of paid days per year (some during the school year, some during the summer) devoted to professional development. In Pennsylvania, that would triple the average number of such days to 12 from four.