On the Fallacy of Value- Added Assessment

James C. Alexander, Ph.D.
Kentucky Wesleyan College

18 August 2008

Abstract

Amidst demands for greater accountability, many methods of ensuring teacher quality have been suggested. One that is currently gaining popularity in some quarters is Value-Added Assessment. In this system, statistical methods are applied to determine which teachers are succeeding, and which are failing in adding "value" to a child's educational attainments. The system has been touted as fair and useful for comparing achievement at the building and district levels, across states, and for evaluating the relative success of teacher education units at colleges and universities. This paper calls the fairness and usefulness of this much acclaimed system into questions and suggest that it is impossible for such a system of assessment to account for the multivariate nature of today's schools and teachers.

On the Fallacy of Value- Added Assessment

James C. Alexander, Ph.D.
Kentucky Wesleyan College

Some months ago, I attended a conference sponsored by a higher education organization that represents most, if not all, of the colleges and universities in Kentucky. Normally, I would not be attending events sponsored by this organization, but this time the topic of the conference grabbed my attention. In large measure, the conference was dealing with assessment issues in higher education and concerns related to standards. Since the topic seemed so closely aligned with the concerns of many college and university departments of education, KACTE, the Kentucky Chapter of the American Association of Colleges of Teacher Education, held a preconference session. The purpose of the preconference session was to update education departments at colleges and universities throughout the state on the most recent and possible future actions of the Educational Professional Standards Board affecting teacher preparation and unit accreditation. The EPBS is a state agency distinct from the Kentucky Department of Education with the responsibility of overseeing standards for teachers and teacher certification.

One of the more discussed topics during the KACTE/EPBS meeting was the possible addition of value-added assessment as part of the evaluation of teacher education programs. It would appear this is a move that has a rather high likelihood of occurring in Kentucky as well as other states. It is part of the larger accountability movement, an attempt to hold stakeholders' "feet to the fire" to insure teacher quality. Most teacher educators agree that teacher quality needs to

remain high. Still, it appeared many at the information session were not eager to embrace value-added assessment and apply the results to outcome assessments for teacher education units.

"Value-Added Assessment: An Accountability Revolution", in *Better Teachers, Better Schools* edited by Kanstroom and Finn (Thomas B. Fordham Foundation, 1999), sets forth the basics of value-added assessment as well as the supposed benefits. It is worthwhile to give the reader a basic overview of the chapter. The chapter begins by stating that all stakeholders want improved standards and quality for schools. Yet, there are two problems with enacting and assessing standards as a method of insuring quality. Standards must be set at a high level to be meaningful. Yet, many students from disadvantaged backgrounds are often penalized by such standards, especially when they are tied to promotion or graduation, while wealthier or otherwise advantaged students often appear fare much better on assessments of such standards.

This leads to the second problem. Since high stakes tests must be reasonably passable and yet remain standards-based, they tend to reflect *minimum high standards*. This may solve some of problems for LEP or lower socioeconomic strata students. Nevertheless, it only serves to depress expectations for the higher achieving students. In addition, for those students who appear unable to achieve at even the minimum high standards, there is an incentive to qualify them as special needs students, thus relieving the schools of some of the anxiety associated with the expected high pass rates.

How then might we fairly address these concerns? Enter value-added assessment. With the value-added approach, schools obtain a base level for each child at the beginning of the school year. Then schools complete an assessment at the conclusion of the year. How much did the child achieve? Perhaps they gained 25 points on some scale. This procedure is completed for each child in a class. From this, one may derive a composite gain for any given class. The resulting gain is, in this case, termed the "value-added" to a given class or subgroup over the course of the year (or other period—the chapter recommends at least 75 days) with a particular teacher. Here the standard is not absolute, but a determination of individual gains attained by each child uses the base level as the starting point.

In this system, it is possible for a child to actually gain a lower score on a valid assessment and, yet, attain more added value than a child who might have scored at a higher level. The chapter recommends various comparisons to assess teacher quality. School districts can assess quality at the building level. One might rank states in terms of how much value they add. The chapter also recommends assessing recent graduates of teacher education programs as an index of how well a given program is doing in preparing teachers compared to another institution. Although the chapter does admit to differences in children beyond a school's control, the authors clearly believe that in comparing the child only to his/her base level instead of an imposed outside standard, in a sense assessment is individualized for everyone. "Outside of school effects" are addressed simply using value-added assessment. The system appears self-contained and valid.

The problem with these proposals is that the notion that pre-existing and exogenous factors play such a minimal role. They do not. The system itself is fundamentally flawed. It sounds good; it sounds fair. It would seem to be logical and useful. Given the right situation, it might be of some limited use. It might be useful to note general trends with a large sample of students. However, it is of limited use at the individual classroom level and even less useful at the subgroup or individual child level. It is completely useless as a method of assessing teacher preparation programs.

It does not take much reflection on how conditions are in "the real world" to see the problem. Take the issue of transience. Students and families move more than was the case twenty years ago—or even ten years ago. Many students do not remain in a class seventy-five days, or they repeatedly leave and return. This significantly reduces the size of the pool from which teachers have to draw a sample to prove they are adding value. A sample size of twenty is significantly better than a sample size of ten when one is using a composite. That situation plays both ways. Value-added computations might be artificially high or low depending on sample size.

An issue that has gained prominence in the United States recently also figures prominently here. What impact will legal immigrants and illegal immigrants have on a value-added paradigm of assessment? Many of these students are LEP students. How will they figure into the equation so as not to skew the results? Consider the case of two teachers in the same school. Teacher A has five LEP

students while teacher B has one. Will the natural course of things work in such a way that the value-added comparisons between these two teachers will be equitable? Especially when so much may be riding on the results?

In Michael Winerip's December 9, 2007 *New York Times* article, "In the Gaps at School, Weighing Family Life," the results of a recent study by the Educational Testing Service (the publishers of such fixtures as the SAT and PRAXIS Series Tests) are discussed. If ETS has expertise at all, it must certainly be in the field of assessment. ETS researchers took four variables, the percent of children living in single parent families, absenteeism, how much preschoolers are read to daily by parents, and the amount of television watching. Using these factors the researchers found they could predict states' results on reading tests in eighth grade with a high degree of accuracy. They determined that these four factors accounted for two-thirds of the variance of results between states.

The report went on to state that by the time many kids start kindergarten, they are already very far behind. A child from a single parent home is 2.5 times more likely to repeat a grade. By age 4, a child from a professional family hears about 35 million more words than his more disadvantaged counterpart.

All of this is significant and, it is understood, little of it is within the school's, much less the teacher's control. Add to this the factors of transience and LEP students already mentioned and it becomes apparent that value-added assessment, though perhaps an improvement in some respects, is far too crude a tool to address something as politically charged as teacher quality and high standards. Schools deal

with a very multivariate clientele, value-added assessment just cannot address the variables (which are growing exponentially) that teachers face every day. Its usefulness requires far less variance. Using value-added assessment to make precise judgments and decisions is akin to using a sledgehammer as a fly swatter. It is just not sensitive enough to the variables of the real world.

The problem gets worse when it is suggested that recent graduates (say the first three years after graduation) are ranked using value-added assessment as a method of judging the effectiveness of teacher preparation programs. It is rather common knowledge, that "somewhere in the ballpark" of a third of new teachers leaves the field by their third year. Almost 50% have left the profession within five years. Why is this?

A study by the National Center for Educational Statistics reveals that 42% leave due to family or personal problems and nearly 30% from dissatisfaction see the NCES web site search page for these results and many more studies on teacher attrition). There are several factors to consider. Those of us who work with preservice teachers can likely attest to most of them. Sometimes a (seemingly) very promising student teacher gets a job and decides in a few months that "it's just not for her." Other times, teachers get bogged-down in the bureaucracy. Others do not like the tension of the test-driven environment. Some leave. Some hang on because, for whatever reason, they cannot leave. Others just lose motivation and interest.

Really, in many ways it is the same problem as that of pinning all the responsibility for a child's success or failure on the teacher (or at least a large part of

it). Motivation, satisfaction and interest after entering the profession are factors simply beyond the teacher preparation unit's control. It is all so multivariate. It seems a bit like scapegoating and shifting the blame to hold teacher colleges responsible for the performance of their graduate's students when there is so much beyond their control.

An April 10, 2008 press release from the Kentucky Education Commissioners Office, states that only 37% of Kentucky schools are on track to make the goal of proficient by the target date of 2014. Yet, the report goes on to cite cut after cut in the education budget. Instead of misguided approaches such as value-added assessment and attempting to fix blame on teachers and the institutions that prepared them, doesn't it make more sense to fund programs addressing the strengthening of families, preschool education, adequate health care for children and families, and finding ways that make teaching a more attractive profession for the best and the brightest? Then we will be making a start in addressing this very complex enterprise called education.