Race to the Top, Value-Added Models, and the Catholic View of Education
Guillermo Montes

Race to the Top is resulting in the widespread adoption of value-added models to measure teacher performance. There are concerns about the reliability and validity of these methods and about the wisdom of the Federal government mandating how to conduct non-federal employees’ performance reviews. (Editor’s note: This article was written in June 2011.)

POLITICAL CONTEXT

In 2009, the Obama administration initiated Race to the Top (RTT), a competition for substantial federal aid ($4.3 billion) to K–12th grade public education at the state level. The program has been immensely successful in promoting substantive change in state-level educational legislation in the direction of facilitating the emergence of public charter schools, the adoption of national standards, and the use of value-added modeling (VAM) in teacher evaluations. On the charter school front, New York, Illinois, and Massachusetts (all heavily unionized states) made legislative changes to remove barriers to increasing the number of public charter schools or to facilitate access to existing charter schools for students in low performing public schools, thus showing that the promise of hundreds of millions of dollars was an incentive powerful enough for legislators to overcome public teachers union resistance to educational reform. Following over twenty years of intense debate regarding national curriculum and standards, the Obama administration has succeeded in having forty-two states, one territory, and the District of Columbia voluntarily adopt the National Governors Association and the Council of Chief State School Officers’ common core state standards for English Language Arts and Mathematics (see www.corestandards.org), essentially a first step towards a national curriculum. In addition, states moved in the direction of teacher evaluations based on student performance, in order to qualify for the funds. Performance of teachers will be measured using relatively new statistical techniques collectively known as value-added models (VAM). It is on these models that this paper concentrates.

The RTT competition had two phases. Some critics have argued that the awards reflected politics rather than merit (Hess 2010, Petrilli 2010,
Herbert 2010). In general, RTT is supported by the Obama administration, with quiet support from many Republicans and conservatives, and opposed by a sizeable portion of the President’s own party and a number of liberal-leaning organizations such as teacher unions, the Urban League, Rainbow Push Coalition, and the Economic Policy Institute.

**VALUE-ADDED MODELS**

Value-added models were initially designed by economist Eric Hanushek (2006) based on an application of the production function to education. In neoclassical theory, a production function postulates that a product is produced by a combination of labor, capital, land, and raw materials, based on technology and on the organization of labor. The production function was a classical tool, mostly dormant during the Keynesian revolution, which was recovered in many areas of micro and macro economics during the 1970s and 1980s (Sandmo 2011). Applied to education, the analogy of production narrowly conceptualizes the product as the achievement test score. The inputs are the ability of the child (considered as a raw material), family inputs into the educational process, school inputs into the process (including teacher quality), and other inputs, such as peer and neighborhood effects.

The econometric specification faced large practical obstacles because school systems did not collect information on many important inputs to the educational process. Hanushek solved the problem, at least partially, by using the child’s prior achievement as the child’s own control group. Thus, if a child had low ability, it would affect the absolute level of performance as measured by a suitable test, but not the change (or growth or value added) in performance between current and previous grade. Because school districts often collect the student’s prior achievement, the value-added approach could be measured in school settings. In theory, VAM provides a better approach to measure the impact of teacher quality for that particular grade than absolute levels of student performance. In particular, in a VAM, the average change of a classroom in a particular achievement test score (e.g., 4th-grade reading) is attributed to the teacher, and called teacher quality. Note that there is no attempt to define what quality is or to measure it directly. To this day, no one knows what pedagogical practices differentiate between high and low VAM teacher quality, and whether such practices are stable, replicable, and applicable to various types of classrooms.

Statistician William Sanders applied this concept to the state of Tennessee in the 1990s. Tennessee developed the Tennessee Value Added Assessment System (TVAAS), the most sophisticated value-added system for teacher performance in the world and undoubtedly one of the assets that contributed to Tennessee’s winning phase I of RTT. TVAAS uses a
sophisticated mixed model VAM that (a) uses tests that are aligned to the curriculum and have enough variability to measure change in performance, (b) uses all available data on the student (not just one prior test result), (c) handles missing data, and (d) uses several years of data to estimate the teacher quality score. Importantly, the TVAAS does not control for students’ demographic or socioeconomic characteristics. In their view, doing so would be equivalent to setting lower expectations for students of color, or for students in poverty. TVAAS has been able to show that teacher quality estimates are uncorrelated with student demographic characteristics (Sanders and Horn 1998; Ballou, Sanders, and Wright 2004; Wright and Sanders 2007; Wright, Horn, and Sanders 1997).

The TVAAS has provided Tennessee with some important information: in particular, teacher ratings are uncorrelated with student demographics; high performers, particularly minority high performers, are least likely to benefit in mathematics; and class size and heterogeneity in ability within the classroom are uncorrelated to student growth.

TVAAS has also been repeatedly criticized. Some of the criticisms show a lack of understanding of the current TVAAS model (e.g., concerns about missing data, selection of tests—which is not done by TVAAS but by the state, lack of transparency on their sophisticated model; Sanders, Wright, Rivers, and Leandro 2009). As mentioned before, the most important criticism is a dispute on philosophy rather than statistics: whether one should include student demographic controls in the model and if so, at what level (student, classroom, or school); and whether doing so constitutes lowering expectations for teachers of children of color or children in poverty. Indeed, whether one can simultaneously maintain—as a teaching norm—equal expectations for all children and yet recognize that teaching children of various backgrounds may not be equally difficult is an unresolved and interesting philosophical issue. In other words, can the same quality of teaching be reasonably expected to yield similar value-added achievement growth in students from markedly different backgrounds? Because VAM defines teacher quality as value-added student achievement, it assumes the answer to this question must be yes. In reality, no one knows because there are no widely accepted measures of teacher quality independent from student achievement.

**GENERAL CRITICISMS OF VALUE-ADDED METHODOLOGY**

There are four types of criticisms of the VAM approach: general concerns about test-based accountability in education, specific concerns that VAM
will not obtain teacher quality reliably, concerns about the roles and responsibilities of teachers, and confusion of correlation with causation.

First, some of the criticisms of VAM are not really arguments with the methodology but with the general accountability framework that has characterized so much of federal educational policy since the 1990s. In this category, one includes criticism of standardized testing as a measure of learning, as well as concerns that high stakes testing results in cheating by students and teachers, thus yielding inaccurate data (Gillum and Bello 2011, Fessenden 2007).

Second, critics point out that VAM does not measure teacher quality fairly. New research has shown that VAM teacher quality is correlated with student characteristics (Newton, Darling-Hammond, Haertel, and Thomas 2010), and that student assignment into classrooms may be based on prior performance (Rothstein 2010), thus resulting in VAM attributing to teachers effects that actually stem from classroom composition. Other critics point out that VAM should not be used for the majority of subjects because it was created for core subjects where hierarchically aligned tests exist (e.g., reading and math). Similarly, attribution of the classroom growth to an individual teacher is difficult in cases where there is team teaching, or where the test is not perfectly aligned with the curriculum, as is often the case. Finally, if students are not tested at the beginning of the year, achievement loss during the summer is attributed to teachers even though it is clearly beyond their control. Thus, VAM estimates may not accurately measure teacher quality.

Additionally, critics claim that VAM estimates are unstable (Baker et al. 2010; Briggs and Domingue 2011; Newton, Darling-Hammond, Haertel, and Thomas 2010) and highly dependent on how the model is specified (and there are many possibilities). One solution, adopted by TVAAS, is to calculate multi-year averages for each teacher. While no one knows whether true teacher quality changes year to year or the oscillations are an artifact of the measurement process, critics argue that teachers cannot change their pedagogy on the basis of such variable estimates. As a result of these practical issues, many researchers on both sides of the VAM debate agree that VAM is not ready for use in a high stakes context (McCaffrey, Lockwood, Koretz, and Hamilton 2003; Braun 2005; Harris 2008; Newton, Darling-Hammond, Haertel, and Thomas 2010; Baker et al. 2010; National Research Council 2010).

Third, at the core of the debate is whether teachers should be responsible for student academic growth. This is often a political and philosophical source of disagreement. Empirically, psychological research on teacher responsibility for student outcomes shows that teachers do not feel respon-
Race to the Top, Value-Added Models, and the Catholic View of Education

sible for the cause of student failure but feel responsible for the remedy (see Lauermann and Karabenick 2011 for an excellent review). VAM does not adopt such nuanced views; rather it assumes that teachers are responsible for student achievement growth.

Finally, do VAM estimates amount to cause and effect, comparable to results obtained by a randomized trial? The answer to this important question is no (Rubin, Stuart, and Zanutto 2004; Newton, Darling-Hammond, Haertel, and Thomas 2010), because prior achievement does not control for many unobservable educational inputs that vary within a single academic year. Unfortunately, we have no empirical comparison between a randomized trial’s experimental teacher effects and VAM’s estimates; thus no one knows whether VAM teacher quality is actually causally attributable to teachers, or merely correlated.

A CATHOLIC PERSPECTIVE

There are multiple ironies in these debates. First, progressives are essentially taking a neoconservative stance by cautioning America about the unintended policy consequences of a federal interventionist approach. Second, it is ironic that after decades of struggle against the legacy of Taylorism in American public schools, the revival of classical economics is resulting in public education being framed as a production process once again. Finally, conservatives who have argued for the classical liberal (as in liberal arts) education tradition and/or the libertarian tradition of free markets find themselves advocating for a legislation that imposes federal intervention, steps towards a national curriculum and a mechanical, factory-like, conceptualization of education with national test scores dominating the content and form of education. This raises the general issue of whether libertarian economics is helping or hindering the promotion of Catholic social teaching.

Undoubtedly, frustration with the protracted resistance of unionized teachers, and occasionally the lack of subordination of public employee union goals to the common good, has resulted in some unusual political alliances and a backlash against public employee unions. Instead of focusing on those schools that need reform, RTT employs a micromanagerial approach to non-federal employee evaluation that would make it harder for school administrators to control the performance evaluation of their own teachers. This is clearly a violation of subsidiarity.

In the Catholic natural law tradition, education is primarily an internal process for which the student takes responsibility as he or she develops the intellectual virtues, in the context of primary external responsibility for education residing in the parents (CCC, nos. 1810, 2223, 2229;
Maritain 1943). Certainly, teachers play an important role by setting the structure, providing content and most importantly by modeling and engaging in reasoning with students, but it would be a violation of subsidiarity for the teacher to take full responsibility for student performance because it would negate students’ agency in the learning process. Fostering such agency is arguably one of the most important teaching tasks. This conceptualization of education is anathema to RTT, which treats students as inert raw materials, and praises or blames the teacher for the improvement or deterioration of the material.

While the role of the state is to orient education to the common good, it also has a duty to preserve the freedom to teach (Maritain 1943). Certainly, the endemic dismal situation of many public schools calls for reform, but the reform should not arrogate the evaluation of teachers; rather it should use its power for legislative reform to restore the principal-teacher relationship by giving principals enough power to orient teachers towards the common good when it is needed.

The underlying logic that VAM will lead to improved student achievement is suspect for at least three reasons: (a) Since public schools have been enormously creative in circumventing the spirit of No Child Left Behind while following the letter of the law, it is reasonable to believe a similar approach will be used to respond to RTT; (b) Most teachers teach in areas where strict VAM cannot be estimated because of the lack of hierarchically aligned tests, so they will be held accountable for the achievement growth of the whole school, thus creating a tragedy of the commons problem; and (c) The lack of pedagogical content of VAM estimates together with their high year-to-year variability provides incentives for teachers to take a wait-and-see approach because there is little evidence that any particular pedagogical action will improve VAM teacher quality estimates.

In sum, the value-added models impose a philosophical analogy of production onto education, which—I think it is fair to say—goes against the Catholic natural law tradition on what it means to be educated, precisely because students are treated as raw materials instead of active persons. In addition, the specification at the federal level of how to conduct non-federal employees’ performance reviews is not only a violation of subsidiary but also a dangerous precedent. If successful, the precedent will be established for the federal government to tell other employers how to evaluate the performance of their employees, and there is no guarantee that such an approach will not be extended to other organizations receiving federal funds. Because education reform will be on the public agenda next year, there is reason to pay attention.
References


Maritain, J. 1943. Education at the Crossroads (New Haven, Conn.: Yale University Press).


