Gazing into the Crystal Ball: I See VAM in Your Future

Roger T. Wilson
Grand Valley State University, wilsoro@gvsu.edu

Follow this and additional works at: http://scholarworks.gvsu.edu/colleagues
Part of the Education Commons

Recommended Citation
Available at: http://scholarworks.gvsu.edu/colleagues/vol12/iss1/15
Evaluating educators has been an integral part of the field for over a century. But increasingly, debate about the rigor of those evaluations, their general value for teachers’ professional development and growth, and their implications for the less instructionally proficient has arisen. This latter issue has gained momentum over the past decade as recommendations for the scope and criteria of those evaluations have evolved. Considerable focus upon evaluations is tied to concerns that there exists a disproportionately high percentage of faculty being awarded tenure and exemplary annual ratings. The worthiness of the entire process as well as some of the recipients has been questioned, especially in districts where student achievement is deemed to be seriously lagging. Consequently, it is being increasingly argued that so-called “high quality” teachers can be determined, in considerable part, by student assessment results, often state standardized test scores. From there, it is a short walk to claims that poor performing students, often in inner-city districts, could approximate their better performing suburban counterparts if only high quality educators were identified or cultivated through rigorous personnel actions. In fact, carried to its illogical conclusion, claims have
Gazing into the Crystal Ball: I See VAM in Your Future

By Roger Wilson, GVSU Faculty
Illustration By Lisa R. Tennant

even arisen that the achievement gap between middle-class white and poor (and often minority) students could be dramatically reduced were underperforming students simply placed with high quality teachers for several consecutive years (Ravitch cited in Haertel, 2013, p. 6).

Background
There has been no shortage of educationally “concerned” groups and organizations weighing in on the issue of educator evaluations. For example, ex-Washington, D.C. superintendent Michelle Rhee’s The New Teacher Project (TNTP), a self-proclaimed “national nonprofit committed to ending the injustice of educational inequality” (TNTP, 2015), found in its own 2009 study that 94 percent of teachers evaluated were placed in the top two assessment categories (i.e., superior and effective/proficient), while less than 1 percent were deemed unsatisfactory. The high assessments were then framed by TNTP as “making it impossible to identify truly exceptional teachers” (Weisberg, Sexton, Mulhern & Keeling, 2009). Later, the organization also “pointed to Florida and Michigan as states that continued to see unlikely high percentages of top-rated teachers” (Vevea, 2013), a statement that followed on the heels of a New York Times article that announced that “In Michigan, 98 percent of teachers were rated effective or better” (Anderson, 2013). Over the years, various media outlets have chimed in on teacher evaluations including with such inflammatory headlines as “Why We Must Fire Bad Teachers” (Thomas, 2010), often citing so-called “damning” evidence from the school districts of the nation’s major cities as justification of the need for improved evaluations. It is the cumulative effect of these kinds of reports in conjunction with the perceived persistence of
student academic underachievement that have fueled the outcry for more rigorous educator accountability.

Along with the installation of the new Obama administration in 2009 came critics of the original NCLB legislation and its focus upon testing and AYP. An alternative was necessary, and the American Recovery and Reinvestment Act of 2009 (ARRA) marked its beginning. One of the goals of the ARRA was to lay the “foundation for education reform by supporting investments in innovative strategies that are most likely to lead to improved results for students” (USDoE, 2009, p. 2). The Act provided $4.35 billion in funding for the now familiar Race to The Top (RTT) program. Whatever those innovative strategies, their focus had to be related to significant improvement in student achievement, and integral to that was the revamping of the evaluation of educators—“revising teacher evaluation, compensation, and retention policies to encourage and reward effectiveness” (The White House, 2009).

When the RTT program was announced, it was done so with three phases in mind. It was immediately apparent that an emphasis upon instructor and administrator quality was a significant consideration for future funding by this program. Michigan was unsuccessful in its first two bids.

**The Case in Michigan**

Under newly elected Governor Snyder, Michigan passed numerous education bills in 2011 including regulations emphasizing more rigorous teacher evaluations that were to include students’ assessment data worth upwards to 50 percent by 2015-16. Then in September 2011, Governor Snyder appointed the Michigan Council on Educator Effectiveness (MCEE) to develop a new state educator evaluation process. Pressure was immediately put upon the committee to report back within 8 months. It did release an interim report in April 2012, but did not release its final report till July, 2013 (MCEE, 2013). The legislature followed by introducing new bills in January, 2014 but the bills struggled in the legislature throughout 2014 and eventually expired. Their reintroduction is imminent.

**The MCEE Report**

The final recommendations of the MCEE represent the culmination of a significant investigation into and assessment of the evaluation of educators. The most contentious aspect of the recommendations is that part of the student growth component referred to as VAM—value-added model. That is, “statistical models that use data from growth and assessment tools to produce estimates of the ‘value added’ by individual educators to student learning” (MCEE, 2013, p. 20). Or put another way, what is the professional impact or contribution of an individual educator to student learning after statistically controlling for the myriad of other factors that also impact that learning. Important to understand is that use of “the term ‘value-added’ … is intended to have the same meaning as the term ‘causal effect’ ” (Briggs & Domingue cited in Haertel, 2013, p. 11).

In fairness to the committee, they addressed the VAM concept and acknowledged the problems associated with it, but then declared that “when comparing the use of VAM data to the alternative of district-developed data models of teaching effects, the MCEE believes that VAMs provide more reliable evidence” (p. 20). This is no small matter. The MCEE is declaring that locally designed models for determining individual teacher impact upon student learning are left significantly wanting, but VAMs, notwithstanding all of their problems, are better. And, while there is an element of truth to that statement, the MCEE is still making an argument favoring the “lesser of two evils.” Teachers should not take much comfort from that declaration because it remains quite contestable as to whether an individual teacher’s attribution to student learning is, in fact, statistically determinable and consistently so over time in the ways that are being claimed. People’s careers may depend upon this.

The argument being put forth by the MCEE is that the myriad of complex social variables that impact student learning are not only identifiable, but adequately so, and are then capable of being statistically teased out in some uniform and reliable fashion so that only a teacher’s instructional impact is left, like tea leaves at the bottom of a strainer. That the nation’s major statistical association does not concur (ASA, 2014) and that longitudinal studies demonstrate historical problems with the reliability of VAMs should concern all educators. Finally, one can only
imagine that given the statistical sophistication necessary to calculate and then interpret such an individualized data portrayal, very few, if any, educators will have the foggiest idea as to the data’s computational process, let alone its accuracy, even if published. How does one confront potential errors from a position of ignorance?

VAMs remain controversial and unreliable as a statistical model because of their attempt to claim student progress in achievement is directly related to a teacher’s unique instructional performance or quality, often referred to as “teacher effects” in the literature, but this confuses correlation with causation (ASA, 2014). And however problematic some of the issues associated with the VAM model may appear to Michigan teachers, perhaps of greater concern should be the MCEE report’s argument in favor of using student scores in subject matter not taught by the particular teacher.

“State-provided VAM or growth data in core content areas may be used in a teacher’s evaluation using information from that teacher’s students, even if the teacher does not teach in one of the core content areas. This means that teachers may be evaluated, in part, for the learning of their own students, even in subject areas that they do not directly teach” (MCEE, 2013, p. 2).

Some Problems With VAMs

There is a lack of clarity about the totality of the effect by teachers upon student achievement. There has been considerable research over the past decade or more that identifies teachers as the greatest “within school” factor for improving student learning. But that implies much more than reality reveals. The question that seems to have been overlooked by too many legislators, policy makers and reform critics is “how large is that potential factor?” And therein lays the crux of the issue. Haertel (2013) suggests that the research generally indicates that the variance in student test score gains attributable to teachers averages approximately 10 percent. Additionally, other research by Nye, Hedges and Konstantopoulus (2004) has pointed out that the methodological design for many of these kinds of studies raised a major problem.

“The advantage of this design is that it does not require the researcher to identify in advance, and measure adequately, the aspects of teacher behavior or other teacher characteristics that are related to achievement. Of course, this design cannot identify the specific characteristics that are responsible for teacher effectiveness” (p. 239).

Consequently, we are left with some models which claim that Teacher A is reportedly more effective than Teacher B, but is then unable to explain just how and why. Presumably, that is where the observation component of the MCEE evaluation model comes into play. Whether the observation will, in fact, identify those characteristics with adequate specificity and link them with an increase in student learning remains to be seen. This writer is doubtful.

Conveniently Overlooked

Lost in this discussion, and conveniently so because of the social and policy implications for legislators, is the elephant in the room—those “outside of school” factors which dwarf the variance in attribution of student learning attributed to the teacher. Those outside factors range
anywhere from 50 to 60 percent (Goldhaber, Brewer, & Anderson cited in Haertel, 2013; Hattie, 2003). And these are factors that education policy cannot really address. Nevertheless, they are not irrelevant. Furthermore, as Haertel (2013) points out, “in the real world of schooling, students are sorted by background and achievement through patterns of residential segregation, and they may also be grouped or tracked within schools” (p. 12). Ignoring this fact has greater consequences for teachers of low-performing students because while “VAM scores do predict important student learning outcomes, [evidence suggests that] these scores nonetheless measure not only how well teachers teach, but also whom and where they teach” (emphasis added, p. 17). One understands that statistically controlling for things like socioeconomic status (SES) are commonplace in social research studies, but one also wonders if VAM models can adequately account for things such as “school climate and resources, teacher peer support, and, of course, the additional instructional support and encouragement students receive both out of school and from other school staff [which] all make the test of teaching much easier for teachers in some schools and harder in others” (Haertel, 2013, p. 11).

Concluding Remarks

I am not opposed to Michigan’s K-12 students having the best possible educators in their classrooms. I am only taking issue with the VAM aspect of the student growth component of the Michigan educator evaluation, and specifically, the idea that student scores on standardized tests speak in some direct way to a given teacher’s performance quality. There is no reliable evidence for this claim. “Teachers whose students show the biggest gains one year are often not the same as those whose students show big gains the next year” (Haertel, 2013, p. 6). Such unreliability over time can produce seemingly illogical scenarios where a veteran teacher in good standing, whose students’ test scores outperform the state means and who is well regarded by the district superintendent is declared effective one year and ineffective the next based on a state-developed VAM (Strauss, 2014). That the teacher in question has sued the state education board with affidavit support from the superintendent should surprise no one. Whether similar scenarios and legal responses are what await Michigan is anyone’s guess. Perhaps, the MCEE pilot study conducted with different vendors and their value-added models might provide some insight. “[E]ven when different VAM scores are … highly correlated across models …, some teachers’ VAM scores will change from statistical model to statistical model … [T]eachers with scores near the established cut points will be especially vulnerable to ratings changes that result from small changes in VAM scores produced by different statistical models” (MCEE pilot, 2013, p. 32).

Clearly, this component of the proposed teacher evaluation model is fraught with problems. That a coalition of groups including Michigan’s largest teacher union and all the state’s administrator organizations came out in support in a December, 2014 editorial is disconcerting (Arellano, Cook, Hayes, Mayes, Melton, Miller & Zdeb-Roper, 2014). One cannot help but wonder if they fully appreciated the intricacies and implications of the VAM.
Education critics, numerous legislators and some members of the public may simply see the model as the means for identifying and easily dismissing “poorly” performing K-12 faculty. But to assume that this form of faculty turnover alone will contribute to significantly improved academic achievement, particularly for our less advantaged student population, is utter folly. Much more is required and more beyond the confines of the school building. Perhaps the most important aspect of this entire discussion relates to any thinking associated with using this model to somehow have all minority students in proximity to the state’s top performing teachers or even have a teaching force comprised entirely of the same. “There is no way to assign all of the top performing teachers to work with minority students or to replace the current teaching force with all top performers. The thought experiment cannot be translated into an actual policy” (Haertel, 2013, p. 7). Given the work that Michigan has put into teacher evaluations, given the political climate in the state, given the public mood and the relative strength of the MEA and AFT, and given the K-12 funding requirements of the federal government, it would seem that the enactment of the MCEE’s recommendations are a matter of “when” not “if.”

NOTE TO READERS: The Haertel (2013) article offers a very good layperson’s explanation of VAMs while Rowan, Schilling, Spain, Bhandari, Berger & Graves (2013) is the description and results of the MCEE teacher evaluation pilot study that educators should become familiar with. It, too, is quite readable.

References


