A More Accurate Means of Measurement

A brief overview of value-added models and considerations for school accountability and Educator Effectiveness

Bradley Carl

The stakes have gotten higher for student achievement data. In many ways, a school's or teacher's reputation is now made or broke based upon student assessment scores. The ability to accurately measure student data is more important than ever. Value-added measures are helping bring student data to a new light — one that, when used correctly, can accurately show student growth.

The term "value-added" refers to a class of growth models used to measure improvement in student performance over time. The advantage of using value-added models is that they can "remove" factors from student test data, such as student poverty, that are beyond the ability of schools or educators to control. This allows school districts to get a more clear and accurate picture of student achievement.

Value-added models, or VAMs, developed by the Value-Added Research Center (VARC) at UW-Madison and other organizations, utilize advanced statistical techniques and rely on high-quality data. They are conceptually fairly easy to understand: a prediction about student performance in the future is made based upon prior achievement and selected student demographic characteristics.

VAMs: A PREDICTION ABOUT STUDENT PERFORMANCE IN THE FUTURE IS MADE **BASED UPON PRIOR ACHIEVEMENT** AND SELECTED STUDENT DEMOGRAPHIC CHARACTERISTICS.

Actual performance is then compared to the predicted performance, resulting in either positive value-added (actual performance is higher than predicted) or negative value-added (actual performance is lower than predicted). In either case, the important point is that performance is measured in relation to schools or educators who serve similar students, as defined by prior achievement and student characteristics.

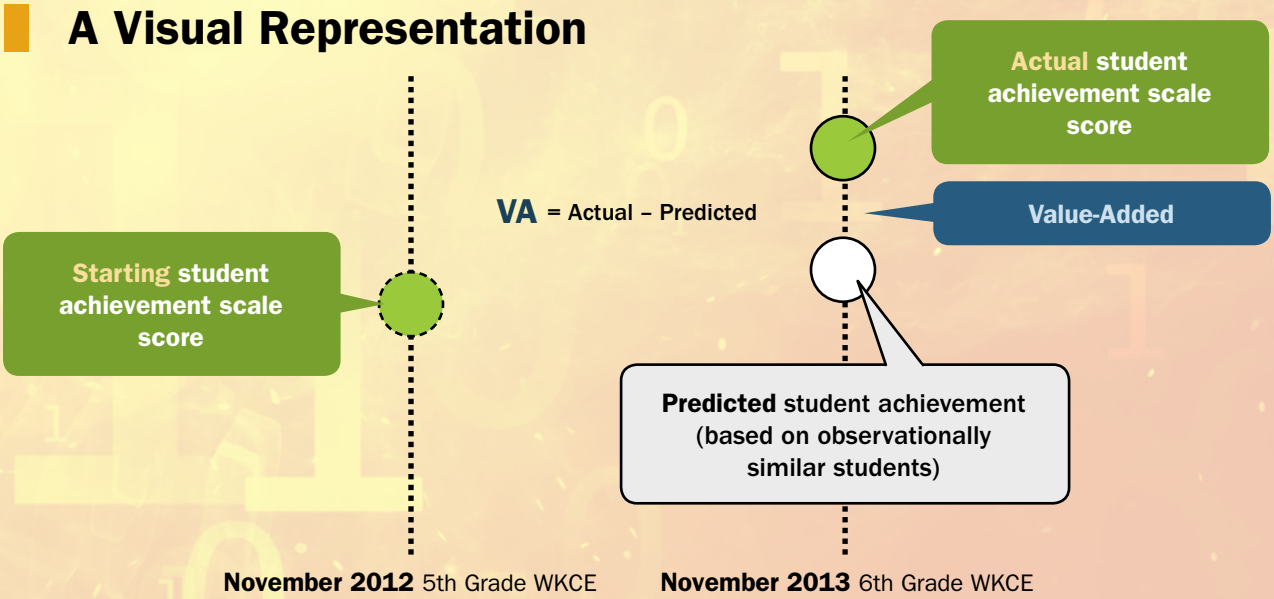
Fairness and Unintended Consequences

Within education, VAMs are used by many states and school districts as one measure of districts, schools, classrooms, and sometimes individual educators, which often serve student populations that vary widely in terms of prior achievement, family income, and other key non-school factors. This approach helps create a more level "playing field" than is typically the case when attainment-based indicators (such as simple proficiency rates) are used to make high-stakes accountability decisions at the school and/or educator level. While attainment measures are very important in reinforcing the need to hold high expectations for the performance of all students, they usually have a strong inverse relationship with student poverty *(higher poverty rate = lower proficiency rate)*.

This raises important policy questions around fairness and unintended consequences such as creating additional incentives for the most able educators to avoid working in the neediest schools. States and districts have moved in recent years to develop "multiple measures" approaches to both school-level accountability systems (school report cards) and new educator effectiveness systems for evaluating teacher and principal performance.

VAMs have most often been used to measure growth on state standardized tests at the late elementary and middle grades; such as the WKCE in Wisconsin. They are also being applied to the next generation of state assessments — Smarter Balanced and PARCC — as well as to commonly administered benchmark assessments such as the

A Visual Representation



Source: *VARC, Value-Added Research Center, University of Wisconsin-Madison*

Measures of Academic Progress, or MAP, the STAR, and the Performance Series. VAMs are also applied in some states and districts to end-of-course exams (using prior test scores and final grades from relevant courses to make predictions), and even in some cases to selected non-cognitive outcomes such as attendance rates.

Looking ahead, as states, including Wisconsin, transition to new assessments, there will likely be new opportunities to measure growth in the high school grades, and to build upon early efforts to apply value-added techniques to measure the performance of institutions of higher education that train teachers and school leaders. In each of these cases, the ability to measure student growth across the entire range of performance, including both low-performing and high-performing students, rather than maintaining a much narrower focus on proficiency, represents an important step forward.

Making it Work

As is the case with all potential measures of school and educator performance, VAMs have limitations and cautions that must be understood and carefully monitored over time. Examples of such cautions include the reliability of VAMs over time, imprecision in the ability of different tests to measure the entire range of student performance, variability in VAM results based on different types of assessments, and the possibility of "narrowing the curriculum" by placing too much emphasis on test-based outcomes that measure only a portion of what we want students to know.

A number of these cautions can be addressed substantially through technical and/or policy decisions. For instance, school districts should use multiple years, rather than a single year, of VAM data to improve reliability. Districts can also avoid test measurement error by applying statistical adjustments to avoid unduly rewarding schools with many low-performing students, which often show very high growth, or penalizing those with many high-performing students, which may show much lower growth.

VARC also recommends the reporting of VAM results with confidence intervals, which convey the precision of the measures and can help avoid over-interpretation of results that may be imprecise in a statistical sense.

Accountability and Educator Effectiveness

At the same time, it is important to keep in mind that there is no single measure of school or educator performance that is without limitations and cautions.

Within the realm of school-level accountability, for example, the use of attainment-based measures generally reflects differences in family background much more than actual differences in the performance of schools and educators.

Within the realm of educator effectiveness decisions, measures of professional practice used by all states require large investments of time on the part of evaluators, rely on the ability of well-trained evaluators to render judgments in an impartial manner, and base this portion of educators' evaluations on a few hours of observation and a limited set of evidence/artifacts.

For grades and subjects that cannot be "covered" by VAMs (more 60 percent of teachers fall into this category), most states are relying at least in part on locally developed assessments for educator effectiveness decisions, which vary widely in terms of technical quality. As such, the point here is to consider the cautions associated with VAMs within a larger context of the limitations that accompany all potential measures of school and educator performance, and to construct (and adjust, as needed) "multiple measures" systems of accountability.

Bradley Carl is associate director and researcher at the Value-Added Research Center, housed within the Wisconsin Center for Education Research at the UW-Madison.