

# Looking Behind the Value-Added Measures Curtain

By Dr. Gregory A. Taranto, Kenneth W. Schrag and Dr. Mark A. Abbondanza



## About the Authors:

**Gregory A. Taranto, Ph.D.**, is serving his 10<sup>th</sup> year as the principal of Canonsburg Middle School, a 2014 Re-Designated Pennsylvania Don Eichhorn and National Schools to Watch Middle School, in the Canon-McMillan School District (S.D.).

**Kenneth W. Schrag** is the assistant principal of Canonsburg Middle School in the Canon-McMillan S.D. He has served in this capacity for the past eight years. Prior to this, he was also a science teacher at the middle school.

**Mark A. Abbondanza, Ph.D.**, is principal of North Strabane Intermediate School in the Canon-McMillan S.D. where he has served as an administrator for the past 10 years.

As children, we all remember that moment when Toto runs behind the curtain and reveals that the great and powerful Wizard of Oz was not what he seemed. He did not have magical powers that could help Dorothy and Toto go back to Kansas. Despite being built up by the community of Oz and all the bells and whistles, he simply was an ordinary man.

As educators, it is our responsibility to dig deep in the research and literature to ensure we are applying the best practices in our schools. The Pennsylvania Department of Education (PDE), as with many state departments of education under the guidance of the U.S. Department of Education, has adopted value-added measures as valid indicators of student progress and educator effectiveness. Just because these agencies have adopted these practices does not excuse us from questioning the validity of such measures. Principals and school leaders have an obligation to exhaust all means to get the facts. We have an obligation to look behind the curtain as our students, teachers and communities deserve nothing less!

Value-added measures (VAM for short; the Pennsylvania Value Added Assessment System or PVAAS in Pennsylvania) have been adopted in most states as a part of the competition for Race to the Top grants or No Child Left Behind (NCLB) waivers. In a nutshell, VAMs, through complicated statistical algorithms, are supposed to be able to predict human behavior. In education, it is being used to predict (or report) how much “value” a school or teacher should add (or has added) to a student’s learning. VAM formulas are also reported to be able to account for many factors that impact student achievement such as poverty. In Pennsylvania, the PA System of School Assessment (PSSA) is the lone student assessment data entered in the formula.

Interestingly, VAM’s start in Pennsylvania was intended for “school evaluation and planning, grade/subject level evaluation and planning and for student level planning” (PVAAS Overview..., 2008). During its rollout in which these intended purposes were stated, PDE also shared that PVAAS was “not being used for teacher-level accountability in Pennsylvania” and even that PVAAS “does not yield a metric on the effectiveness of a teacher” (PVAAS Overview..., 2008). These statements regarding what PVAAS does and does not do were shared during PDE’s trainings provided at intermediate units and school districts. This echoed information was shared in a PDE Penn-Link the prior year, stating “...PVAAS is a powerful tool for school improvement...It is the intent of the Pennsylvania Department of Education that this data will be used solely as a school improvement tool. The data should be used to make decisions pertaining to curriculum and instructional materials, but should not be used for the purpose of evaluating or compensating teachers” (Zahorchak, 2007).

As the popularity of VAM increases and its use is now being expanded to evaluate educators, administrators and schools, more and more research is revealing that VAM is not all-powerful and that predicting human behavior is not something a math formula can do. As principals and leaders of schools, we need to question what exactly VAM does for us.

In a joint statement by the National Association of Elementary School Principals (NAESP) and the National Association of Secondary School Principals (NASSP), the associations warned the chairman of the Senate’s Health, Education, Labor and Pensions Committee that they did not want a proposed educator preparation program evaluated with VAM stating (2014):

NAESP and NASSP have concerns about any proposed grant program that would require states to rate all preparation programs using value-added measures. We



believe there are multiple ways that states can strengthen accountability and standards for preparation programs without developing such a rating system, which will conflate the multiple concerns that research has exposed in terms of the efficacy of VAM with teacher and principal evaluations (NAESP/NASSP, para. 8, 2014).

VAM has also caught criticism from the organization that represents statisticians across the country, the American Statistical Association (ASA) (2014), warns, "VAMs are generally based on standardized test scores, and do not directly measure potential teacher contributions toward other student outcomes" (para. 2). The ASA later goes on to say that VAM studies find that teachers account for 1% to 14% of the variability in test scores. This is a stern warning that ranking educators by their VAM scores can have unintended consequences that reduce teacher quality.

Even in a study prepared for the U.S. Department of Education by Schochet and Chiang (2010), part of the conclusion states that their results were "largely driven by findings from the literature and new analyses that more than 90 percent of the variation in student gain scores is due to the variation in *student-level* factors that are not under the control of the teacher" (p. 35). This same study concluded that VAMs used at the teacher level "are subject to a considerable degree of random error when based on the amount of data that are typically used in practice for estimation" (p. 35).

The reliability and validity of VAM has raised red flags by our national associations and many researchers. Scherrer (2011) states, "Questions of validity arise when one starts to make inferences based on value-added scores. To begin with, labeling a teacher as "effective" based on mathematics and English Language Arts scores is invalid" (p. 130).

Fortunately, the state only includes a small portion of an educator's evaluation on VAM; however, that portion is consuming mass amounts of time due

*"Students, teachers, principals and schools are so much more than a test score."*

to collecting evidence for the evaluation. In the end, the unreliable nature of VAM is causing educators to lose precious time that could be well-spent on academic planning (Darling-Hammond, 2010; McCaffrey, Lockwood, Koretz & Hamilton, 2003; Ravitch, 2013). Moreover, the VAM training from the state cautions us that VAM "does not infer causal relationships between educational variables and student growth!" When it comes to crunching the numbers to measure teacher effectiveness, the multitude of tests that are used as the basis of a VAM each produce different results. If the data were reliable, teachers' VAM scores would not vary from one year to the next. Obviously, this would lead one to question the reliability of the mathematical calculations behind VAMs.

Why are VAMs accepted by so many? VAM formulas are complex and confusing. The complexity of VAMs leads to confusion and as a result intimidation by the people being evaluated with such measures (Ewing, 2011). Unfortunately, this complexity results in many simply accepting the numbers as valid. This is a mistake, and educators need to look to the literature and research as guidance and support.

Students, teachers, principals and schools are so much more than a test score. Working with young human be-

ings who are undergoing rapid changes is unique to the field of education. In every school, the experiences offered to students enable them to grow socially and emotionally, which is not measured by a VAM. There is not a mathematical formula that can measure the effectiveness of all the wonderful things that occur in our schools each and every day. No VAM formula can account for a jazz band program that a school adds, the incorporation of a morning intramurals program, a peer-to-peer buddy program that supports students with autism or a charity drive orchestrated by students. However, if a school narrows the curriculum for test prep purposes, which frequently happens, and eliminates opportunities such as the arts, physical education, exploratory, etc., then a school's VAM score may increase especially since VAM data are ultimately derived from a single test. In the end, is it worth it? Did the school improve the *overall* learning experience of those students? School leaders and teachers are the only ones that we can depend on to hold the line of what is right and what works best for children. Darling-Hammond, Amrein-Beardsley, Haertel & Rothstein (2013) write,

Using VAMs for individual teacher evaluation is based on the belief that measured achievement gains for a specific teacher's students reflect that teacher's 'effectiveness.' This attribution, however, assumes that student learning is measured well by a given test, is influenced by the teacher alone and is independent from the growth of classmates and other aspects of the classroom context. None of these assumptions is well supported by current evidence (p. 8).

Simply put, the use of a tool that was designed to measure student growth based on high-stakes assessments is not effective, nor is it realistic to estimate an educator's causal growth.

The literature that does support the use of VAM states that they should be used as part of a tool for improvement and not determining an educator's contribution to student growth. This is a

**“Our data so far show decreased behavior referrals, increased parent involvement in student problems and improvement in teacher/student relationships...”**

struggle. Research, and my gut, told me that creating the time was a non-negotiable and that has proven to be correct. Expecting teachers to stay after school or create the needed time to make the collaboration work would have given their meeting time a completely different tone.

**Lesson #4: Never take for granted that the kids know it.**

Our model is, true to the State College way, uniquely our own. Instead of a pullout transition class, we created an integrated, infused curriculum of study skills, behavior expectations, peer relationships, citizenship and academic planning topics that are negotiated by each team of teachers. They determine who teaches what lessons and the other teachers in the cohort back up the lessons by knowing what was taught and using the same language and expectations in their own content courses. Resoundingly the teachers are saying, “Why did we think they came to us knowing this?” They have seen over and over that intentionally establishing expectations for learning transition skills has shown the kids didn’t know how to negotiate learning and study skills. Our “good” students perhaps did, but many of our students did not. This year they are learning.

**Lesson #5: We have only just begun.**

I recently went to a professional learning community conference, and as I sat through the first session, my thought was: “Wow...we’ve started this right.” I moved on to

the second session and my thoughts turned to: “Wow...we have a long way to go.” Both were correct. We have come so far in two years, but now we have to start tracking and determining if what we are doing is working to improve student learning. Our data so far show decreased behavior referrals, increased parent involvement in student problems and improvement in teacher/student relationships (teachers by way of common conversations about what works for a given child feel they are getting to know their students at a deeper level). However, we can’t stop here. Is the 9LC improving academic student learning? To get the answer to this we are collecting data to compare to our pre-9LC year, refining our curriculum so that it is aligned to what students need to be active and invested learners and continuing to make time and relationship a priority.

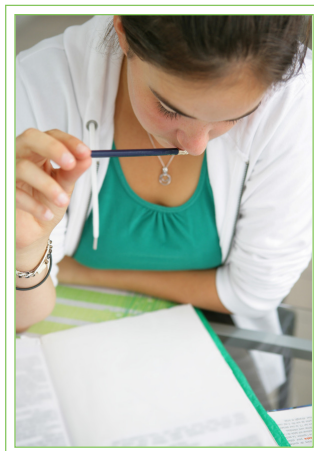
As our school moves forward in planning our new educational model for the 10<sup>th</sup>- through 12<sup>th</sup>-grade students, it will be exciting to see how the lessons we’ve learned can be adapted and used as a resource for the learning communities to come. We have many exciting, yet demanding, days ahead of us. I imagine by the end of the year I will have more lessons learned than I can possibly write about in an article. I look forward to watching and learning as we continue to strive to ensure our students learn.

For more information, contact the author at [kay15@scasd.org](mailto:kay15@scasd.org).

## Looking Behind the Value-Added Measures Curtain

Continued from page 21

clear warning against using them for high-stakes decision making. For some reason, the warnings are being ignored. This is especially troubling considering the increasing use of VAMs when forming policy and making decisions with far-reaching consequences, including but not limited to: individual educator’s employment; scores, grades or labels placed on schools; even landmark decisions on teacher tenure. Again, as educational leaders, we must be informed of the flaws of this tool, implications of relying on this (or blindly accepting it), as well as the growing body of literature that does not support VAM.



Like Toto, who revealed that the Wizard of Oz was nothing more than a deceptive ordinary man, school leaders must take action by ‘pulling the curtain’ back on VAMs to understand the detrimental impact they can have on their educational community. As school leaders, we cannot let the effectiveness of our teachers and the culture of our schools be determined by a ‘magical’ mathematical formula that does not calculate humanity in the equation!

For more information, contact Dr. Taranto at [tarantog@cmsd.k12.pa.us](mailto:tarantog@cmsd.k12.pa.us).



Continued from previous page

The year ended on an even higher note for the students. In July, the College Board published stunning AP scores for South Side High School. The school had a record number of AP Scholars and as an added bonus, as “The American Experience” was designed as an 11<sup>th</sup>-grade course offering, they were mostly juniors! With that, enrollment in senior-level AP courses ballooned. The number of students who had achieved a “3” or higher on multiple exams tripled. Many of these students had never taken an AP course before and probably thought they didn’t have what it takes. *They know better now.* These results have been a game-changer for the students of South Side.

The progressive and innovative thinking that led to these results is typically not associated with small, rural school districts. With a poverty level close to 30% and an estimated 10% of district residents or fewer holding a bachelor’s degree or higher, college has become a reality, rather than a dream in this community. Thoughtfully managed, collaboration produces results.

For more information, contact Mr. Paull at [acp@sssd.k12.pa.us](mailto:acp@sssd.k12.pa.us).

## Building Teacher Capacity: Our *Only* Chance To Provide Ongoing *Instructional Leadership*

Continued from page 26

Additionally, our school achievement data can also be substantiated by a reduction of failure rates and a drastic increase in honors’ and distinguished honors’ recipients using our local assessment data such as report cards and cumulative promotion rates. Combined with a steady increase in our state required Pennsylvania System of School Assessment (PSSA) data, it completes a full picture of improvement in academic achievement as well as cultural data improvement.

In 2013, the National Association of Middle School Principals (NAMSP) recognized our school improvement challenges and achievements when we were chosen as a distinguished runner-up for the NAMSP Middle School of the Year. In honor of this distinction, the NAMSP presented



Northeastern Middle School was selected as a national runner-up for the National Association of Middle School Principals annual Middle School of the Year during the 2013-14 school year.

NMS with a large foyer banner in recognition of our continued growth and high achievement (with close to a 20% Individualized Education Program (non-gifted) and 50% economically disadvantaged student population).

For information related to building teacher capacity in your school, or any of the ideas included in this article, contact the author at [alessanm@nesd.k12.pa.us](mailto:alessanm@nesd.k12.pa.us).

### References

- Combs, J.P., Edmonson, S. & Harris, S. *The Trust Factor*. Larchmont, NY. 2013.  
Schmoker, M. *FOCUS*. Alexandria, VA. ASCD. 2011.  
Wahlstrom, Karen. *Using Data to Improve Student Achievement* (October, 2002).

## Looking Behind the Value-Added Measures Curtain

Continued from page 23

### References

- ASA statement on using value-added models for educational assessment. (2014, April 8).  
Connelly, G. & Bartoletti, J. (2014, June 23). [Letter]. National Association of Elementary School Principals & National Association of Secondary School Principals.  
Darling-Hammond, L. (2011, May 24). Value-Added assessment is too unreliable to be useful. Retrieved from New York Times website: <http://www.nytimes.com/roomfordebate/2010/09/06/assessing-a-teachers-value/value-added-assessment-is-too-unreliable-to-be-useful>.  
Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E. & Rothstein, J. (2012). Evaluating teacher evaluation: Popular modes of evaluating teachers are fraught with inaccuracies and inconsistencies, but the field has identified better approaches. *Phi Delta Kappan*, 93(6), 8-15.  
Ewing, J. (2011). Mathematical intimidation: Driven by the data. *Notices of the American Mathematical Society*, 58(5), 667-673.  
McCaffrey, D., Koretz, D., Lockwood, J.R. & Hamilton, L. (2003). Evaluating value-added models for teacher evaluation. *RAND Corporation*. Retrieved from [http://www.rand.org/content/dam/rand/pubs/monographs/2004/RAND\\_MG158.pdf](http://www.rand.org/content/dam/rand/pubs/monographs/2004/RAND_MG158.pdf).  
Pennsylvania Department of Education. (2008, November). *PVAAS overview; Evaluating growth, projecting performance [PowerPoint Slides]*.  
Ravitch, D. (2013). *Reign of error: The hoax of the privatization movement and the danger to America's public schools*. New York: Alfred A. Knopf.  
Scherrer, J. (2011). Measuring teaching using value-added modeling: The imperfect panacea. *National Association of Secondary School Principals Bulletin*, 95(2), 122-140. <http://dx.doi.org/10.1177/0192636511410052>.  
Schochet, Peter Z. & Chiang, Hanley S. (2010). *Error rates in measuring teacher and school performance based on student test scores gains*. (NCEE 2010-4004). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.  
Pennsylvania Department of Education. (n.d.). *Penn-Link: Appropriate use of PVAAS data* (G. Zahorchak, Author).