

[Home](#) [Communities](#) [Members](#) [Resources](#) [Granters](#) [FAQs](#) [Contact](#)

[MY ACCOUNT](#) [LOG OUT](#)

[Home](#) » [Value-Added and Other Measures](#)

## TIF Spotlight: Interview with Maricopa County

**Grantee:** Rewarding Excellence in Instruction and Leadership (Maricopa County)

*Greetings TIF Grantees! It is an honor for us to have the opportunity to highlight some of the TIF exemplary work that is transpiring throughout the country. This portion of the newsletter gives grantees the opportunity to learn in more detail best practices, strategic approaches, and specific action steps that have proven to be effective and successful for TIF grantees as they implement the Core Elements. It is our hope that you will learn more about the excellent educator quality reform work that is taking place in our TIF districts and schools.*

Wesley Williams II

Senior Research Associate  
Westat

### Maricopa County

Lori Renfro, TIF Project Director

Theme: Value-Added

**Inquiry:** Describe your Value-Added Model/s (VAM). Please also share the strategic approach taken to secure buy-in from the educational community (TIF staff, the local union, parents, educational stakeholder, etc.) once the model/s was finalized and approved.

REIL's (Rewarding Excellence in Instruction and Leadership) VAM is described by Basis Policy Research (BPR) as "a covariate adjustment value-added model that serves to estimate teacher and school effects." Basis Policy Research is the firm that MCESA has contracted with to develop and support the implementation of the VAM. Drs. Keke Liu and David Stuit from BPR and Dr. Cory Koedel from the University of Missouri have formed the core team working on this project. In addition, Dr. Christopher Thorn, who was with VARC at the University of Wisconsin-Madison, has provided additional support via TIF-provided technical assistance. BPR also states: "The model regresses students' current year test scores on prior year test scores and a number of student- and classroom-level covariates. Separate regressions are run for each grade, subject, and year. After each regression, students' predicted test scores are subtracted from their actual test scores to determine their "residual" gains. The average residual within a teacher's class serves as her value-added estimate. If a teacher's students systematically exceed their predictions, this teacher will have a positive value-added estimate. Conversely, if students perform worse than their predictions, on average, this teacher will have a negative value-added estimate."

To account for measurement error, the recommended model will use the standard errors of measurement (SEMs) from students' post-test and pre-test scores to assign more weight to students with more precise scores, and less weight to students with less precise scores. REIL's VAM will also combine the estimated effect of an individual teacher (or school) with the average estimate of all teachers (or schools) evaluated: the more precise an individual estimate is, the more weight it receives. By shrinking the estimates, teachers or schools with small sample sizes are moved toward the mean estimate of all teachers or schools in the sample, thus reducing the likelihood that they will end up with extreme high or low VAM estimates due to sampling error.

With REIL's VAM, teachers will receive a rating on a 1 to 5 scale based on their shrunken VAM estimates. The confidence interval (CI) of each shrunken VAM estimate is incorporated to further improve the accuracy of performance designation. Preliminary criteria for the five performance designations are as follows:

- 1 (Ineffective): The *upper bound* of 99<sup>th</sup> CI is lower than the sample average estimate, the teacher is classified in the lowest performing group.
- 2 (Suggestively Ineffective): The *upper bound* of 95<sup>th</sup> CI is lower than the sample-average estimate but the *upper bound* of the 65<sup>th</sup> CI is still above the sample average estimate.
- 3 (Not distinguishable from Average): The 65<sup>th</sup> CI overlaps the average sample estimate.
- 4 (Suggestively Effective): The *lower bound* of 65<sup>th</sup> CI is above the sample-average but the *lower bound* of 99<sup>th</sup> CI is below the sample average estimate.
- 5 (Effective): The *lower bound* of 99<sup>th</sup> CI is above the sample-average estimate.

SEARCH THIS SITE:

Search

SEARCH TITLE

SEARCH AUTHOR

FILTER BY COMMUNITY

FILTER BY GRANTEE

Apply

Leave your feedback  
Improve the community

VISIT CECR WEBSITE

TIF TALK

**tif** Teacher Incentive Fund

Added Estimates

Designation	Ineffective	Partially Effective	Effective 1	Effective 2	Highly Effective
Account	\$0	\$0	\$3000	\$4000	\$5000

12 teachers will receive reading and math scores on the 1 to 5 scale. For teachers in schools that administer both AIMS/SAT-10 and Galileo K-12 (e.g., K-5, K-8), the average of their reading and math scores (rounded to the tenth decimal) will account for 15 percent of the REIL Score in 2012-13, with Galileo growth results making up the remaining 10 percent of their school growth component.

With respect to buy-in from the educational community, we would like to provide the following background information:

- 1) The state of Arizona passed legislation (Senate Bill 1040) in 2010 that required the State Board of Education (SBE) to adopt and maintain a model framework for a teacher and principal evaluation instrument that used quantitative data on student academic progress for at least 33-50% of the evaluation outcomes.
- 2) The submitted TIF3 grant application wrote specifically to a competitive preference priority to use value-added measures of student achievement. Our state's accountability system was using a value-added measure in its labeling of schools, and our alliance partners felt that value-added measures were a fair way to look at the individual contribution of educators to a student's overall performance. Therefore, our grant application required that the proposed PBCS would use a value-added measure of the impact on student growth as a significant factor in calculating differential levels of compensation provided to educators.
- 3) Each district is supported by a REIL Field Specialist who is a MCESA employee hired specifically to assist with grant implementation in a specific district. In addition, MCESA hired a REIL Cross-District Field Specialist to help support transition team work across all districts (transition teams are described below).

Upon award of the TIF3 grant, REIL continued the process of engaging stakeholders in the discussion of the VAM development. This engagement occurred through several activities as part REIL's communication plan. Examples of this are highlighted below.

1) Transition Teams: Each district had a data management and assessment transition team that helped guide the work in relation to development of the VAM, as well as the assessment development process. The transition teams comprised representative stakeholders (e.g., superintendent, REIL implementation contact, curriculum director, primary teacher, intermediate/middle school teacher (or content areas/dept chairs for high school), principal, representation from IT department, human resources/payroll; REIL field specialist). REIL field specialists have facilitated Data Management and Assessment Transition Team (DMATT) Meetings over the course of grant implementation. Through these meetings, we established a common language regarding data management, assessment, and value-added. For example, transition team members learned about the concept of value-added using VARC's *Oak Tree Analogy* presentation. Transition team members also helped us understand the depth and breadth of educator concerns/ideas regarding assessment development and delivery and the connection to value-added. Through a series of activities called "Stump the System," we worked together to identify all the factors that needed to be considered/addressed as we engaged with the work. As we uncovered these concepts (e.g., teachers sharing students, impact of student attendance, developing fair and reliable assessments for non-tested areas), we were able to engage the stakeholders in in-depth conversations and activities. For example, part of the work of these teams was to engage in the concept of identification of covariates that could be used to include in the VAM (e.g., prior test scores, ELL status, mobility). Lists were created and then compiled across districts to look for commonalities.

2) Cross-District DMATT, Advisory Council and Superintendent & Loads Meetings: These teams also helped inform the work of the VAM. A cross-district DMAT contained representatives from each district, which allowed for the extension of common understandings within districts to common understandings across districts. This team also dealt with related issues in support of VAM implementation including the critical element of student-teacher linkages. This group helped support the districts' involvement in the state's Course-Walk Pilot Project, becoming the state's first districts to go through the process of mapping their course catalogs to the state's master course catalog, which was the first step in automating the student-teacher course connection for future years. Advisory Council Members, superintendents, and other central office staff (as well as transition team members) participated in activities to begin to make meaning of REIL's concepts of a "REIL Score"

Small Group Calibration Exercise

Assign a Designation of Effective to the following schools.

Designation	Ineffective	Partially Effective	Effective 1	Effective 2	Highly Effective
Account	\$0	\$0	\$3000	\$4000	\$5000

REIL Growth Composite Rating

Observation Composite Rating	1	2	3	4	5
1					
2					
3					
4					
5					

POPULAR CONTENT

- Value-Added Paper (1,735)
- Uplift Education - Synthesis of Domain Standa... (1,165)
- Data Capacity and Quality Considerations Pres... (1,063)
- Training Principals to promote growth (550)
- Transition from traditional salary scale (467)
- I am looking for anyone who has scoring put t... (447)

Grantee Spotlight

- FAQs
- Resources

HOT DISCUSSIONS

- Data Quality
  - Replies: 1 | Views: 36
- Principal Evaluation Standards and Rubrics
  - Replies: 3 | Views: 23
- Performance-Based Payouts, AZ
  - Replies: 3 | Views: 91
- Sustainability
  - Replies: 1 | Views: 26
- Sustainability

2013 TIF Community

and the resulting performance award continuum. In November 2011, David Stuit from BPR facilitated a special session with REIL superintendents and leads on the calibration of the REIL Score and the performance award continuum. Key questions for this session included: "What does it mean to be "effective"?" and "Who should be eligible for performance awards?" During this event, the superintendents and leads engaged in an activity informed by the New Haven Public Schools Summative Evaluation Rating Matrix. This involved labeling each cell on a matrix (25 cells) with 1 of 5 performance classifications. This activity was also replicated in each district with the data management and assessment transition teams, as well as other transition teams. Policy questions were identified and addressed (e.g., Can/should outstanding performance in classroom observations make up for average or less-than-average performance in value-added and vice versa?). BPR used this information to create REIL Score prototypes (see example below) for ongoing discussion, and to finalize the methodology behind the calculation of the REIL Score. During January-April 2012, the cross district human resources team was also supported in developing understanding of the performance award continuum and how this would align to the state's teacher evaluation reporting requirements. A transition plan was developed to assist REIL districts in reporting performance classifications of teachers prior to implementation of the REIL Score.

**Guidebooks**

We have Guidebooks in place for the 2012-13 school year. You can access these on the MCESA website at: <http://mcesa.schoolwires.net/Page/324>

**Inquiry: How will VAM be used for non-tested grades and subjects, and what exams are used to calculate VAM in the tested grades and subjects?**

For non-tested grades and subjects, the model will be the same as for tested grades and subjects. Galileo/ATI is used by all of the participating districts, and we participated in their pre- post-piloting of instructional effectiveness assessments. The chart below represents the grades and subject areas that will be represented.

	K	1st	2nd	3rd	4th	5th	6th	7th	8th
Reading	X	X	X	X	X	X	X	X	X
Math		X	X	X	X	X	X	X	X
Science						X	X	X	X
Writing						X	X	X	X

In addition, assessments are being developed for other areas (see chart below) through a partnership with WestEd.

	3rd	4th	5th	6th	7th	8th	9th	10th	11th	12th
Art	X					X	X	X	X	
Physical Education	X					X	X	X	X	X
Music	X									
Choir					X	X				
Band					X	X				
Dance							X	X	X	X
Theatre							X	X	X	X
Social Studies				X	X	X	X	X	X	X

**Inquiry: In order to effectively gain knowledge and understanding about VAM, what kind of professional development did the TIF staff: teachers, principals, teacher leaders, and other TIF staff receive on the VAM.**

**Background Information:**

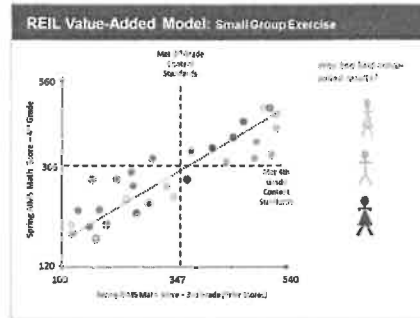
REIL educators will receive their first performance incentives in fall of 2013. Teachers will receive payouts based on their REIL Score, which will be based on observation scores (75%) and school-level VAM (25%) for SY 2012-13. The following two years, teachers will receive payouts based on a REIL Score that will be based on observation scores (50%), individual VAM (40%), team-level VAM (5%), and school-level VAM (5%) for SY 2013-14/2014-15. Significant professional development related to VAM is still up and coming.

**Current PD:**

Professional development on the VAM, up until this point, has been incorporated in multiple ways. MCESA and district-supported Roadshows have been used to canvass large groups of educators in school settings. A recent example includes school-level roadshows where teachers and principals practiced calculating a REIL Score using a scorecard activity where they had to use growth scores and point ranges to learn about the composite scoring process. During roadshows, educators have also learned about the covariates in the model and were exposed to

the guidebook and where to locate this resource on the website. Our governing board members, central office staff, MCESA staff, and advisory council members have also learned about the VAM through our annual governing board symposiums. Many of the presentations from these are on the website. [http://mcesa.schoolwires.net/Page/328]. We also recently hosted a Principal Symposium where our district principals came together and received training on the VAM. We also consider the work that is done in transition teams as professional development. Transition team members engaged in chapter studies and activities with *Value-Added Measures in Education: What Every Educator Needs to Know* by Douglas Harris. They also received presentations on growth measures, achievement vs growth, and comparing like groups of students, things to consider, common misconceptions, covariates, etc. One example of an activity included having the participants look at a scatterplot representing students of three sample teachers and then try to identify the teacher that had the best value-added results (see picture of slide below). This activity really helped people understand what value-added results would mean.

In addition, including stakeholders in the assessment development process is another way to build understanding related to VAM. Teachers have been included in the assessment development process, which includes activities such as development of blueprints and item specifications, writing of items, bias review, field testing, etc.



**Inquiry: How is VAM weighted in the overall Educator Evaluation System since it is one of the multiple measures used to determine teacher and/or principal effectiveness?**

The tables below show the weighting for SY 2012-13, which includes only a school-level growth measure, and for the next two years, which add in team-level and individual-level growth measures.

**SY 2012-13**

	Classroom-Level Growth	Team-Level Growth	School-Level Growth	Observation*
Teachers	NA	NA	25%	75%
	Individual- Level Growth		District-Level Growth	Observation
Building-Level Administrators	45%	5%		50%

Note: For SY 2012-13, K-12 teachers will receive reading and math scores on the 1 to 5 scale. For teachers in schools that administer both AIMS/SAT-10 and Galileo K-12 (e.g., K-5, K-8), the average of their reading and math scores (rounded to the tenth decimal) will account for 15 percent of the REIL Score in 2012-13, with Galileo growth results making up the remaining 10 percent of their school growth component.

**SY 2013-14 and 2014-15**

	Classroom-Level Growth	Team-Level Growth	School-Level Growth	Observation*
Teachers	40%	5%	5%	50%
	Individual- Level Growth		District-Level Growth	Observation
Building-Level Administrators	45%	5%		50%

**Inquiry: What were your greatest successes in determining, piloting and implementing your VAM?**

1) We spent significant amounts of time creating communication tools that allow us to always have printed materials and online resources to support stakeholder understanding and outside agency understanding as well (e.g., Arizona Department of Education, TIF, contractual services). We have guidebooks, handbooks, roadshow materials, webcasts, etc. The effort spent putting these together (which is massive) has paid off in the end. [Most of these materials are posted on our website].

2) Partnering with contractual services can lend much needed support and credibility to the project. For example, our work with WestEd, which had prior experience developing assessments, resulted in the development of quality procedures and practices (e.g., field testing, bias review) to ensure quality assessments. BPR has worked extensively to support the VAM work, consulting with national people in the field and conducting the necessary research to support our choices.

3) Including teachers in the assessment development process (as opposed to solely using contractual services to do this work) has built trust in the overall system and has developed a cadre of teacher leaders to prepare for future assessment delivery in non-tested areas.

4) Simplifying the complex VAM into steps (BPR identified "5 steps" to help our educators understand the concept) helps build acceptance of the model, and a more "technical manual" can be developed and available for those who need them.

5) Gathering cross-district input on the selection of covariates for the VAM was a very productive activity that allowed stakeholders to engage in dialogue about factors that would be considered when "leveling the playing field" for students. Through this process, stakeholders learned what the state was already accounting for in the state-administered AIMS assessment and identify possible additions for future inclusion. A VAM was run during the 2011-12 school year that included the following covariates: prior test score in the same subject as the post-test, prior test score in different subject than post-test, student race/ethnicity, gender, free and reduced-price lunch status, special education status, Limited English Proficiency, mobility, migrant status, homeless status, and student attendance from the prior school year. These covariates were measured at the student and classroom levels.

6) Our goal in designing the REIL measurement system was to ensure that our measurement approach aligned with REIL stakeholders' conceptual definition of effective teaching. So our first step was to understand stakeholders' preferences and priorities regarding what should be required for a teacher to be considered ineffective, effective, etc. We used the matrix activity to gather this information and then proceeded to design a system for aggregating the measures in such a way as to accurately reflect their collective "definition". The stakeholders made clear that they desired an approach to aggregating the measures that allows full compensation such that low performance on one measure could be offset by high performance on another measure. This is why we are using an additive aggregation (weighted sum) rather than a criterion matrix or some sort of multiplicative aggregation. They also made clear that the majority of the overall performance score (REIL Score) should be driven by measures of teachers' individual performance, but also reflect their contributions to their team and school. Stakeholders also indicated they wanted each measure to have clear objective performance benchmarks so that teachers can determine what was required in order to earn a performance award and be designated as effective or highly effective. This is why we convert all measures to the 1-5 scale instead of just normalizing the measures, which would create moving targets. The criteria for the 1-5 scale of each measure were calibrated using prior data so the scales are approximately interval, which makes the "weighted sum" aggregation more defensible. We lose some information by converting continuous measures to the 1-5 scales, but our sensitivity analyses shows this has minimal consequences on teachers' overall results. Moreover, we gain transparency in that teachers can "back out" their scores at the end of the year and can monitor their progress on each measure during the year. This also makes it easier to add new measures down the road.

7) The alignment of VAM with a categorical growth measure has allowed us to include pre- and post-assessment results for a wider group of teachers as part of the VAM (see the answer to question 2, as well as the guidebook, for more information).

8) Our educators have really embraced the concept of "growth" vs. "achievement" and looking at an "apples to apples" approach as far as what they have control over within the classroom. At first, the media that is out there cautioning against value-added affected their understanding of the concept. Now they have a better understanding and like the fact that the model accounts for factors outside of their control. It also resonates with educators that the VAM speaks to a focus on moving all students and not just focusing on a specific proficiency level via an "attainment" score.

### **Inquiry: What were your learned lessons in determining, piloting, and implementing your VAM?**

Lessons we have learned include:

1) Building-level administrators are crucial communicators with respect to everything, including the VAM. It is critical to make sure your principals and other school leaders can explain the principles and concepts underlying the VAM in order to build acceptance among teachers. We recently held a principal symposium where we took principals through a series of activities, including calculation of a REIL Score, so they would feel more equipped to support and generate teacher understanding at their schools.

2) School districts use legal services that sometimes advise them in ways that are unaligned to current education reform efforts. We are currently exploring some ways that we might engage with our district's varying legal counsels so that we can develop some common understandings about education reform in the context of statutory requirements.

3) Determining methods to assist teachers and principals in projecting individual student growth trajectories would help educators' effectiveness when it comes to student growth and would help develop educator's understanding of the VAM.

**Inquiry: What insights, suggestions, and cautions would you share with school districts and states that are at the infancy stage of engaging in this work to use VAM as a component of their redesigned Educator Evaluation System?**

We have several comments that we think might help inform others engaging in this work.

- 1) Procure contractual services or engage in partnerships to assist with the work and to establish or enhance credibility with stakeholders.
- 2) If you can, participate with a consortia or use statewide assessments to ensure there is a large enough student and teacher count to effectively attribute student growth results to educators and effectively calculate value-added scores.
- 3) Develop a common language among stakeholders as soon as possible. Words like growth, progress, achievement, value-added, and formative and summative assessment have different meanings to different people. Identification of specific resources to help define and refine these definitions is critical. For example, VARC's *Oak Tree Analogy* presentation was extremely helpful because it takes a concept and develops meaning outside of the educational context. As another example, an article we used was "Seven Misconceptions about Value-Added Measures" by Douglas Harris.
- 4) Acknowledge the strengths and challenges of value-added and other models for evaluating educator effectiveness. Grappling with these builds understanding and helps people better communicate with additional stakeholders. As an example, our MCESA team grappled with the specific differences between growth models and VAM.
- 5) Specific assessments used to measure educator effectiveness is a critical piece of the VAM work. We realized within the first couple of months that a major part of the VAM discussion is actually what assessments will be used as part of the model. And since we were going to include individual-level growth results for non-tested as well as tested grade levels and subject areas, we had to engage in conversations about what that would look like. Assessments had to have "face validity" across all types of assessments.
- 6) Establish data sharing agreements as soon as possible among all appropriate parties. We had to establish data sharing agreements between MCESA and each district, between MCESA and Arizona Department of Education, between MCESA and Basis Policy Research, between Basis Policy Research and Arizona Department of Education, and between the district's assessment vendor(s) and MCESA/ADE/BPR.
- 7) There are a lot of pieces to data extraction and information needed from districts. You have to get way out in front of it and know who the appropriate contacts are.
- 8) Empowering teachers and leaders to identify students and the percentage of instructional time helps build confidence in calculations and improves acceptance of the VAM.
- 9) Everyone knows that communication is important, but you have to think outside the box about how to communicate effectively and frequently. And don't assume your communication is getting where it needs to go. Communicate ahead of time and at all stages.
- 10) Feedback and input on the VAM can be collected in many ways, some more effective than others, and strong facilitation of the work is optimal. We recommend in engaging in authentic processes that will assist you in driving the work forward and uncovering scenarios that help you avoid pitfalls. For example, when we engaged in the *Stump the System* activities, we were able to identify issues that we may not have been able to otherwise. For example, what if a teacher teaches four sections of ceramics and one section of algebra? What if a teacher teaches a specialized course, and he/she is the only one?

**Related Communities:** Value-Added and Other Measures

Tags: [Value-Added and Other Measures](#)

LATEST COMMUNITY CONVERSATIONS

	4 weeks 1 day ago	TIF 3 grantees- please share information about your experiences with the evaluation...
	2 weeks 22 hours ago	TIF Spotlight Greetings TIF Grantees! It is an honor for us to have the opportunity...