**SPECIAL ISSUE**
**Value-Added: What America's Policymakers Need to Know and Understand**

education policy analysis archives

A peer-reviewed, independent, open access, multilingual journal

epaa aape

Arizona State University

Volume 21 Number 9 January 31st 2013 ISSN 1068-2341

# Sentinels Guarding the Grail: Value-Added Measurement and the Quest for Education Reform

*Rachael Gabriel*
University of Connecticut

*Jessica Nina Lester*
Washington State University
United States of America

**Citation**: Gabriel, R. & Lester, J. N. (2012). Sentinels guarding the grail: Value-added measurement and the quest for education reform. *Education Policy Analysis Archives,* 21(9). This article is part of EPAA/AAPE's Special Issue on *Value-Added: What America's Policymakers Need to Know and Understand*, Guest Edited by Dr. Audrey Amrein-Beardsley and Assistant Editors Dr. Clarin Collins, Dr. Sarah Polasky, and Ed Sloat. Retrieved [date], from http://epaa.asu.edu/ojs/article/view/1165

**Abstract**: Since the beginning of the federal Race To The Top grant competition, Value-Added Measurement (VAM) has captured the attention of the American public through high-profile media representations of the tool and the controversy that surrounds it. In this paper, we build upon investigations of constructions of VAM in the media and present a discourse analysis of the policymaking process within the meetings of Tennessee's Teacher Evaluation Advisory Committee (TEAC), a 15-member panel appointed by the Tennessee governor to develop a new teacher evaluation policy under Race to the Top. The data included audio recordings of public meetings from March, 2010 through the end of the committee's work in April, 2011. As we analyzed the talk of the TEAC, we oriented to the particular version of VAM worked up within these conversations in relation to a descriptive metaphor in which VAM is compared to a "sentinel of trust." We present examples to illustrate three patterns in the construction of VAM as the sentinel of trust within teacher evaluation: (1) VAM alone defines effectiveness; (2) VAM is the only objective option; and

Journal website: http://epaa.asu.edu/ojs/
Facebook: /EPAAA
Twitter: @epaa_aape

Manuscript received: 7/6/2012
Revisions received: 12/2/2012
Accepted: 12/31/2012

(3) concerns about VAM are minimized. We discuss the implications of this way of thinking and talking about VAM and contrast it with other possibilities, including those constructed by teachers, researchers, and the media.
**Keywords**: value-added measurement; teacher evaluation; discourse analysis; policy making.

**Centinelas guardando el Santo Grial: Medida sobre el valor añadido y la Búsqueda para reformar la educación**

**Resumen**: Desde el principio de el programa federal *Carrera a la Cumbre* (*Race to the Top*) los modelos de valor añadido (MVA) han capturado la atención del público americano a través de presentaciones sobre el instrumento y los debates que lo rodean en medios de comunicación prominentes. En este trabajo, nos basamos en investigaciones sobre la generación de MVAs en los medios y presentamos un análisis de discurso del proceso de generación de políticas dentro de las reuniones del Comité asesor de Evaluación de Profesores de Tennessee (TEAC), un panel compuesto por 15 miembros designados por el gobernador de Tennessee para desarrollar una nueva política de evaluación docente bajo el programa *Carrera a la Cumbre*. Los datos incluyeron grabaciones de audio de reuniones públicas desde marzo de 2010 hasta finales del trabajo del comité en abril de 2011. Cuando analizamos los discursos del TEAC, nos orientamos a la versión particular de MVAs desarrollado dentro de estas conversaciones con relación a una metáfora descriptiva en la cual MVAs es comparado con un "centinela de confianza." Presentamos ejemplos para ilustrar tres modelos en la construcción de MVAs como centinela de confianza dentro de la evaluación docente: (1) MVA solo define la eficacia; (2) MVA es la única opción objetiva; y (3) las preocupaciones por MVA son minimizadas. Discutimos las implicaciones de esta forma de pensar y hablar de MVA y la contrastamos con otras posibilidades, como las construidas por profesores, investigadores, y los medios de comunicación.
**Palabras clave:** modelos de valor añadido (MVA); evaluación docente; análisis de discurso; elaboración de políticas.

**Sentinelas custodiando o *cálice*: medidas do valor agregado e a busca de reforma da educação**

**Resumo:** Desde o início do programa federal *Corrida para o Topo* (Race to the Top) os modelos de valor agregado (MVA) capturaram a atenção do público americano através de apresentações do instrumento e debates em torno dele na grande mídia. Neste artigo, nos concentramos nas pesquisas sobre os MVA na mídia e apresentamos uma análise do discurso do processo de geração de política dentro das reuniões do Comitê Consultivo de Avaliação do Professor do Tennessee (TEAC). O TEAC é um painel de 15 membros nomeados pelo governador do Tennessee para desenvolver uma nova política de avaliação de professores de acordo com o programa *Corrida para o Topo*. Os dados incluem gravações em áudio de reuniões públicas entre março de 2010 e o fim do trabalho da comissão, em abril de 2011. Ao analisar os discursos de TEAC, vamos nos concentrar na versão particular de MVA desenvolvida dentro dessas conversas a respeito de uma metáfora descritiva em que MVA é comparado a um "sentinela de confiança." Nós apresentamos exemplos para ilustrar três modelos na construção MVA como sentinela de confiança em avaliação de professores: (1) MVA define sozinho a eficácia, (2) MVA é a única opção objetiva, e (3) preocupações MVA são minimizados. Nós discutimos as implicações deste pensar e falar sobre MVA e contrastamos com outras possibilidades, como aquelas construídos por professores, pesquisadores e pela mídia.
**Palavras-chave**: modelos de valor agregado (MVA); avaliação de professores; análise do discurso; o desenvolvimento de políticas.

## Introduction

Since the beginning of the federal Race To The Top grant competition, Value-Added Measurement (VAM) has captured the attention of the American public through high-profile media representations and the controversy that surrounds its use (Amrein-Beardsley, 2012; Baker et al., 2010; Ewing, 2012). Despite controversies surrounding the wisdom and practicality of VAM in high-stakes teacher and program evaluations, federal and state policies have included references and mandates for the use of VAM at an alarming rate. Federal policy statements, including the Race to the Top (RTTT) criteria and the Higher Education Act, have explicitly referenced the use of VAM in high-stakes teacher evaluations, as well as the evaluation of teacher preparation programs (US Department of Education, 2011). Likewise many states have considered or enacted policies that require VAM to be used in annual, high-stakes teacher evaluations and the evaluation of teacher preparation programs.

The use of VAM for teacher evaluation has recently become a key debate in discussions surrounding education reform. In August 2010, *The LA Times* published a series of articles about VAM, along with a database of 6,000 Los Angeles Unified School District teachers' value-added scores, sparking a controversy that echoed on radio, Internet, and newspaper media outlets across the country. This was shortly followed by news coverage of similar attempts to make VAM rankings public in New York City. VAM remains a topic of lively debate in the media as more states have incorporated it into laws regarding evaluation systems as part of RTTT proposals or No Child Left Behind waiver applications. Ewing (2011), President of Math for America, noted that VAM has taken on almost mythic proportions and been (mis)used to intimidate those who do not claim expert knowledge of statistics. He wrote: "As the popular press promoted value-added models with ever-increasing zeal, there was a parallel, much less visible scholarly conversation about the limitations of value-added models," noting that the public has been "bamboozled" by complex mathematics into accepting an "imperfect substitute" for the tool they think VAM could be (p. 672). Ewing also highlighted the alarming disconnect between the scholarly conversations endorsing the cautious and limited use of VAM in high-stakes decisions of any kind, as well as the media frenzy demanding the public release of individual teachers' scores.

We have argued elsewhere (Gabriel & Lester, 2010, in press) that VAM has been presented to the public as a simple and unassailable answer to a range of policy problems, including teacher evaluation, preparation, certification, and tenure. Specifically, we have suggested that the media has constructed a story of contemporary education reform where the "dragons of ineffectiveness" (aka bad teachers) can only be rooted out by "heroes" (aka reformers) that wield VAM the perfect weapon for vanquishing ineffective teachers (Gabriel & Lester, in press). The idea that teacher effectiveness is the single most important school-based factor in student achievement has cast the ability to accurately and decisively root out "the ineffectives" by identifying and selecting effective teachers as being the holy grail of education reform. In this way, VAM has been positioned as an important weapon in the quest for better, more equal schools, with those who disagree with its use positioned as anti-reform, pro-union, and unscientific. Still, there has been minimal exploration of how policymakers, who may be influenced by both the media and research findings, go about talking about and making sense of VAM.

In this paper, we extend our previous analyses of media representations of VAM by attending to the ways in which VAM is made relevant by policymakers themselves in the course of policy development meetings. We examine how VAM was constructed and positioned in the talk of a governor-appointed committee and a state board of education responsible for developing and approving state-based teacher evaluation policies under RTTT. We conducted a discourse analysis of

the meetings of Tennessee's Teacher Evaluation Advisory Committee (TEAC), a 15-member panel appointed by the governor to develop guidelines and criteria for a new teacher evaluation policy under RTTT. The data included audio recordings of public meetings from March 2010 through the end of the committee's work in April 2011.

## A Brief Review of Research

A review of available research on VAM in education can roughly be grouped into investigations around four major methodological questions: (1) error rates; (2) test effects; (3) model biases; and (4) inherent problems of achievement tests. Though value-added estimates may be able to estimate which groups of teachers are more or less likely to have students who achieve more or less on state tests, they seem less likely to be able to reliably and accurately predict which individual teachers will do so.

### Error Rates

As Lockwood, Louis, and McCaffrey (2002) have shown, more than half of teachers in the top ten percent of *estimated* value-added scores, based on several years of data, are actually in that top ten percent (or decile) of value-added scores. There is substantial uncertainty in the estimates of a teacher's value-added, with a large number of classification errors occurring among teachers who serve students of different populations (Goldhaber & Hansen, 2008). Lockwood, Louis, and McCaffrey (2002) concluded that, "student characteristics are likely to confound estimated teacher effects when schools serve distinctly different populations" (p. 256).

Several scholars have highlighted the problem of error rates associated with VAM, specifically if used in high-stakes decision-making (Baker et al., 2010; Darling-Hammond, Amrein-Beardsley, Haertel, Rothstein, 2012; Schochet & Chiang, 2010). While proponents of VAM are quick to point out that any statistical calculation has error of one kind or another, some statisticians have set out to specify the source and nature of error in various VAM models as they have important practical implications. Error in VAM could mean the difference between merit pay, promotion, or job termination for a percentage of teachers every year. For example, Schochet and Chiang reported that error rates for comparing an individual teacher's score to the average score are likely to be 25%, using three years of data and 35% using one year of data. In other words, one in four teachers will be misclassified (and possibly promoted, rewarded, or fired) when using three years of data in the model, and one in three will be misclassified when using a model that considers only a single year of data.

### Test Effects

Researchers have also investigated whether it matters what sort of test is used when calculating a teacher effect score. Lockwood et al. (2009) used data from two different subtests of a mathematics achievement test (procedures and problem solving) to see if teacher effect scores would be similar in both areas. Using 20 different models, they found very low correlations between effect scores generated from each subtest (.01 to .46, with an average of .255). This result either indicates that VAM is an unreliable indicator or that teachers have a different impact on various areas of mathematics. Papay (2011) replicated this study using three different measures of reading achievement and two subtests of math achievement given at three different times during the school year. He found that both the measure and the time of year were statistically and practically different enough to have an impact in a system that used VAM to award merit pay. For instance, if tests or timing were switched, almost 50% of teachers would be in different pay categories. He concluded that, "using different achievement tests produces substantially different estimates of individual

teacher effectiveness" and "that the timing of the test alone may produce substantial variation in teacher productivity estimates across outcome measures" (p. 166).

Corcoran, Jennings, and Beveridge (2011) compared value-added scores of teachers from high-stakes test data and low-stakes test data using eight years of data, and similarly found what they described as an unsatisfactory correlation between them (.52 for reading, .59 for math). The Gates Foundation's Measures of Effective Teaching Project (MET) has similarly compared value-added scores generated from state tests (most often multiple choice) and open-ended tests, also finding low correlations (.37 for Language Arts and .22 for Math). Taken together these studies offer a strong indication that the time, content, and stakes of a test make a statistical and practical difference in a teacher's value-added score. Such sensitivity makes VAM an unstable metric from which to base decisions about promotion, termination or merit pay.

**Model Bias**

One of the larger concerns surrounding the use of VAM is the possibility that some biases are inherent in the statistical models themselves. One source of suspicion for bias is that the scores of minority teachers tend to be stable one year to the next, while VAM in general, is described as being highly variable year to year (McCaffery, Sass, Lockwoord & Mihaly, 2009). The question of bias seems to depend on which variables are included in the model, with models that control for aggregate data, prior test scores, and demographics (e.g. Chetty, Friedman, & Rockoff, 2012; Kane, & Staiger, 2008) showing less correlation to background variables. The nonrandom assignment of students to teachers may also be a source of bias (Rothstein, 2009).

**Achievement Testing**

When it comes to the inherent problems of achievement tests, the old adage "garbage in, garbage out" is a useful summary. The literature surrounding VAM mostly attends to internal issues of the method, while sometimes making a peripheral nod to the underlying assumptions of the data used in VAM and the possibility that such assumptions should not go unexamined. The general consensus within the field, however, followed the same "it's the best we have" logic that is so often echoed in media representations of VAM (Gabriel & Lester, 2010, in press). This logic is one which standardized achievement tests are positioned as the best (and only) available data, and are therefore good enough simply by virtue of being the only option.

## Context of the Study

Tennessee uses TVAAS scores as a *growth measure*, which means that growth is equated with learning (growth in knowledge). A positive value-added score indicates that a student "grew" or "learned" more with that teacher than they do on average. This is based on the assumption that learning or growth (a) are captured in the scaled score of a standardized test, and (b) increase at a consistent rate over time and across developmental stages. Within TVAAS, teacher effect scores are calculated by comparing individual students' test scores over time. Individual students' value-added scores are averaged together to come up with a teacher effect score. It is a measure of expected versus actual performance on standardized tests based on average scores from a three to five year period. The system uses scaled scores from a standardized test (most often state tests in math and reading) and compares the current year's score to the scores from previous years. If this score is higher than the average, there is a positive value-added and the teacher is in the $4^{th}$ or $5^{th}$ quintile of effectiveness depending on how much higher this year's score was from the average. If the score is the same as the previous average, there is zero value-added, which is equated to "a year's worth of growth for a year's instruction," or the third quintile of effectiveness. If the score is lower than the

average of the last few years there is negative value added, which is interpreted to mean that a student "grew" less with that teacher than they have proved themselves able to in other years (1st and 2nd quintiles of effectiveness).

It is important, however, to be clear that this is a *teacher-level* score, not necessarily the effect of a single teacher on a student's trajectory of achievement. This means that a teacher effect score represents the contribution that cannot otherwise be explained by school-level or individual-level factors in the model. Thus there are both methodological and practical problems with linking student scores to an individual teacher. The most obvious is that students are taught skills that are assessed on math and reading tests by multiple teachers every year. In elementary schools there are specialists, content area teachers, support staff, aides, substitute teachers, etc., not to mention out-of-school instruction from parents, tutors or mentors. In middle and high school, students may only be assigned to one math teacher, but may learn and practice mathematics in their science and social studies classes, as well as in out-of-school settings. The extent to which instruction is distributed across adults makes it difficult to hold only one teacher of record accountable for the success or failure of the students on their rosters. In an article called "Where's the action?" Croninger and Valli (2009) explained that 4th and 5th grade reading instruction is found throughout the school day across courses and classrooms. They further noted that the distribution of responsibility for the teaching of reading only widens as high-stakes tests approach and content area teachers turn into reading tutors or test prep facilitators. In many ways, the distribution of responsibility for math and literacy instruction is required for best practices in remediation (e.g. Response to Intervention) and differentiation. Ironically, these best practices for student support complicate, if not confound, efforts to calculate an individual teacher's effect score.

Even if it were clear which teacher should be responsible for a certain portion of a student's score, the actual process of linking teachers to students is administratively challenging, error-prone, and time consuming. Districts in which there is a high degree of student mobility would need to develop systems to calculate how many days each student was present in each teacher's class. Inclusion classrooms with co-teachers would have to decide which students each teacher would claim and which they would share credit for. Thus, the process of linking and claiming students is an ongoing process that requires automated systems, along with teacher reports or confirmation at several points during the year. In the past, Tennessee teachers participated in a "claiming day" at the end of the year to indicate which students' scores should be linked to their names. In some cities, students were linked to teachers only if they were present in school on a specific "linking day" or if they appear on the roster after a certain point in the year. Unfortunately, this makes it easy for certain students to be pushed or left out of calculations. It also contributes to the amount of time it takes for teacher effect scores to be calculated.

One of the first large-scale data systems capable of calculating teacher effect data using VAM was the Tennessee Value-Added Assessment System (TVAAS). TVAAS was developed at the University of Tennessee by then professor of statistics in the College of Agriculture, Dr. William Sanders in the 1990s and is the first iteration of the Education Value-Added Assessment System (EVAAS), which is currently managed by Sanders at SAS, a statistical consulting firm. TVAAS and EVAAS represent one of many possible models for value-added analyses, each with their own assumptions, safeguards, sources of error, and possible biases. EVAAS has since been marketed commercially and used by entire states (e.g. North Carolina, Ohio, Pennsylvania) as well as individual districts across the country. Sanders continues to hold an exclusive contract with the state of Tennessee, which includes access to a database of over 20 years of TVAAS scores (a database that has been called the largest of its kind in the world). Though value-added data has been generated for Tennessee teachers and schools for over twenty years, it was never intended to be an

evaluative tool for individual teachers.

Tennessee's winning RTTT application emphasized their unique database and existing capacity to calculate a teacher's effect on student's achievement. In fact, the phrase TVAAS is mentioned no fewer than 54 times in the text of Tennessee's RTTT application. Since the creation of teacher evaluation policies that would differentiate teachers by effectiveness was worth up to 30% of available points in the RTTT competition, the existence of TVAAS was a major factor in Tennessee's first-round win. The state earned more points in the category labeled "Great Teachers and Leaders" than any other state.

As a show of good faith, and as insurance against reform rollback at the end of his term, then Tennessee Governor Bredesen convened an emergency meeting of the state legislature in order to pass his RTTT application proposals as state law in early 2010. The reforms described in the RTTT application were passed in their entirety as the "First to the Top Act." This act specified the formation of the TEAC, its scope of work, and made the inclusion of TVAAS scores in annual evaluations a matter of state law. The First to the Top Act requires all teachers to be evaluated every year, with 50% of the evaluation made up of student achievement data, at least 35% of which must be TVAAS scores where available. Comparable measures where TVAAS scores are unavailable, as well as the acceptable measures for the "other 50%", specific guidelines, and procedures for use were left to the TEAC to describe.

As of the 2011-2012 school year, Tennessee state law requires teacher evaluations to incorporate TVAAS (for 35-50% of the overall evaluation), but does nothing to address the fact that this data cannot, in its present form, be made available in time for such use. According to representatives of the Tennessee Department of Education, the timeline for generating TVAAS scores based on student achievement data from Spring standardized tests involves a long "lag time" between when the tests are taken, and the scores are returned to districts. Up until the 2011-2012 school year, TVAAS scores for one year were returned several months into the next school year. This lag time is due to a number of steps taken in the collection, cleaning, and data quality confirmation prior to sending test data to SAS for extensive calculations. In order to make "human capital decisions" (hiring, firing, promotion, tenure) or to use TVAAS data to improve practice or inform professional development decisions, TVAAS scores need to be returned within the school year they reflect. Though there are many suggestions for speeding up the timeline, including online testing, an earlier testing period, and different systems for linking and cleaning data, most suggestions would require several years to implement.

Even if scores could be calculated in time to make decisions before a new school year begins, individual teacher effect data currently do not exist for most teachers. In Tennessee, value-added scores are currently only calculated for subjects and grades that are tested under the Tennessee Comprehensive Assessment Program (TCAP) of state standardized tests. These include reading and math tests in 3rd through 8th grade, as well as end of course exams in certain core high school courses. This leaves more than 60% of teachers without individual teacher effect data. In order to use value-added data for the annual evaluation of every teacher every year, new tests will need to be created or adopted in order to generate standardized test data for every teacher's annual evaluation. This would require the creation of more than 300 standardized tests for currently untested grades and courses (Tennessee Department of Education, 2010 ). In the meantime, school-wide averages may be used along with observation scores to calculate an overall evaluation rating for teachers of untested subjects and grades.

While one could argue that school-wide averages might, for example, encourage all teachers in a K-3 school to work together towards higher 3rd grade scores, it becomes less and less meaningful to hold teachers accountable for the scores of students they may not have taught (e.g.,

transfers) or have not been involved with for several years. As with any system that averages across individuals, it may also hold teachers responsible for the extreme performance of outliers in their schools. From this perspective, assigning teachers an averaged score serves as an incentive for principals to stock certain positions (3rd grade) with their most effective teachers, while placing less experienced or less effective teachers into earlier, untested grades.

Given these methodological and practical concerns, we were interested in how policymakers described and made sense of VAM as a tool for individual teacher evaluation. Specifically, we were interested in the ways in which uses of VAM, and TVAAS specifically, came to be taken-for-granted and naturalized aspects of teacher evaluation. In the section that follows, we describe our theoretical and methodological approach to investigating the discursive construction of VAM in the meetings of the TEAC.

## Theoretical Perspective

In order to investigate the discourse surrounding VAM within a policy-making process, we took up a theoretical perspective that privileged the action-oriented nature of talk (Edwards & Potter, 1993). More particularly, we presumed that language, in all its forms, was constitutive rather than representative or reflective of a social reality (Edwards, 1997). Discourse choices can be assumed to always be *doing* something, even while being informed (yet not determined) by broader structures and institutionalized practices. Drawing upon discursive psychology (Edwards & Potter, 1993?) and conversation analysis (Sacks, 1992), we undertook a microanalysis of talk in order to analyze how certain versions of VAM were worked up and made relevant within conversations.

Given our approach to the action-oriented nature of language, we took up a critical social constructionist perspective (Jorgensen & Phillips, 2002), and viewed the process of data analysis as constructive in and of itself. Thus, as we analyzed and interpreted the data, we presumed that:

> …by representing a qualified (that is, scientific) and *different* account of reality from those which are otherwise available, research knowledge can hopefully contribute to the addition of new perspective to public debate. As social constructionists, we do not have the right endowed by possession of a final truth. But we do have the right that all people, in principle, have to intervene in democratic debate with a truth that can be discussed, in order to further our visions for a better society. (Jorgensen & Phillips, p. 211)

Thus, we did not assume that our interpretations of TEAC interactions were a "final truth." Rather, we acknowledge that we are participants in the ongoing conversation that surrounds teacher evaluation, VAM, and teacher effectiveness. We thus position this study, like many other discourse studies, as a reflexive product in that we too are part of the discourse we critique. Accordingly, we recognize that the interpretations we proffer are always situated and exist among many other possibilities (Taylor, 2001).

We oriented to our work from a critical perspective, viewing the very idea of "critical" as an orientation that seeks to distinguish and tease out "complexity", while "denying easy, dichotomous explanations" of a given phenomenon (Wodak, 1999, p. 186). Thus, as we worked to make sense of the data, we sought to deconstruct ideas and assumptions that were positioned as objective and natural, and presumed that each idea was layered, potentially contradictory, and indeed complex. Furthermore, throughout our study, we drew upon ideas from narrative policy analysis (Roe, 1994), specifically the idea that policy issues are often described and deployed in ways that can be identified as following familiar patterns of myths or other cultural narratives. This allowed us to frame our line-by-line micro-analysis of talk within more macro understandings of the social implications of discourse.

## Data Sources

This study is part of a larger investigation of the discursive construction of teacher effectiveness in the public meetings of the Tennessee TEAC (Gabriel, 2011). The first author attended and audio recorded every meeting of the committee from May 13, 2010 through April 15, 2011, as well as the meetings of the State Board of Education in which teacher evaluation was discussed. A total of twelve meetings and one open conference call were attended and recorded with the permission of the committee chair. This constituted approximately 24 hours of recorded data. Since these meetings were considered open, public events, no pseudonyms were used in accordance with our Institutional Review Board's guidance. Under the Tennessee Open Meetings Act of 1999, all public meetings must be announced to the public and the media in advance, with minutes taken by a public servant and made available to the public at large. These minutes, along with handouts and presentation materials, were all posted online at the First to the Top Act website (http://www.tn.gov/firsttothetop/). Handouts and other presentation materials are considered supporting documents, but are not representative of the committee talk and therefore were not included within our data set.

**Data Analysis**

We employed a form of discourse analysis (Woods & Kroger, 2000) situated within a discursive psychology framework (Edwards & Potter, 1993) informed by conversation analysis to some extent (Sacks, 1992). Our analysis was iterative and emergent, involving a four-step process. First, using the transcripts of the TEAC meetings that were transcribed for an earlier analysis (Gabriel, 2011), we selected extracts of discussions that specifically related to VAM and TVAAS. We collected all of those extracts specific to discussions about VAM and TVAAS, thereby narrowing the focus of our analysis. This included approximately 7 hours of conversation. Second, we created theoretical and analytical memos for the extracts of focus, taking note of the discursive patterns. As we created these memos, the following three analytic questions informed how we made sense of the data: (1) What is the action orientation of this discourse? (2) How is the discourse constructed to do this? and (3) What discursive resources are present and currently used to perform this activity? (Potter, 2003; Wood & Kroger, 2000). The memos served to create an audit trail (Creswell & Miller, 2000) of our theoretical and analytical ideas and decision-making process. Third, we engaged in a line-by-line analysis of the extracts of focus, drawing upon the analytic tools of conversation analysis (Sacks, 1992). This analysis resulted in the production of three overarching patterns. As we analyzed the extracts, we assumed that variability existed within the discourse, and sought to identify alternative interpretations (Antaki, Billig, Edwards & Potter, 2003). Fourth, we generated explanations linked to the three overarching patterns, drawing upon narrative policy analysis (Roe, 1994), as well as existing studies of the media constructions of VAM (Gabriel & Lester, 2010, in press; Ewing, 2011).

We pursued several strategies to validate and warrant our claims. First, we acknowledged that our interpretation is situated and partial, standing as one of many interpretations (Taylor, 2001). We used Atlas.ti (Muhr, 1993) to systematize and make the analysis process more transparent. Third, we intentionally sought out alternative cases and explanations (Potter, 2004), with the hope of deepening and verifying our interpretations. Fourth, throughout our discussion of the findings, we supported our interpretations with a detailed line-by-line analysis of each extract. As is the case of many studies claiming to draw upon discursive psychology and/or conversation analysis, we worked to transparently support each of our claims with detailed explanations of the selected extracts. Fifth, in that discourse analysis studies place great emphasis upon reader evaluation (Potter, 1996), we reproduce below several illustrative extracts. These extracts were drawn from a larger set of extracts.

The extracts we included are intended to highlight the ways in which a particular pattern was made evident in the talk. Within the transcripts, we included a 'light' layer of Jeffersonian transcription symbols that served to highlight the ways in which the conversations unfolded at a micro-level. Rather than assuming that pauses, interruptions, and overlapping speech, for instance, are meaningless, we assumed that the micro-features of the talk were potentially significant and therefore made efforts to represent them in each included extract. Further, because we grounded our claims in the talk itself, we included line numbers so that we could more easily refer back to the actual conversational turn(s) to which we were referring. From a discursive psychology perspective we did not aim to analyze the speaker's intent; rather, we aimed to highlight how the discourse choices produce particular versions of the world. We invite readers to analyze these extracts along with us in an effort to make our own analysis transparent and to leave room for other possible interpretations. Finally, we carefully considered Tracy's (1995) argument that a fruitful discourse study offers "productive ways to reframe old issues," creating "links between previously related issues," and raises "new questions" (p. 210). As such, as we share our findings, we aim to articulate our claims in relation to the broader scholarly discussions surrounding VAM.

## Findings

In the meetings of the TEAC, the four major methodological questions (e.g. error rates, test effects, model biases, and inherent problems of achievement tests) surrounding VAM were rarely, if ever, made relevant. Instead, discussions about TVAAS centered around a set of practical problems which included: (1) the problem of efficiently and accurately linking individual students with the teachers deemed responsible for their progress; (2) the lagging timeline required for valid VAM data; (3) weighing and comparing VAM scores with other measures included in the overall evaluation instrument (observation, achievement scores, surveys etc.); and (4) the lack of individual teacher effect data for more than 60% of teachers.

Throughout discussions within the TEAC and related presentations to the Tennessee's State Board of Education (SBOE), VAM was constructed as being the singular, stable, objective core of any teacher evaluation. The claim of objectivity was all but unquestioned by the committee, as this objectivity was cast as the only legally defensible, decisive factor within the teacher evaluation instrument. Similarly, throughout the conversations at the TEAC meetings, the words "statistics", "research", "objective", "valid", and "reliable" were positioned as being synonymous and unquestionably good. On the other hand, words such as "subjective", "opinion", and "qualitative" were positioned as questionable and difficult to defend. As we analyzed the talk of the TEAC, we began to orient to the particular version of VAM worked up by the committee in relation to a descriptive metaphor in which VAM is compared to a "sentinel of trust." The phrase "sentinel of trust" (Miller, 2012) was first used to describe the role of VAM in media representations, yet we found it to be useful for describing the committee's construction and positioning of VAM as an evaluation tool.

The metaphor of a sentinel is, in this case, particularly apt both because of the popular use of the word to describe a guard and the Latin root of the word, *sentire*, literally meaning *to feel.* Because of its etymology, the word sentinel has been used in science fiction to describe mythical creatures with enhanced senses that give them the supernatural power to know the future or essence of what is before them. In the three sections that follow, we present examples to illustrate three patterns in the construction of VAM as the *sentinel of trust* within teacher evaluation. These include patterns in which: (1) VAM alone defines effectiveness; (2) VAM is the only objective option; and (3) concerns about VAM are minimized.

**VAM Alone Defines Effectiveness**

Across the TEAC meetings, we noted that VAM and other quantitatively-oriented measures were positioned as the "gold standard" of "objectivity," and "objectivity" the goal of teacher evaluation. The first extract is taken from a discussion about labels for different rating levels of teacher effectiveness that can be earned in the new evaluation system; this extract is one of many examples in which VAM was positioned as being the primary definer of teacher effectiveness. Tennessee state law requires that a teacher's TVAAS score, if available, will be used for at least 35% of their overall evaluation score, with 15% based on some other measure of student achievement. The other 50% can be a qualitative component, such as an observation or survey. In Extract 1, Katie Cour, a consultant from Education First who was hired to facilitate meetings, explained that in order to be at the "needs improvement" evaluation rating level, the 35% TVAAS score must be low. A state senator and two teachers (Judy and Pam) continued the conversation about how the 35% TVAAS component works to balance out other variables. (See Appendix A for transcription symbols.)

**Extract 1**

SENATOR GRESHAM: So is highly effective, effective, or (in)effective going to be predicated on what those what those scores are what that TVAAS data shows and what and what uh other specific mathematical [uh
KATIE: Yeah] so it'll be the so yeah so you'll if I'm a highly if I'm a needs improvement teacher that means that 35% of my evaluation in TVAAS was () you know not not that great
SENATOR: [Yeah but then you have to then you have to
KATIE: It all incorporates in]
SENATOR: Define what not that great is
KATIE: Right
SENATOR: Right
KATIE: And we can't do that now this is [um
SENATOR: Right okay]
KATIE: That's that's something that will have to happen after we've we've field tested and we know what we're looking for
JUDY: But you could be a three or a two because you don't show up for bus duty or you don't work well with others, but you might have some really good test scores.
KATIE: Mm hmm
JUDY: But so
PAM: Then that calculation would throw you into the effective hopefully because student student growth and achievement is 50%, correct?
KATIE: Mm hmm
PAM: That's that's not subjective that's objective.
KATIE: That's right

The extract above highlights the way in which the presumed role of value-added in both calculating and validating a teacher's overall evaluation rating was worked up in and through the talk of those attending the meeting. It is worth 35% of the overall evaluation; therefore, one could not earn the lowest "needs improvement" status unless a TVAAS score was also "not that great." In addition, TVAAS is presumed to protect teachers from earning low scores for things such as bus duty or working well with others, activities positioned as less relevant to an overall rating. Though Katie and Senator Gresham agreed that they cannot define the levels of effectiveness now (lines 9-

13), Katie also pointed out that the field test will demonstrate/determine (lines 14-15) what should count as "great" teaching. In other words, once they have had a chance to compare value-added scores with scores from qualitative measures, they will know which measures align with different levels of effectiveness.

As the Senator noted, "…highly effective, effective, or effective going to be predicated on what those what those scores are" (line 1-2). In this case, VAM was constructed as the stable and accurate center of a teacher's effectiveness. This accuracy was implicitly linked to the notion of "objectivity", as Pam noted (line 23). Though the committee members demonstrate agreement (lines 10, 11, 13) that they cannot define what great teaching is, they trust value-added scores to show them what should count as effective teaching (line 23). VAM is thus positioned as guarding the trustworthiness of the entire evaluation instrument.

As Judy stated (lines 16-17), there are several reasons that a teacher might earn a certain score that is sometimes measured by evaluation rubrics. Pam pointed out that even these individual differences would be balanced because TVAAS could "throw you" into the right category (lines 20-21). This description constructed a version of effectiveness that relied upon VAM to identify whether someone is effective or not, and positioned it as a tool that could accurately balance out all other streams of evidence that feed into an overall evaluation. The trustworthiness of the entire evaluation was therefore bound up in the existence and balancing ability of VAM.

The next extract further illustrates how the talk within the TEAC meetings served to position teacher effectiveness as definable only in the presence of value-added scores. In this extract, John Barker, an administrator from the Memphis Public Schools, was responding to a question about whether there is a ceiling effect in which a teacher or school could "top out" of VAM. Committee members had reported hearing concerns that VAM would be biased against teachers of already high-performing students because they would not grow at greater rates each year. They suggested that students might reach a "ceiling" on their rate of growth, which would make teachers look as though they were less effective when it actually merely reflects consistent learning at a high rate. Barker addressed this concern by giving a specific example of an "optional school" in his district that functioned as a magnet school with strict admission requirements and an enrichment and college preparatory curriculum.

**Extract 2**

1 BARKER: The idea of value-add being topped out I think you get that
2 from a lot of people across the way. But I've got a school in Memphis, White Station
3 Middle School, and you check it out on the website. It's White
4 Station Middle school, big school. One thousand kids grades six through eight. And
5 what they do is they take kids who are optional scoring kids and take them even
6 higher. They have one of the highest TVAAS scores in the state. So now we've got
7 TVAAS from Mathematica but value-added is value-added. And the way that this is
8 measured they have A's and B's in value add and they're taking optional kids who
9 are absolutely projected at 26 and 27 on the ACT and taking them up. So it's
10 sometimes a spurious argument and I will have to say there is sometimes a
11 ceiling effect and I will absolutely give you that but there are some schools that
12 are really doing that. Come down and see what they're doing. They got teachers
13 who are outstanding.

In Extract 2, Barker used a concrete example to describe a teacher as outstanding based on their value-added scores. Yet, similar to the discussion in Extract 1, John failed to describe or define teacher effectiveness outside of value-added scores. In everyday conversations, personal stories and

experiences are often used to build a case, as they are difficult to question and verify (Barnes, Palmary, & Durrheim, 2001). The use of this concrete example makes it difficult to argue that the committee does not need to worry about the possibility a ceiling effect. Barker also used extreme case formulations (Edwards, 2000; Pomerantz, 1986) to emphasize the exceptionality of this example, creating an even stronger argument for ignoring the possibility that ceiling effects would undermine the validity of VAM. Extreme case formulations (ECF) are often used to bolster the factuality of someone's claim and are typically employed when speakers anticipate a counter-claim. Adjectives and adverbs like "even higher" (line 5), "highest" (line 6), "absolutely" (line 9), "outstanding" (line 13), "really" (line 12), and a phrase like "optional scoring kids" (line 5) (a label for students who score high enough to be admitted to an enrichment/college preparatory magnet school) are each examples of an ECF. This particular discursive feature makes the example of White Station Middle School appear even more powerful, as it is the most extreme of examples. Therefore, in this way, the admission that a ceiling effect sometimes exists (line 11) can be said without seeming to contradict the argument for the use of VAM.

The admission, "I will absolutely give you that" also contains the adverb "absolutely", but with a different effect. The "absolutely" in line nine functioned to emphasize Barker's confidence that students at this school would score well on the SAT, constructing White Station Middle as an extreme case. In line 11, the word "absolutely" emphasized his willingness to admit the possibility of a ceiling effect, which worked to position him as reasonable. The words, "I will…give you that" (line 11), positioned the speaker between the audience and the admission of the possibility of a ceiling effect. In other words, he will "give" credence to the possibility of a ceiling effect, even though he has a strong example to the contrary. This constructs Barker as reasonable because he has specific knowledge, along with a willingness to accept multiple possibilities. In previous statements, Barker also alluded to his doctoral degree from Vanderbilt and the statistics classes he took in the very building where the committee meetings were held, positioning him as an expert of sorts. His admission that ceiling effects are possible worked to increase his trustworthiness as an expert on statistics, with both academic and practical knowledge about the topic. Though the admission that there is "sometimes a ceiling effect" contradicts his example, it makes him seem more trustworthy.

Extract 2 also captures one of many examples of logistical difficulties with VAM, as delivered by TVAAS. The data Barker presented was generated by Mathematica Policy Research as part of a grant-funded contract, but similar data were not available in the same format from SAS. He argued elsewhere that individual teacher effect data is vital to teacher evaluation and policy decisions, but was not being made available in its most useful form from SAS. Still, in Extract 2 he noted that "value added is value added." This equivalency or tautology is one of many examples across the data set in which speakers set up an equivalence (e.g. "it is what it is" or "call a spade a spade") in order minimize other possibilities. In this case, as in others, the phrase "value added is value added" worked to construct a version of value-added that is stable and unassailable – rather than variable and error-prone. Though different companies may arrive at a value-added score in different ways, or report it in a different format, phrases like "value added is value added" indicate that such differences do not affect the essence of the measurement: it is what it is either way you calculate it.

Ironically, researchers have demonstrated that all methods for generating and reporting value-added scores are not equivalent, and that there can be large practical differences when scores are calculated based on different tests (Lockwood et al, 2009) at different times of year (Papay, 2011), and by different statisticians or organizations (Briggs & Domingue, 2011). The equivalency phrase (line 7) also works to naturalize a certain version of VAM by inserting it into a familiar phrase/structure. It is similar to the saying "it is what it is", which implies that something exists in a certain way, and either defies or does not require explanation. Within this extract, Barker's

combination of extreme case formulations, specific stories, and tautologies about value-added seem to successfully ward off the question of validity, as the ceiling effect was never again mentioned in a TEAC meeting.

**VAM is the Only Objective Option**

This notion that VAM is the only stable, defensible center of a teacher evaluation was also highlighted by SBOE when the TEAC's recommendations were brought up for a vote. On the subject of effectiveness rating levels (as discussed in Extract 1), one board member scoffed: "I think that you know either a teacher is effective or they're not so those five levels are kind of uh subjective to say the least." This statement worked up a version of effectiveness that is binary in nature, rather than continuous (you are or you are not effective, no in between), as well as a version that can be definitively measured. This binary construction is also found in media representations of VAM (Gabriel & Lester, 2010, in press), and is used to position VAM as definitive enough to do what many are afraid to do: make a definite ruling on teacher effectiveness, rather than accepting a gray area or acknowledging room for error. This binary construction of VAM is also used to position the speaker as willing to take a risk in labeling a teacher ineffective, rather than hiding behind uncertainty because of weakness or conflicting sympathies (Gabriel & Lester, in press).

According to this board member, any attempt to create levels beyond effective/ineffective is described as subjective, and thus carries less weight. The phrase, "to say the least", makes it difficult to argue that levels are subjective by implying that they are something more than subjective. This construction of subjectivity as something that is negative and undesirable in an evaluation tool was a common pattern noted across the meetings of the TEAC and the SBOE.

Extract 3 is another example of this version of subjectivity taken from a SBOE meeting in which the final policy was discussed and approved. It includes a comment by one board member, Dick Ray, to another board member, Gary Nixon, who was also a TEAC member and served as their representative to the SBOE.

**Extract 3**

RAY: Just a kind of an overall observation. Do you outside of the those teachers that are uh affected by TVAAS you have a hundred percent subjectivity on this evaluation so especially if you're a librarian, band director or whatever and even those that are affected by TVAAS you have well up to you have up to sixty-five percent subjectivity. Do you have any problem with the amount of subjectivity that's involved there and whether or not that creates more havoc than it does good
NIXON: I think if you reflect back to the current practice it's almost 100% subjectivity. Um the I think the key as as we approve 15% of other criteria is that that not become a larger percent that that stay a focused list of items that will help to drive student performance not a long list of things that are in fact subjective. Um from the 50% that's in the qual in the qualitative evaluation of the house I I don't know exactly how you would try to remove subjectivity I think that we'll you'll look at the process and you'll see the field test results they will be certifying evaluators and observers they can implement with fidelity () observation uh but I still think that about half will be subjective I I agree.
UNIDENTIFIABLE MALE VOICE: Mr. Chairman
CHAIR: Yes sir
MALE VOICE: I just want to go on record uh I'll kind of follow up here with Mr. Wright I this will be a step in the right direction but I have great concern about the

20 small portion of this that is absolutely objective. I know how these processes work. I
21 know people that are observing and uh judging effectiveness on a subjective basis for
22 people that they work with every day and live maybe as neighbors uh I have great
23 concern for it. Hopefully this will work out. No question about it it's better than
24 what we had I mean just about everybody will agree to that but um I have great
25 concern it's gonna get us to a true measurement of performance with the current
26 formula.

In Ray's opening statement, he characterized all components of the evaluation tool besides the minimum 35% value-added component as subjective. Subjectivity was described as potentially creating "more havoc than good" and as something Nixon could "have a problem with." Rather than personally accusing the committee of choosing to be subjective, Ray presented his comment as "just" an observation. Asking, rather than stating, whether the tool is too subjective distanced the speaker from reproach by suggesting that someone *might* have a problem with it rather than stating that someone does. Asking instead of telling someone a critique is a way of being polite in conversations (Brown & Levinson, 1987; McHoul, 1987). Given the formal setting of this interaction, being polite makes the speaker's comments seem reasonable and appropriate, though they call the wisdom of the state law's (the component percentages are required by law) into question. Interestingly, Ray did not say that subjectivity raises issues of fairness or accuracy. Instead, he described the downside of subjectivity as the possibility of creating "havoc": confusion and disorder. Subjectivity, then, was constructed as destructive and undesirable, while objectivity was positioned as the opposite.

Beyond positioning himself as reasonable and distanced from the critique of 65% of the evaluation, Ray's construction of subjectivity (line 6) positioned him as someone who aims to guard against havoc; one who protects order and clarity. This in turn positioned those who would defend the subjective components of TVAAS as unaware or un-phased by the potential for havoc. Indeed, a fellow board member followed Ray's comments by going "on record" about his "great concern about the small portion of this that is absolutely objective." This extract includes repetition ("I have great concern") and extreme case formulations ("absolutely objective," "true") to emphasize the shortfall of objectivity in the tool and underscore the objectivity of the TVAAS component. Like Nixon, Ray pointed out that the inclusion of TVAAS, rather than pure observation, is "a step in the right direction." Still, observation was constructed as unable to "get us a true measurement of performance" because it was "subjective."

Interestingly, this speaker linked the subjectivity of observation ratings to human nature, not to the observation instruments themselves. Ray noted that observation was compromised because it was done by people who "work together everyday" and "may live together as neighbors." The notion that this compromises their ability to be objective is naturalized by framing it as common practice and common knowledge. For example, the repeated "I know" and "I know how this works" implied that observation always works the same way, and that this is identifiable by outside eyes. It also removed the blame of observer bias from individuals by pointing out that this is just "how it works" when people evaluate those they work and live closely with. This functioned to naturalize and defend the version of evaluations in which observations cannot be anything but subjective; evaluators are always compromised in their ratings, and such subjectivity is described as greatly concerning. Moreover, this extract constructs a version of evaluation in which (1) there is a possibility of an absolute "true" measurement of performance, (2) the performance can be measured by a formula, and (3) the non-TVAAS components compromise an "absolutely objective" rating. According to this construction of teacher evaluation, the TVAAS component now required by law is still too small, but the only redeeming (objective) feature of the overall evaluation. On a similar note,

in an earlier TEAC meeting, State Representative Harry Brooks also described TVAAS as the only defensible component of the overall evaluation. Extract 4 comes from a discussion of the troublesome timeline for generating individual teacher effect scores. Under the system in use, TVAAS scores are calculated over the summer and delivered to the state to be distributed midway through the following year. In order to include teacher effect scores in annual evaluations, as required by Tennessee law, this timeline had to be sped up, but state department representatives reported that efforts to speed the timeline might take several years to implement. Below, Brooks pointed out that the TVAAS component will be important when it comes to legal challenges to teachers' ratings, even in the first years of the policy.

**Extract 4**

1 BROOKS: Because ultimately I I mean if we wind up in court we're going to
2 have to
3 have TVAAS data for teacher reviews.
4 KATIE: Let let's talk about a couple different options that we could look at I
5 mean I think that's the ideal ()
6 BROOKS: () I mean I would
7 KATIE: Yeah () and we'll well walk that-
8 BROOKS: Having been having been in court for civil services for many years () I
9 can hear those lawyers
10 KATIE: Absolutely
11 BROOKS: Saying state statute said xyz
12 KATIE: I understand
13 BROOKS: You make a decision without that information, we'll go to court. Do you
14 know what I'm saying?

In the extract above, Brooks made two important points about the inclusion of TVAAS scores in a teacher's evaluation. First, districts will have it if a teacher sues over their rating. This constructs a version of VAM that stands up in court more effectively than any other component of the evaluation. Second, he asserted that a teacher evaluation rating without a TVAAS component would definitely be contested. He equated making "a decision without that information" with "going to court" (line 13), as if one automatically causes the other. The familiar rhetorical question, "do you know what I'm saying", also worked to naturalize his comments as if he has said something people were expected to know, understand, or agree with. Phrases like this ("you know what I'm saying") set up predictable, preferred responses in conversations (Jefferson, 1974); in this case agreement that one does know what he is saying is the preferred response. If we were to disagree, we would have to acknowledge that we understand before explaining any dissent (e.g. "I know what you're saying, but…"). Katie's short interjections are another example of agreement, in this case strong agreement ("absolutely" rather than "yes" or "mmhmm"), which adds to the weight of Brooks' argument.

Extract 4 displayed strong agreement, making it appear natural for TVAAS to be the strongest evidence of effectiveness, the arbiter of any dispute, and the most essential part of an overall evaluation. Like Extract 1, such claims about the stability and reliability of TVAAS are constructed, supported, and reified by committee members in and through conversations about multiple aspects of the evaluation from labels for different rating levels to legal disputes.

In the section that follows, we present two cases in which the proposed use of TVAAS is called into question. Such moments were rare across the data, almost always limited to visiting presenters, and, as we show below, accompanied by a great deal of conversational difficulty.

**Concerns about VAM are minimized**

As noted in the literature review, there are a number of methodological concerns (e.g., error rates and bias) and logistical challenges (e.g., getting scores returned in ample time) associated with the use of VAM. Thus, we were particularly interested in examining the ways in which the members of the TEAC dealt with the concerns in and through their interactions. Overall, we noted that potential deterrents to the use of value-added data were minimized, if mentioned at all. In other words, anything that could have positioned TVAAS as a less-than-perfect option was rarely made relevant, with many speakers displaying great difficulty questioning the use of TVAAS.

The following extract is one of very few examples from the data in which a speaker openly questioned or expressed concerns about TVAAS. This is an example of variability within the data set. More specifically, it came just after the same discussion about the timeline for releasing TVAAS scores, as was displayed in Extract 4. Across the data, members of the TEAC deemed a summer release of scores "completely unacceptable," in part because a July release was too close to the beginning of Tennessee's school year for an evaluation to trigger promotion or termination. One committee member, Principal Jimmy Bailey, asked if the score reporting timeline would be addressed as early as May 2010; yet, the reality of the TVAAS timeline did not come up for full discussion until four months later, eight meetings into the committee's process. In fact, logistical concerns were only discussed after the SBOE had approved the first reading of the draft policy.

Extract 5 includes Paul Tsangas, head of research for Metro Nashville Public Schools, who was brought in to discuss "data quality issues" that slow down the timeline of generating TVAAS data, and to make suggestions for improving efficiency. At the very end of his six and a half minute opening statement, Tsangas explained:

**Extract 5**

TSANGAS: Uh I will say that with the three year averages uh I understand that the timing is an issue uh I have mixed feelings because we look at trend data on year-by-year basis for any of the school improvement planning that we do. And I would hope that we're making decisions about teachers we're looking at those trends to see that a teacher who may have been struggling a few years ago is moving in the right direction. Maybe the three year average isn't what it should be and I don't know if there's a way to do a weighted analysis in this process to to account for some of that but uh but also know that that there's research that's come out very recently showing that even with three year averages there've been a significant percentage of teachers that are over or underestimated. We've got to have a process in place to to dig into the data to look for those inconsistencies between the different measures and and take those trends into account. Uh that's primarily where we are.

In the above extract, Tsangas does not directly question the validity of TVAAS, but described "mixed feelings" about its proposed use and suggested that a "process" is needed "to look for those inconsistencies between the different measures." Thus VAM, as a tool, was never questioned, but the differences between data points it generates was positioned as something that needed to be monitored. The phrase, "dig into the data", constructed a version of potential inconsistencies that were difficult to see on the surface, but could be managed in a systematic way. This characterization contrasts with the version presented by scholars who have described error in VAM as inherent within particular models (Rothstein, 2009) and as a threat to reliability and stability (Baker et al., 2010; Darling-Hammond et al., 2012).

It is noteworthy that Tsangas left this statement for the end of his lengthy remarks, and framed concerns as personal, rather than methodological. He began with "I will say", rather than "research has shown" or " recent literature suggests." This framed the concern about using a single year's data instead of three-year averages as something he would say, rather than something peer-reviewed research has warned about. Moreover, he described this as his personal feelings ("mixed"), rather than a professional opinion. He positioned himself as having some expertise, noting that he knows about research (which is not described or named). Yet, this research was minimized when he explained that it only came "out very recently", thus allowing those who have supported VAM in the past to save face as the information did not exist before and may yet be discredited. Tsangas also demonstrated a great deal of hedging ("maybe", "I don't know", "but"), which is an example of displaying uncertainty in conversation and functions to soften the critique of the proposed use of VAM.

In terms of methodological concerns, Tsangas made two primary aspects of TVAAS relevant: (1) the departure from the procedure Sanders endorsed of using three-year averages in order to provide an annual score (lines 1, 3, 6), and (2) recent research that questions the reliability of the methodology in general (lines 11-12). Earlier in his remarks, Tsangas attributed issues of reliability to the data itself, not to VAM, pointing out that it is difficult to ensure teachers are linked to their students and that data is not missing, mis-entered, or compromised. In fact his very place on the agenda was described as a discussion of "data quality issues", rather than a discussion about the issues with the use of VAM or the quality of TVAAS scores. Indeed the primacy and accuracy of TVAAS was rarely questioned by the committee members or facilitators. Even after Tsangas's statement above, none of the TEAC members took up his concerns in conversation. His comments simply laid in isolation, without question or response from the committee.

As a result of the committee's extended conversation about timing, Nixon suggested that the committee follow the example of Southwest Airlines and hire an outside expert to consult with them. He explained that, several years ago, Southwest Airlines could not find any consultants within their industry to help them improve their boarding efficiency, so they brought in consultants from NASCAR pit crews to suggest improvements. Nixon explained that Southwest now leads the industry in efficiency. Nixon thus framed the problem of data quality and TVAAS timing as one of simple logistics: moving a stack of tests from point A to point B without losing identifying information or accuracy. This version of reality resisted the alternatives: that TVAAS is too cumbersome; that it requires at least three years of data; that standardized test scores are unreliable; that it is difficult to ensure teachers are only being held accountable for their own students, etc. The concern about three-year averages was ignored and the problem of data quality rested with logistics, not with standardized tests or statistical processes (i.e., TVAAS) that rely on them. As a result of this conversation, the consultant from Education First pledged to call FedEx headquarters herself to ask if they might consult on transporting tests back and forth to be checked and scored. That promise was the first and last mention of hiring other outside consultants in any of the public meetings.

Like Extract 5, Extract 6 is another example of an outside presenter raising a concern about TVAAS. It was rare for concerns or questions about the tool to be raised by committee members. In almost every case, such statements were addressed or minimized within a single conversational turn. Extract 6 was drawn from a long speech made by Barker when answering a question about the number of new tests that would need to be created in order to be able to generate TVAAS for all teachers. At the time, Barker's district was involved in the MET project's attempt to identify measures that aligned well with value-added in order to find the most useful and efficient measures of effective teaching (MET, 2010, 2012).

**Extract 6**

1 BARKER: (I) know what the state law says. Know how we want to use that
2 but conceivably we're going to have enough variables in the model to say that
3 looking at value-add doesn't matter. I know that's kind of heretical to say but
4 if you're able to ask a student in the in the classroom a set of questions, three
5 or four questions that have a really strong correlation with a value-add score,
6 you don't need a value add score. That's a really really interesting statistical
7 kind of leap at this point…Now again that's just something to consider cuz in
8 the laws of statistics if you have something that's a pretty strong correlate
9 with another thing you don't have to know both of those you just need to
10 know one of them. So just to kind of lay it out here this afternoon we are absolutely
11 wanting to use TVAAS data…

Within this extract, Barker suggested that TVAAS could someday be replaced by a close correlate in order to avoid the creation of more than 300 additional standardized tests for every grade (even Kindergarten) and every subject. The suggestion that TVAAS may eventually be unnecessary requires a great deal of conversational repair work (Schegloff, Jefferson & Sacks, 1977). First, Barker admitted that it's "kind of" (a conversational hedge) "heretical" (an extreme case formulation). The lexical choice, "heretical," is most closely associated with religious figures who are assassinated for their beliefs and constitutes quite an extreme example. Barker began by acknowledging that the idea is extreme enough to be punishable by death, and then explained it as belonging to another extreme: a "really really interesting statistical leap." He has thus invoked the language of religion (heretics), statistics (correlates), and the philosophy of science (laws of statistics) to shore up his suggestion as if preempting resistance. In keeping with the idea of VAM as the sentinel of trust, the laws of statistics, from which VAM derives its powers, are positioned as making a way for other measures to have similar power by virtue of their association with VAM.

Barker invoked the "laws of statistics" to assert "you don't have to know both of those you just need to know one of them." This is immediately followed, however, with an assurance that Memphis Public Schools want to use TVAAS data, lest they be accused of trying to dodge its objective glare. Throughout the speech, he reiterated his district's desire to use TVAAS, even though he pointed to the possibility of an alternative. Barker described the idea of replacing TVAAS with a close correlate as "something to consider", a "statistical leap", and the expression of "the laws of statistics." Each represents a different strategy for mitigating the dilemma of seeming to speak out against the use of TVAAS. In the first case, he used a hedge to soften the assertion by constructing it as a suggestion ("something to consider"). In the second, he admitted that it was a leap, but described it as statistical, thus positioning the idea within the objective discipline of statistics and quantitative reasoning. In the third case, he invoked the "laws of statistics" to provide credence to the (heretical) idea that if some measure was found to be correlated with VAM, the measure would suffice on its own.

The collection of defensive rhetoric used in the next extract demonstrates the need for politeness when it comes to TVAAS, particularly in a state where officials have long had faith in a system that they could neither describe nor understand. Extract 7 came from a comment made by then commissioner of education, Timothy Webb. In this extract, he expressed concern that there would ever be a measure "comparable" to a TVAAS score; one that would be of equal weight and authority in the teacher evaluations of untested subjects and grades.

**Extract 7**

1 WEBB: ...you've got a system in place that where all of the data are processed

> and all of the growth calculations are gone through some scientific algorithm that's stood the test of time by some entity somewhere who's said this is the growth of this teacher this is your gr- this is what your growth is now we're going to find another growth calculation that we might do locally. Is that gonna stand that same litmus test of I'm I'm paying him more than I'm paying you when I'm on a different growth measure than than yours?...

In the above extract, Webb questioned whether any other growth calculation would stand the "litmus test" of merit pay decisions. He characterized TVAAS scores as "calculations" that have "gone through some scientific algorithm that's stood the test of time by some entity somewhere." Within this description, he constructed TVAAS scores as: (1) scientific; (2) derived from an algorithm; (3) having withstood the test of time; and (4) confirmed or authorized ("by some entity that exists somewhere"). It is important to note that he said "some entity somewhere", rather than stating something more specific like "Dr. William Sanders" or "SAS." Those who have followed Sanders' work since the 1990s know that he is well known for being vague and elusive when it comes to explaining his calculations and methods. He routinely tells audiences that he could explain TVAAS to them, but they would not understand it. In this case, however, the vague and untenable substance of TVAAS seems to be the very source of its authority. As Channell (1994) has pointed out, vagueness in conversations often functions to cover gaps in knowledge and/or refute potential counter-arguments.

Webb also pointed out that TVAAS it is not only historic, but scientific as well. All this makes it difficult to imagine that any other measure could be its equal. This version of reality, one in which TVAAS cannot be replaced, let alone be questioned, resonates with the idea of a sentinel of trust. It also resonates with the very genre of science fiction in which real technological capabilities are exaggerated beyond their actual capabilities and used to create a different reality. Convinced in part by Webb's argument, committee members chose to create or adopt more tests rather than investigate the possibility of replacing TVAAS with a proxy or some other measure.

## Discussion and Conclusion

In this paper, we have demonstrated how a certain version of VAM was worked up in the conversations of one state's policymakers. We described what aspects of the tool were made relevant, and which versions of the tool (including those that acknowledge methodological concerns and teachers' experiences) were acknowledged and/or resisted. Throughout the conversations of the TEAC, VAM was held up as the beacon of decisive stability within an otherwise subjective evaluation rating. This sentinel of trust version held even though TVAAS data was not available for most teachers, nor available in time for annual evaluations. The sentinel of trust version of VAM is not, however, the only possible version.

Researchers agree that VAM should not be used as a single indicator of effectiveness (e.g. Amrein-Beardsley, 2008; Baker et al., 2011; Braun, Chudowsky, & Koenig, 2010; McCaffery, 2003). As Harris (2009) noted, "many of the key assumptions of value added analysis have been rejected by empirical analysis" (p. 319). Yet, in the conversations of the TEAC, it is frequently positioned as the cornerstone objectivity and the arbiter of accuracy. Rarely, if ever, did Tennessee policymakers question VAM's complex relationship to reliability and validity in practical applications. Instead, committee members reified the version of VAM as a sentinel of trust, positioning it as the only defensible component of the evaluation that granted other components their accuracy. This stands in sharp contrast to the version worked up by researchers, as well as by teachers who work in systems where VAM has been applied.

As Amrein-Beardsley and Collins (2012) have written, teachers working in systems where VAM is used to calculate merit pay view VAM as chaotic, unfair, and unpredictable. Some teachers even compared the chance involved in earning merit pay due to EVAAS scores to winning the lottery. Others were convinced that their student population had more to do with their EVAAS score than their teaching. In their study of teachers who had been terminated because of low value-added scores, Amrein-Beardsley and Collins found that:

> From these teachers' perspectives, it seems that many district administrators are more trusting of SAS® EVAAS® and are skewing [observation] data to match. This makes sense in theory, as the SAS® EVAAS® is the objective system that the district has purchased, and traditional observational scores are increasingly being dismissed as subjective. (p. 19)

This reported pattern among administrators is similar to how VAM was constructed in the conversations of the TEAC in which it was positioned as a tool to balance or validate other evidence. According to the TEAC members, even when VAM is used with other, or "multiple measures" of effectiveness, it directs and supersedes them.

We suggest that the construction of VAM as the sentinel of trust made the work of the TEAC and the SBOE easier, because it was presumed to make up for the weaknesses and biases inherent to other measures. Without a stable and unassailable core, or an explicit definition of effectiveness in teaching, no evaluation tool can hold. We argue that the absence of a single, monolithic explanation of effectiveness in teaching makes a single, accurate, and objective tool for measuring it part of the realm of science fiction, rather than reality. No statistical tool can tell us what makes a good teacher if it us unclear what, beyond standardized test scores, should count as evidence good or bad teaching. We further suggest that this "sentinel of trust" construction, however inaccurate, has led policymakers to largely put the horse before the cart on using VAM for teacher evaluations. Methodological concerns do not support the use of VAM to generate individual teacher effect scores annually, especially when using only one year of data. Testing logistics, in most cases, will not be ideal for several years. And, since VAM is only as good as the data that goes into the model, buying or making hundreds of new tests for untested subjects and grades is likely to do more harm than good to the accuracy of evaluation and the quality of education. Still, the media, the public, and both appointed and elected policymakers in this case (the TEAC and the SBOE) hold fast to the myth of objectivity, and the mystique of a statistical tool for the definitive measurement of teacher effectiveness.

The disconnect between versions of VAM worked up by those who use it (statisticians), experience it (teachers), and legislate it (policymakers) is cause for concern; so too are the patterns of conversational difficulty noted when policymakers discussed questions surrounding VAM's use or authority over other measures. We argue in this case that the cultural power of VAM as constructed by these policymakers has generated relative comfort around troubling and complex decisions required for defining, identifying, and measuring teacher effectiveness in much the same way that the power of the archetypal romance (Frye, 1957) allowed *The LA Times* to promote a certain version of VAM to the public (Gabriel & Lester, in press). Archetypal stories and familiar tropes, like the mythical sentinel and the quest for school reform, lull us into complacent acceptance that VAM is what we wish it to be. Meanwhile, the myth of objectivity obfuscates a hornet's nest of difficulties with a tool that is perhaps best suited for higher-level analyses of groups of teachers, and school-level policy decisions (Harris, 2009). As policymakers engage in conversations about VAM as a tool for teacher evaluation, a heightened level of awareness about the ease with which concerns about VAM can be minimized may encourage more complex and perhaps even more informed conversations about VAM as a policy issue.

Teachers and researchers have both presented compelling counter-stories to the sentinel of trust version of VAM, constructing it as everything from fatally flawed to promising if applied with caution. Yet, these perspectives were not made relevant in the conversations of the TEAC. We suggest that the very rhetorical strategies used to work up a trustworthy version of VAM simultaneously resist other versions, positioning VAM-detractors as unscientific, subjective, and unwilling to "bite the bullet" and "call a spade a spade" when it comes to labeling a teacher "ineffective." Like other discursive constructions, the sentinel of trust version can only be interrupted when such strategies are made explicit through analysis and deconstruction.

## References

Amrein-Beardsley, A. (2012) "Value-added measures in education: The best of the alternatives is simply not good enough. *Teachers College Record*, Date Published: January 12, 2012 http://www.tcrecord.org ID Number: 16648, retrieved: 3/19/2012.

Amrein-Beardsley, A. (2008). Methodological concerns about the Education Value-Added Assessment System (EVAAS). *Educational Researcher, 37*(2), 65-75.

Antaki, C., Billig, M. G., Edwards, D., & Potter, J. A. (2003). Discourse analysis means doing analysis: A critique of six analytic shortcomings. Discourse Analysis Online, 1. Available from: http://extra.shu.ac.uk/daol/articles/open/2002/002/antaki2002002-paper.html

Atkinson, P. (1990). *The ethnographic imagination*: *Textual constructions of reality*. London: Routledge.

Baker, E., Barton, P., Darling-Hammond, L., Haertel, E., Ladd, H., Linn, R., et al. (2010). *Problems with the Use of student test scores to evaluate teachers*. Washington, DC: Economic Policy Institute.

Barnes, B., Palmary, I., & Durrheim, K. (2001). The denial of racism: The role of humor, personal experience, and self-censorship. *Journal of Language and Social Psychology*, *20*(3), 321-338.

Bratton, S., Horn, S., & Wright, S. (1996). *Using and interpreting Tennessee's Value-added Assessement System: a primer for teachers and principals*. Retrieved February 26, 2011, from Shearon For Schools: www.shearonforschools.com/documents/tvaas.html

Braun, H. (2010) *Using student progress to evaluate teachers: A primer on value-added models*. Princeton, NJ: Educational Testing Service Policy Perspectives. Retrieved from: http://www.ets.org/Media/Research/pdf/

Braun, H.; Chudowsky, N. & Koenig, J. (2010). *Getting value out of value-added: Report of a workshop*. Washington, DC: Committee on value-added methodology for instructional improvement, program evaluation, and accountability, National Research Center. Retrieved from: 0. http://www.nap.edu/catalog/12820.html

Briggs, D. & Domingue, B. (2011). *Due Diligence and the evaluation of teachers: A review of the value-added analysis underlying the effectiveness rankings of Los Angeles Univfied School District teachers by the* Los Angeles Times. Boulder: National Education Policy Center.

Brown, P., & Levinson, S. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.

Channell, J. (1994). *Vague language*. Oxford: Oxford University Press.

Chetty, R., Friedman, J.N., & Rockoff, J.E. (2012). *The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood.* Working Paper 17699. Cambridge, MA: National Bureau of Economic Research.

Corcoran, S., Jennings, J., & Beveridge, A. (2011). *Teacher Effectiveness on High- and Low-Stakes Tests.* Paper presented at the annual conference of the Association for Education Finance and Policy in Seattle, WA.

Creswell, J., & Miller, D. (2000). Determining validity in qualitative inquiry. *Theory into Practice*, *39*, 124-130.

Croninger, R., & Valli, L. (2009). "Where is the action?" Challenges to studying the teaching of reading in elementary classrooms. *Educational Researcher, 38*(2), 100-108.

Darling-Hammond, L.; Amrein-Beardsley, A., Haertel, E., Rothstein, J. (2012) "Evaluating teacher evaluation." *Phi Delta Kappan, 93*(6), 8-15.

Edwards, D. (1997). *Discourse and cognition*. London: Sage.

Edwards D (2000) Extreme case formulations: Softeners, investment and doing non-literal. *Research on Language and Social Interaction*, *33*(4), 347–373.

Edwards, D., & Potter, J. (1993). Language and causation: A discursive action model of description and attribution. *Psychological Review*, *100*(1), 23-41.

Ewing, D. (2011). *Leading mathematician debunks 'value-added.'* Washington Post, April 5, 2011.

Frye, N. (1957). *The anatomy of criticism*. Princeton, NJ: Princeton University Press.

Gabriel, R. E. (2011). Tennessee teacher evaluation policies under Race To The Top: A Discursive Investigation. Knoxville, TN: University of Tennessee. http://trace.tennessee.edu/utk_graddiss/971

Gabriel, R., & Lester, J. (2010). Gabriel, R., & Lester, J. (2010, December). Public displays of teacher effectiveness. EducationWeek, 30(15). http://www.edweek.org/ew/articles/2010/12/15/15gabriel.h30.html

Gabriel, R., & Lester, J. N. (in press). The romance quest of education reform: A discourse analysis of *The LA Times*' reports on value-added measurement teacher effectiveness. *Teachers College Record.*

Goldhaber, D., & Hansen, M. (2008). *Is It Just a Bad Class? Assessing the Stability of Measured Teacher Performance.* Center for Reinventing Public Education, working paper # 2008_5. University of Washington.

Harris, D. (2009). Would Accountability Based on Teacher Value Added Be Smart Policy? An Examination of the Statistical Properties and Policy Alternatives. *Education Finance and Policy. 4*(4), 319-350.

Jorgensen, M., & Phillips, L. (2002). *Discourse analysis as theory and method*. London, UK:

Kane, T.J., & Staiger, D.O. (2008). Estimating teacher impacts on student achievement: an experimental evaluation. Working Paper 14607. Cambridge, MA: National Bureau of Economics.

Laclau, E., & Mouffe, C. (1985). *Hegemony and socialist strategy*. London: Verso.

Lockwood, J.R., Louis, T.A., & McCaffrey, D.F. (2002). Uncertainty in Rank Estimation: Implications for Value-Added Modeling Accountability Systems. Journal of educational and behavioral statistics, *27*(3), 255-270.

Lockwood, J.R., McCaffrey, D.F., Hamilton, L.S., Stecher, B.M., Le, V., & Martinez, F. (2007). The Sensitivity of Value-Added Teacher Effect Estimates to Different Mathematics Achievement Measures. *Journal of Educational Measurement, 44*(1), 47-67.

McCaffery, D. (2003) *Evaluating value-added models for teacher accountability.* Santa Monica: Rand publishing.

McCaffery, D., Sass, T., Lockwood, J., & Mihaly, K. (2009). The intertemporal variability of teacher effect estimates. *Education Finance and Policy. 4*(4), 572-606.

McHoul, A.W. (1987). Why there are no guarantees for interrogators. *Journal of Pragmatics*, *11*, 455-471.

Measures of Effective Teaching Project. (2012). *Learning about Teaching Initial Findings from the Measures of Effective Teaching Project.* Seattle: Bill and Melinda Gates Foundation.

Measures of Effective Teaching Project. (2010). *MET Project.* Retrieved February 7, 2011, from http://www.metproject.org/project.

Miller, R. (2012, April). *Discussants comments*. Annual meeting of the American Educational Research Association, Vancouver, BC.

Muhr, T. (2004). User's manual for ATLAS.ti 5.0. Berlin: ATLAS.ti Scientific Software Development GmbH.

Papay, J. (2011). Different Tests, Different Answers: The Stability of Teacher Value-Added Estimates Across Outcome Measures. *American Educational Research Journal, 48*(1), 163-193.

Pomerantz, A. (1986) Extreme case formulations: A way of legitimizing claims. *Human Studies* 9, 219–229.

Potter, J. (1996). *Representing reality: Discourse, rhetoric and social construction.* London: Sage.

Roe, E. (1994). *Narrative policy analysis: Theory and practice*. Durham, NC: Duke University Press.

Rothstein, J. (2009). Student sorting and bias in value-added estimation: selection on observables and unobservables. *Education Finance and Policy, 4*(4), 537-571.

Sacks, H. (1992). *Lectures on Conversation*. Oxford: Blackwell.

Sanders, B., & Rivers, J. (1996). *Cumulative and Residual Effects of Teachers on Future Student Academic Achievement.* Knoxville: University of Tennessee Value Added Research and Assessment Center.

Schegloff, E. A., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, *53*, 361-382.

Schochet, P.Z., & Chiang, H.S. (2010). *Error Rates in Measuring Teacher and School Performance Based on Student Test Score Gains* (NCEE 2010-4004). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Taylor, S. (2001). Locating and conducting discourse analytic research. In M. Wetherell, S. Taylor & S. J. Yates (Eds.), *Discourse as data: A guide for analysis* (pp. 5-48). London: Sage.

Tennessee Department of Education (2010). Meeting Minutes from 5/27. Retrieved from: http://tn.gov/firsttothetop/docs/TEAC_minutes/May%2027%202010.PDFTennessee

Tennessee Department of Education (2011). Educators overview: New Teacher and Principal evaluation. Retrieved from: http://www.tn.gov/firsttothetop/docs/Educators_Overview.pdf

Tracy, K. (1995). Action-implicative discourse analysis. *Journal of Language and Social Psychology*, *14*(1-2), 195-215.

Wodak, R. (1999). Critical discourse analysis at the end of the 20th century. *Research on Language and Social Interaction*, *32*(1&2), 185-193.

Wood, L. A., & Kroger, R. O. (2000). *Doing discourse analysis: Methods for studying action in talk and text*. Thousand Oaks, CA: Sage.

## About the Authors

Rachael Gabriel
University of Connecticut
rachael.gabriel@uconn.edu
Rachael Gabriel is an Assistant Professor of Reading Education at the University of Connecticut. Rachael's career in education began as a middle school literacy teacher in an urban charter school. She has since worked as a literacy specialist, pursued a reading specialist certification and a Ph.D. in Education with a focus on literacy studies. She holds graduate certificates in both quantitative and qualitative research methods in education, and she is a former fellow of the Baker Center for Public Policy. As a researcher, Rachael has focused on teacher preparation, development and evaluation with a specific interest in related policy and a continued interest in literacy instruction, and disability studies.

Jessica Nina Lester
Washington State University
jessica.lester@tricity.wsu.edu
Jessica Lester is an Assistant Professor at Washington State University. She began her career in education as a middle school math and science teacher. She has since worked as a special educator in the US and Colombia, South America. She holds a Ph.D. in Educational Psychology, with a focus on theoretical and methodological foundations of qualitative methodologies and cultural studies in education. Her main research interests lie at the intersection of culture, psychological constructs as produced in and through discourse (e.g., learning, motivation, emotions, etc.), and education, particularly as related to the education of targeted youth.

## About the Guest Editor and Assistant Guest Editors

Guest Editor
Dr. Audrey Amrein-Beardsley
Arizona State University
audrey.beardsley@asu.edu
Dr. Amrein-Beardsley is currently an Associate Professor in the Mary Lou Fulton Teachers College at Arizona State University. Audrey's research interests include educational policy, research methods, and more specifically, high-stakes tests and value-added measurements and systems. In addition, she researches aspects of teacher quality and teacher education. She is also the creator and host of a show titled Inside the Academy during which she interviews some of the top educational researchers in the academy. For more information please see: http://insidetheacademy.asu.edu.

Assistant Guest Editor
Dr. Clarin Collins
Virginia G. Piper Charitable Trust
clarin.collins@asu.edu
Clarin Collins recently completed her Ph.D. in Educational Policy and Evaluation from Arizona State University, with an emphasis in research methods. Via her dissertation, she examined teachers' understandings of and experiences with the SAS Education Value-Added Assessment System (EVAAS) in the Houston Independent School District where it is used to evaluate teacher

effectiveness. Clarin is currently a Research and Evaluation Officer at the Virginia G. Piper Charitable Trust in Phoenix.

Assistant Guest Editor
Dr. Sarah Polasky
Arizona State University
sarah.polasky@asu.edu
Dr. Sarah Polasky is the Value-Added Specialist for the Arizona Ready-for-Rigor Project, a Teacher Incentive Fund Grant, within the Mary Lou Fulton Teachers College. Her current research interests include the development and implementation of value-added measurements and systems using high-stakes test data, assessment in early childhood education, the use of alternative achievement (e.g., district benchmarks, formative assessments) and non-achievement (i.e., developmental) data for value-added analysis, as well as the impact of socioemotional and neurological development of young children on their short- and long-term academic achievement.

Assistant Guest Editor
Edward F. Sloat
Mary Lou Fulton Teachers College, Arizona State University; Dysart Unified School District, Surprise, Arizona
esloat@asu.edu
Mr. Sloat is currently employed as the Director of Research and Accountability at Dysart Unified School District located in Surprise, Arizona and a doctoral student in the Leadership and Innovation Program within the Mary Lou Fulton Teachers College, Arizona State University. Mr. Sloat has served as Deputy Associate Superintendent for Research and Evaluation within the Arizona Department of Education, the Director of Research, Planning, and Assessment for the Peoria (Arizona) Unified School District, and as Director of Research and Assessment at the Glendale (Arizona) Elementary School District. He regularly contributes to state technical and policy working/advisory groups concerning assessment design and accountability systems and is past President of the Arizona Education Research Organization. Mr. Sloat holds a Master's Degree in Applied Economics from the University of Arizona, concentrating in econometric methods and management information systems. His academic interests focus on value-added modeling, education accountability and evaluation systems, data-driven instructional planning, applications of measurement theory, and research methods.

**SPECIAL ISSUE**
**Value-Added: What America's Policymakers Need to Know and Understand**

education policy analysis archives

Volume 21 Number 9 January 31st 2013 ISSN 1068-2341

Please contribute commentaries at http://epaa.info/wordpress/ and send errata notes to Gustavo E. Fischman fischman@asu.edu

**Join EPAA's Facebook community** at https://www.facebook.com/EPAAAAPE and **Twitter feed** @epaa_aape.

# education policy analysis archives
# editorial board

*Education Policy Analysis Archives Vol. 21 No. 9* **SPECIAL ISSUE**

# arquivos analíticos de políticas educativas
## conselho editorial