
education policy analysis archives

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



Arizona State University

Volume 22 Number 98 October 27th, 2014

ISSN 1068-2341

Houston, We Have a Problem: Teachers Find No Value in the SAS Education Value-Added Assessment System (EVAAS®)

Clarín Collins
Private Foundation
Phoenix, Arizona
United States

Citation: Collins, C. (2014). Houston, we have a problem: Teachers find no value in the SAS education value-added assessment system (EVAAS®). *Education Policy Analysis Archives*, 22(98). <http://dx.doi.org/10.14507/epaa.v22.1594>

Abstract: This study examined the SAS Education Value-Added Assessment System (EVAAS®) in practice, as perceived and experienced by teachers in the Southwest School District (SSD). To evaluate teacher effectiveness, SSD is using SAS EVAAS® for high-stakes consequences more than any other district or state in the country. A mixed-method design including a large-scale electronic survey was used to investigate the model's reliability and validity; to determine whether teachers used the SAS EVAAS® data in formative ways as intended; to gather teachers' opinions on SAS EVAAS®'s claimed benefits and statements; and to understand the unintended consequences that occurred as a result of SAS EVAAS® use in SSD. Results revealed that the reliability of the SAS EVAAS® model produced split and inconsistent results among teacher participants, and teachers indicated that students biased the SAS EVAAS® results. The majority of teachers disagreed with the company's marketing claims and did not report similar SAS EVAAS® and principal observation scores, reducing the criterion-related validity of both measures of teacher quality. Many unintended consequences associated with the high-stakes use of SAS EVAAS® emerged through teachers' responses, which revealed among others that teachers felt heightened pressure and competition, which reduced morale and collaboration, and encouraged cheating or teaching to the test in attempt

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

Manuscript received: 1/8/2014
Revisions received: 6/17/2014
Accepted: 7/30/2014

to raise SAS EVAAS® scores. The results of this study, one of the first to investigate how the SAS EVAAS® model works in practice, should be considered by policymakers, researchers, and districts when considering implementing the SAS EVAAS®, or any value-added model for teacher evaluation.

Keywords: Value-added models (VAMs); teacher effectiveness; teacher quality; teacher evaluation; accountability; education policy.

Houston, tenemos un problema: Los docentes no le encuentran valor al sistema SAS de evaluación de valor agregado de la educación.

Resumen: Este estudio examinó la implementación del sistema SAS de evaluación de valor agregado de la educación (EVAAS®), tal como fue percibido y experimentado por docentes en el Distrito Escolar del Suroeste (SSD). El SSD, está utilizando el SAS EVAAS® con consecuencias severas. más que cualquier otro distrito o estado en el país. Un diseño de métodos mixtos que incluye una encuesta electrónica a gran escala, se utilizó para investigar: la fiabilidad y la validez del modelo; para determinar si los profesores utilizan los datos SAS EVAAS® en maneras formativas según lo previsto; para reunir opiniones de los docentes sobre los beneficios y de acuerdo a lo que promociona la empresa que produce SAS EVAAS®; y para comprender las consecuencias no deseadas que se produjeron como consecuencia de la utilización SAS EVAAS® en SSD. Los resultados revelaron que la fiabilidad del modelo SAS EVAAS® produce resultados inconsistentes entre los participantes, y los profesores indicaron que los estudiantes sesgaban los resultados del SAS EVAAS®. La mayoría de los profesores no estuvo de acuerdo con los beneficios que el marketing de la empresa promocionaba y los puntajes de SAS EVAAS® y de observación no eran similares a los que los directores de escuelas recibían, reduciendo la validez de criterio de ambas medidas de calidad. Muchas consecuencias no deseadas asociadas con las consecuencias severas de SAS EVAAS® surgieron a través de las respuestas de los docentes, que revelaron que la presión y la competencia, redujo la moral y la colaboración, y estimulo a hacer trampa o enseñar para responder exámenes en un intento de aumentar puntuaciones de SAS EVAAS®. Los resultados de este estudio, uno de los primeros en investigar cómo funciona el modelo SAS EVAAS® en la práctica, deben ser considerados por los responsables políticos, los investigadores y los distritos de la hora de considerar la aplicación de la SAS EVAAS®, o cualquier modelo de valor agregado para la evaluación de los maestros.

Palabras clave: modelos de valor agregado (VAM); efectividad de los maestros; calidad de los maestros; evaluación de los docentes; rendición de cuentas; la política de educación.

Houston, temos um problema: Os professores não acham valor na avaliação do sistema SAS de valor agregado da educação.

Resumo: O presente estudo analisou a implementação do sistema SAS para a avaliação do sistema SAS de valor agregado da educação (EVAAS®), como foi percebido e experimentado pelos professores do Distrito Escolar do Sudoeste (SSD). O SSD está usando SAS EVAAS® com consequências graves. mais do que qualquer outro distrito ou estado do país. Um projeto de métodos mistos, que inclui uma pesquisa on-line de grande escala, foi utilizado para investigar: a confiabilidade e a validade do modelo; para determinar se os professores utilizam SAS maneiras EVAAS® dados em formação como o planejado; para recolher feedback dos professores sobre os benefícios e de acordo com o que promove a empresa produtora EVAAS® SAS; e compreender as consequências não intencionais que ocorreram como resultado do uso de SAS SSD EVAAS®. Os resultados revelaram que a confiabilidade do modelo SAS produz resultados inconsistentes EVAAS® entre os participantes e os professores indicaram que os estudantes sesgaban resultados

EVAAS® SAS. A maioria dos professores não concordavam com os benefícios que a empresa de marketing promovidas e SAS pontuação EVAAS® e observação não foram semelhantes aos diretores recebido, reduzindo a validade do critério de ambas as medidas qualidade. Muitas consequências indesejadas associadas com as graves consequências da SAS EVAAS® surgiu através de respostas dos professores, que revelou que a pressão e a competição, reduziu a moral e colaboração, e incentivo para trapacear ou exames de ensino para responder uma tentativa de aumentar a pontuação EVAAS® SAS. Os resultados deste estudo, um dos primeiros a investigar a forma como o modelo SAS EVAAS® funciona na prática, deve ser considerado pelos decisores políticos, investigadores e distritos quando se considera a aplicação da SAS EVAAS® ou qualquer modelo de valor agregado para a avaliação de professores.

Palavras-chave: modelos de valor agregado (VAM); eficácia do professor; qualidade dos professores; avaliação dos professores; prestação de contas; política de educação.

Introduction

Stemming from *A Nation at Risk* (1983), our nation has been focused on accountability and academic reform for the past four decades—holding students, teachers, principals and schools responsible for measureable, quantitative evidence of student achievement (Holloway-Libell & Collins, 2014). As one reform effort is replaced by another in the constant quest for academic quality, the most recent movement is focused on teacher effectiveness. Currently, our nation appears to be speeding around the teacher accountability racetrack—racing to find teacher accountability systems that states and districts can use, like a one-size-fits-all model that can easily determine teacher quality and effectiveness. It is hoped that such models will identify and remove ineffective teachers, leave in place effective, high quality teachers and as a result, student learning will improve and student achievement will increase (see Holloway-Libell & Collins, 2014).

The federal government, in particular under the leadership of U.S. Secretary of Education, Arne Duncan, has been the leading force of this race with the Race to the Top (RttT) challenge. To date, 18 states, the District of Colombia (DC), and 21 school districts have received RttT awards which require that teacher effectiveness be measured using student performance data as a “significant factor” (USDOE, 2009, 2012a, 2012b, 2013). In addition, 44 states and DC have applied for waivers put into place by the U.S. Department of Education to excuse states from meeting No Child Left Behind’s prior goal that 100% of the students in their schools be academically proficient by the year 2014 (Philips, 2012). In exchange for waivers, states must adopt even stronger teacher accountability mechanisms, using student growth scores as an integral component of their teacher evaluation systems.

Currently, 40 states and DC (80%) are using, piloting, or developing statewide growth or value-added models (Collins & Amrein-Beardsley, 2014). Individual states are developing state legislation and policies in addition to the federal requirements for teacher accountability (i.e., NCLB waivers and RttT grant requirements). A recent study indicated that 30 states and DC (61%) now have legislation or policies that require student achievement data be used to “significantly” inform the criteria for the evaluation of teacher effectiveness and subsequent decision-making efforts (Collins & Amrein-Beardsley, 2014).

Purpose of Study

This study expands upon research previously published in EPAA (Amrein-Beardsley & Collins, 2012), which included a case study of four teachers who were terminated from a school

district, at least in part, due to low SAS Education Value-Added Assessment System (EVAAS[®]) scores. The research addressed the teachers' evaluations and identified issues with the model's reliability, validity, and fairness. Such findings warranted further investigation into teachers' perspectives of and experiences with SAS EVAAS[®] in the same school district, specifically those directly evaluated by SAS EVAAS[®] (EVAAS[®]-eligible teachers). As such, the purpose of this study was to further examine the reliability, validity and formative use of SAS EVAAS[®] data as experienced by EVAAS[®]-eligible teachers in the K12 district using SAS EVAAS[®] for more high-stakes consequential decision making than any other district in the country (Harris, 2011; Lowrey, 2012; Sparks, 2011). Additionally, this study investigated the consequential validity associated with SAS EVAAS[®]: the intended consequences—the benefits and outcomes as promoted by SAS, and the unintended consequences that occurred as a result of SAS EVAAS[®] use and implementation (Messick, 1989).

SAS EVAAS[®]

The SAS EVAAS[®] was one of the first value-added models (VAMs) and is self-proclaimed to be “the most robust and reliable” system available, providing “precise, reliable and unbiased results that other simplistic models found in the market today cannot provide” (SAS, 2012a). As advertised, the system “provides valuable diagnostic information about [instructional] practices,” helps educators to make more “sound instructional choices,” and use “resources more strategically to ensure that every student has the chance to succeed” (SAS, 2012a). The SAS EVAAS[®], like most other VAMs, has shown to be more accurate at analyzing student academic progress than traditional end-of-year “snapshot” or Adequate Yearly Progress (AYP) reports, and the SAS EVAAS[®] is probably the best or “least bad” VAM in existence (Amrein-Beardsley, 2008; Economic Policy Institute [EPI], 2010) given its statistical sophistication.

Today, SAS EVAAS[®] is the most widely used VAM in the country, and North Carolina, Ohio, Pennsylvania and Tennessee use the model state-wide (Collins & Amrein-Beardsley, 2014). Despite widespread popularity of the SAS EVAAS[®], however, no research has been done from the perspective of teachers to examine how their practices are impacted by this methodology that professedly identifies effective and ineffective teachers. Even more disconcerting is that districts and states are tying consequences to the data generated from the SAS EVAAS[®], entrusting the sophisticated methodologies to produce accurate, consistent, and reliable data, when it remains unknown how the model actually works in practice.

Existing research on value-added tends to be largely quantitative, focusing only on the data generated from the models. Lacking from the research base are (more descriptive, qualitative) studies about the relationships between VAM scores and the teaching qualities they are assumed to measure (Hill, Kapitula, & Umland, 2011), and analyses of how these models actually impact teacher practices. Whether teachers use VAM-based data to reflect on and improve their instruction is also unknown, though widely assumed to be the case. To determine if the SAS EVAAS[®] reliably and validly identifies teacher qualities and whether the model produces usable data for teachers to inform their practices, it is necessary to bring the invaluable perspectives and experiences of the teachers evaluated by the model into the national conversation. Failing to do so essentially means we are taking the “product” at face value without really understanding if it works as expected and promoted. As such, this is one of the first studies to investigate how the SAS EVAAS[®] works in practice, from the perspectives of EVAAS[®]-eligible teachers in one of the country's largest school districts.

Summary of District

The district in this study is one of the largest school districts in the country, located in a city in the Southwest United States. For the purposes of this article, the district will be referred to as Southwest School District (SSD). SSD consists of approximately 300 schools, 204,000 students and 11,000 teachers. The district primarily serves high-needs students, with 62% of students labeled at-risk, 92% from racial minority backgrounds, 80% on the federal free-and-reduced lunch program, and 57% classified as English Language Learners, Limited English Proficiency, or bilingual.

In 2007, SSD administrators contracted with the SAS software company to use their SAS EVAAS[®] system to measure student progress as a component in their newly created program to recognize and celebrate great teaching. The district pays approximately \$500,000 per year for the contract with the SAS software company.

At the time of this study (2012), SSD had two main teacher evaluation and accountability systems: 1) a program in which the district used one year of SAS EVAAS[®] scores to rank order teachers throughout the district as criteria for a merit pay program, and 2) a professional development and evaluation system where classroom observations were used to evaluate teachers on eight different domains of teacher performance. EVAAS[®]-eligible teachers were evaluated by both systems each year.¹

Methods

Given the large volume of K-12 teachers ($n \approx 11,000$) in SSD, of which approximately 30% or 3,300 were EVAAS[®]-eligible (core subject teachers in grades 3-8; see also Harris, 2011), the researcher chose a survey research method which allowed for the simultaneous “examination of hundreds or even thousands of survey respondents” (Babbie, 1990, p. 41). The survey method allowed interplay between qualitative and quantitative measures, which helped to portray the stories and lived experiences of the participants through findings and results (Greene, 2008).

The overarching research question was as follows: How does the SAS EVAAS[®] work in practice, as experienced by SSD EVAAS[®]-eligible teachers? The researcher designed an electronic survey questionnaire containing four different constructs with sub-questions regarding: (a) Reliability—Are EVAAS[®] scores consistent over time? (b) Validity—Do EVAAS[®] scores match other indicators of teacher quality? (c) Formative uses and consequences—Do teachers use EVAAS[®] data to inform their instruction? (d) Intended consequences and claimed benefits of EVAAS[®]—Do teachers agree with EVAAS[®] marketing claims and statements, and, from their perspectives, are these claims and statements realized in practice?

The researcher derived these questions using an analytical framework aligned with the “Standards for Educational and Psychological Testing” designated by the leading associations in educational measurement and testing (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999). The Standards were designed for use by professional test developers and consumers to evaluate tests and the use of tests (AERA, APA, & NCME, 1999, p. 3). More specifically, the researcher constructed a list of questions aligned with the aforementioned

¹ Given that one teacher evaluation format focused on student achievement data and the other on observed teacher performance in the classroom, it was possible for SSD teachers to receive contradictory feedback via the two systems; for example, a teacher may have received a bonus through high SAS EVAAS[®] scores while they simultaneously received low remarks on the classroom observation evaluation system and vice versa.

measurement constructs, compiled a list of SAS EVAAS[®] statements from the SAS website and literature, and added statements about the school environment/culture as influenced by SAS EVAAS[®], to generate a greater series of Likert-type items. The researcher also included a series of open-ended questions to better capture teachers' individual experiences, beliefs, and opinions about the SAS EVAAS[®] and its use in SSD.

Response Rate

Over 1,300 SSD teachers participated in this study. However, the intent of the researcher was to include only those teachers who were EVAAS[®]-eligible. After removing all non-EVAAS[®] eligible teachers, 882 complete responses remained, with a response rate within the range of 14% to 27%².

Generalizability

The low response-rate reduces the ability to generalize the findings of this study beyond SSD and this sample of teachers. Although the researcher determined that the sample size needed to support generalization was achieved (Creative Research Systems, n.d.), it could still logically be argued that only the most vocal or opinionated teachers participated in this study, which makes it unclear whether the results of the study are generalizable, especially considering the high-stakes consequences that SSD attaches to the SAS EVAAS[®] output (termination and merit-pay). However, situational generalizability is more the intent with this study rather than statistical demographic representativeness of all SSD teachers and teachers in general (Horsburgh, 2003; Popay, Rogers, & Williams, 1998). Readers might make naturalistic generalizations from the findings within their own contexts and experiences with VAMs (Stake & Trumbull, 1982), and transferability of this research can be achieved by the readers (Lincoln & Guba, 1985) as they internalize trends and rich qualitative description from within this one particular setting and group studied (Maxwell, 1941).

Data Analyses

Approximately two-thirds (69.4%) of the respondents were members of the teachers' union. There is a strong union presence in SSD, whereas approximately half of the SSD teachers are members of the local federation of teachers (Z. Capo, personal communication, August 13, 2012). Given the media attention on the opposition of teachers' unions to student performance based teacher evaluations and merit pay tied to such evaluations (e.g., Florida, New York City, Boston, Chicago to name a few), the researcher calculated chi-square analyses for each of the Likert-type items which examined whether the perceptions of SAS EVAAS[®] among the two mutually exclusive groups, union and non-union teachers were significantly different (Gay, 1996). All analyses illustrated that there was no statistically significant difference between the responses of union members and non-union members. In other words, non-union and union members had essentially the same thoughts and beliefs about the SAS EVAAS[®] system and its use within the district. Being a member of the union did not significantly bias respondents' one way or another per issue of interest.

² EVAAS[®]-eligible teachers were not identified by the district; therefore, the researcher emailed all K-8 teachers (N=6,292), and reviewed each completed survey to remove all teachers who were not grade 3-8 math, English/language arts, science or social studies teachers. The response rate range was determined using two denominators: N=6,292 as the denominator representing all K-8 teachers who received the survey, and N=3,300 as the denominator representing the approximate 30% of all 11,000 K-12 teachers in the district who would be EVAAS[®] eligible based on research estimates (Harris, 2011; Nelson, 2011).

The researcher analyzed all quantitative data using IBM SPSS to calculate descriptive statistics for each question (Gay, 1996). The researcher manually coded the qualitative data—4,594 unique responses to the open-ended questions—analyzing responses line-by-line (Strauss & Corbin, 1998), using open codes and keeping track of the number of respondents for each code. The researcher let the qualitative data speak for themselves (Lincoln & Guba, 1985; Glaser & Strauss, 1967) while looking for similarities and differences across the data and codes, as well as “cutting and sorting” to collapse codes from each individual question into larger subgroups of codes and subsequent findings.

The researcher gathered documents related to SAS EVAAS[®] use and teacher evaluations in SSD, and had one 45-minute conversation with SSD’s assistant superintendent of research and accountability to clarify and verify collected information. The researcher exchanged approximately 10 emails with SSD employees in the department of research and accountability, and approximately five emails with the director of the local teachers’ union to authenticate information. The researcher used teacher quotations as often as possible in presenting the findings to demonstrate participants’ experiences and to ensure their voices can be heard throughout the results (Creswell, 2003; Strauss & Corbin, 1998). Additionally, the researcher completed member checks (Lincoln & Guba, 1985), which allowed a group of five, self-selected teacher participants to read and respond to the overall findings of the study.

Results

Demographics and Description of Sample

The majority of the teacher participants was female ($n = 648/871$; 74.4%) and identified as Caucasian/White ($n = 306/868$; 35.3%), African American/Black ($n = 237/868$; 27.3%), or Hispanic/Latino ($n = 231/868$; 26.6%). The average respondent was 37 years old, with the oldest 78 and the youngest 24.

The plurality of the 882 teachers who responded to the survey had taught in SSD for 6-10 years ($n = 226/878$; 25.7%) and had taught in total for 21+ years ($n = 171/879$; 27.3%). Most of the teachers ($n = 312/882$; 35.4%) had received five years of individual SAS EVAAS[®] scores, as depicted in Figure 1.

Most teacher respondents earned their teaching certificates from public universities in-state ($n = 302/881$; 35.0%) or through alternative certification programs ($n = 266/881$; 30.2%). Almost half of the teachers had taught third ($n = 412/882$; 46.7%) or fourth grade ($n = 417/882$; 47.3%), with the average teacher having taught 3.15 different grade levels (see Figure 2).

The majority of teacher respondents taught in the core areas of reading/English language arts, mathematics, science, and social studies, in that order, with the average teacher having taught 3.56 different subject areas in SSD (see Figure 3).

More than three quarters of the teachers indicated that the students they taught were of high socioeconomic needs ($n = 692/874$; 79.2%), which fits with district demographics (i.e., 63% of students labeled at risk, 92% from racial minority backgrounds, 80% on the federal free-or-reduced lunch program, and 58% classified as English Language Learners (ELL), Limited English Proficiency (LEP), or bilingual).

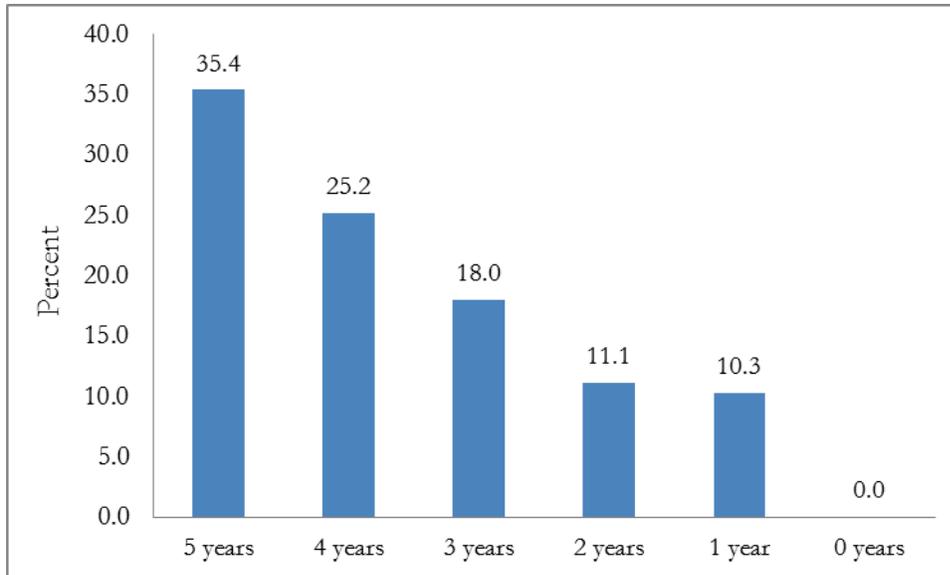


Figure 1. Number of years for which individual SAS EVAAS® scores were received.

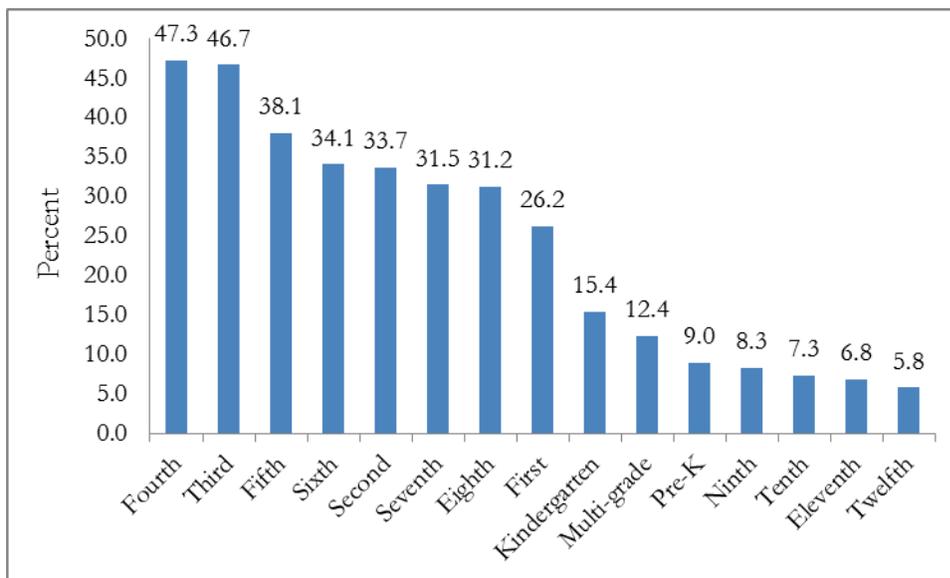


Figure 2. Proportion of grade levels ever taught in the Southwest School District.

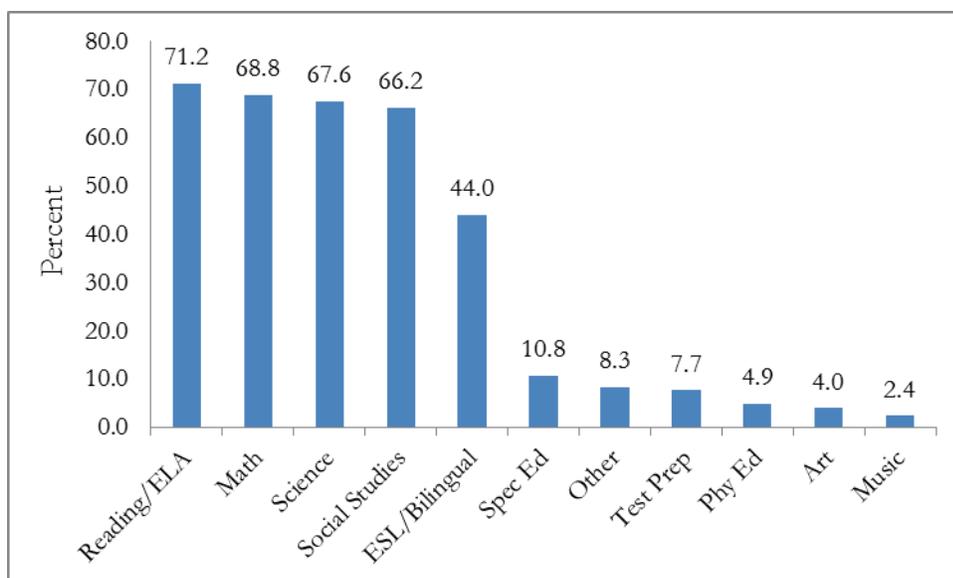


Figure 3. Proportion of subject areas ever taught in the Southwest School District.

Reliability

Among participants in this study, more teachers indicated that their SAS EVAAS[®] scores were inconsistent ($n = 404/874$; 46.2%) year-to-year than those who reported consistent scores ($n = 371/874$; 42.4%). To investigate further, the teachers who reported inconsistent scores ($n = 404/874$; 46.2%) were asked to provide explanations, and 348 teachers did so generating 381 substantive responses³ as to why their SAS EVAAS[®] scores were inconsistent year-to-year. Over one third of these responses ($n = 150/381$; 39.4%) simply indicated that scores varied, but further explanations for the variation were not provided. For example, one teacher replied, “In three years, I was above average, below average and average.” Another teacher responded, “I have taught 4th grade for the last 8 years. My scores have been ‘green’ some years and ‘red’ other years.”

Among the teachers who did provide an explanation for the fluctuation of their SAS EVAAS[®] scores, 24.4% ($n = 93/381$) reported the inconsistencies were caused by the different types of students they taught, and specifically referenced ELL and transition⁴ students as well as high achieving and gifted students as those responsible for score inconsistencies. As one ELL teacher put it, “Since I am teaching 5th grade ELL, I have been categorize[d] as ineffective because my students don't grow when coming from 4th grade all Spanish to 5th grade all English.” A teacher of gifted students explained:

The first year, they were ok. Then as I began to teach the gifted students, the scores continued to show negative growth. For the 2010-2011 school year, the Principal even told me that my scores revealed that I was one of the worst teachers in the school. The School Improvement Officer observed my teaching and reported that my teaching did not reflect the downward spiral in the scores.

³ Some teachers provided more than one substantive response for each open-ended question, here and throughout the rest of the study. Therefore, the number of substantive responses differs from the number of respondents per question. Superfluous answers (e.g., “No opinion,” “Already answered it in previous question,” etc.) were discarded and not counted as responses or used in qualitative coding.

⁴ In Southwest School District, students are transitioned from bilingual classrooms to English-only classrooms typically in the fourth grade. Teachers referred to these students as “transition” students.

Other teachers ($n = 48/381$; 12.6%) described scenarios of switching grade levels or content areas, which impacted their EVAAS[®] scores as they adjusted to new situations. One new teacher attributed the change in scores to her own growth as a teacher, “My second year's score was higher than my first year's score. I attribute this to professional growth and experience.” A fourth grade teacher reflected back on her past four years of SAS EVAAS[®] scores, revealing what she learned along the way:

My first years of teaching I was still learning the ropes. Therefore, those scores were lower; however, over the years I understand that you must teach to the test to get the scores you want. To do well, the students must not only be intimate with the objectives, but also the lay-out and the verbiage on the test. Especially the ELL students. They need to know the wording of the questions beforehand so that they can be sure that they grasp what the question asks.

Reliability Across Grade Levels and Subject Areas

The researcher also asked the teachers whether their scores were consistent if they had taught more than one grade level and/or more than one subject area. Out of the teachers who reported having taught more than one grade level ($n = 559/873$; 64.0%), 51.3% ($n = 287/559$) reported having inconsistent scores across grade levels. Out of the 577 teachers who had taught more than one subject area, 49.6% ($n = 272/577$) reported that their scores were not consistent across subject areas either. Once again, the reliability of scores was pretty equally split; about half of the teachers reported consistent scores, while the other half did not.

While some teachers named specific grade levels ($n = 196$) or subject areas ($n = 209$) that they believed were responsible for inconsistent SAS EVAAS[®] scores, there was no consensus on a specific grade or subject area. One multi-grade teacher explained:

I did a 7th and 8th grade split one year. 7th grade didn't grow, and [the students] were shown to regress a little bit. 8th grade grew. Was it me? Was it them? Was it both? I tend to think it was them. Chemistry in the classroom can affect the growth, too. And I don't know how that would be measurable by any instrument.

Given the emphasis that standardized testing places on subjects such as math and English language arts/reading, several teachers mentioned these subject areas received more resources. A middle school teacher claimed, “Certain subject areas such as reading and math are given the priority in resources,” and another explained, “My scores tend to be high in math, reading, & writing; but low in science & social studies because we have no or limited materials for those subjects.”

Other teachers described that some subject areas had curricula that were less aligned with the tests than others, specifically those subjects (e.g., history, social studies, and science) that relied on the nationally norm-referenced Stanford test. One teacher who could not pinpoint the reason for her score fluctuations recalled, “I receive higher scores in some subjects than in others. Sometimes the most is in my certified field (math) and other times I receive nothing for math but receive bonuses in other subjects.” Another teacher found better success with her SAS EVAAS[®] scores when she taught history:

When I taught 8th grade history the scores rose significantly one year and stayed consistent for two years. I did nothing different in my approach to teaching. This last year I moved to sixth grade math and the scores took a dip.

Reliability Across Student Characteristics

The final reliability question included in this section of the survey instrument asked teachers if they received consistent SAS EVAAS[®] scores despite the varied proportions of different types of students (i.e., ELL, gifted, special education, low/high income) they taught. Among the teachers

who indicated they did teach different types of students year-to-year ($n = 710/877$; 81.0%), 52.5% ($n = 373/710$) responded that their SAS EVAAS[®] scores were inconsistent, yet again. This group of teachers generated 282 substantive responses in explanation.

The plurality of these teachers ($n = 106/282$; 37.6%) responded that all students are different, and that issues such as motivation, prior academic preparation, behavior and external factors such as home life and family situations greatly influenced student performance and inherently teacher SAS EVAAS[®] scores. A reading teacher replied, “[SAS EVAAS] depends a lot on home support, background knowledge, current family situation, lack of sleep, whether parents are at home, in jail, etc. [There are t]oo many outside factors – behavior issues, etc.”

Other teachers specifically referenced certain student groups whom they believed were responsible for impacting their SAS EVAAS[®] scores. Gifted and advanced students were seen by teachers ($n = 49/282$; 17.4%) as high scorers on tests that left little to no room for growth that could be measured by SAS EVAAS[®].

Validity

The survey contained several questions to investigate the validity of the SAS EVAAS[®] model and scores as well. To examine content-related evidence of validity, the researcher included questions to investigate if the student data used to calculate individual teacher SAS EVAAS[®] scores were appropriate. The researcher also included questions for teachers to compare their SAS EVAAS[®] scores to other indicators of teacher quality to examine criterion-related evidence of validity. The responses generated from the validity questions indicated evidence of an overall issue with construct-related evidence of validity as well.

Content Validity

First, teachers were asked if they had ever been evaluated by SAS EVAAS[®] for a grade level for which they were not the teacher of record. Only 9.1% ($n = 80/875$) of teachers indicated this had happened to them, and these teachers reported discrepancies with how their teaching responsibilities during student instruction time were allocated to them as part of the data linkage process (i.e., teaching more than one grade level in a given year but only receiving SAS EVAAS[®] scores for students in one certain grade level).

A similar minority of teachers ($n = 84/874$; 9.6%) indicated that they had been evaluated with SAS EVAAS[®] scores for a subject for which they were not teacher of record, with the majority of these teachers indicating they taught in a departmentalized or team-teaching situation or they were a lab teacher, which resulted in inaccurate allocations of student instruction time included in their SAS EVAAS[®] data.

A slight increase was noted when teachers ($n = 152/871$; 17.5%) were asked if they had ever been evaluated with SAS EVAAS[®] scores for students for whom they were not the teacher of record. This group described situations where students were placed in their classrooms only within weeks of the standardized test used to determine SAS EVAAS[®] scores, or scenarios where teachers had students removed from their classroom early in the year for disciplinary reasons to attend alternative schools but still had those students' scores show up on their SAS EVAAS[®] reports. A middle school English teacher described such a situation:

I'm not sure how I get evaluated for a student who is only in my class for one month and then goes into CEP [community education partners for disciplinary alternative education]. I'm still considered the teacher of record even though he spent 5-6 months out of my classroom.

Criterion-related Evidence of Validity

Teachers were asked if their SAS EVAAS[®] scores typically produced similar findings to their principal or supervisor observation scores, following the assumption in SSD that both represent accurate measures of teacher quality. More than half ($n = 497/863$; 57.6%) of the sample indicated their SAS EVAAS[®] scores do not typically match their observation scores. Out of this group, 367 teachers generated 340 substantive responses explaining these issues further.

The plurality of teachers ($n = 159/340$; 46.8%) replied that their observation scores were always higher than their SAS EVAAS[®] scores, whereas conversely 9.1% ($n = 31/340$) of teachers indicated that their SAS EVAAS[®] scores were always higher than their observation scores. Regardless of which score was higher, the frequently conflicting SAS EVAAS[®] and observation scores seemed to send teachers mixed messages. A fifth grade teacher explained:

Based on the EVAAS system, I am considered below the standards, but based on my principal's observation and state test scores, I am a great teacher...Because on one hand you're meeting the State's testing requirements, but if you're [not] doing well according to EVAAS, then you have two contradicting sets of evaluations.

A seventh grade math teacher responded:

I have always received positive - even glowing - observation and evaluation scores from my principal and evaluator. I have been asked to serve as a lead teacher on campus and I have mentored others - but my negative [EVAAS] growth score does not reflect that.

Others ($n = 43/340$; 12.6%) responded that their observation scores were consistent year-to-year while their SAS EVAAS[®] scores fluctuated. In fact, a fair amount of teachers ($n = 41/340$; 12.1%) indicated that the principal evaluation portion of the observation was very subjective, and that principals based their evaluations on their relationships with teachers. An English teacher explained:

If you're 'in' you'll be rated well. If you're not, you won't. The EVAAS scores are nice in that they are purely data driven, and sometimes (if a teacher is [in] a bad way with the principal) they can be a relief.

Some of the same teachers ($n = 41/340$; 12.1%) described how principals would switch their observation scores if dissimilar to reflect their SAS EVAAS[®] scores. An elementary teacher said, "Evaluation scores are subjective. One principal told me one year that even though I had high SAKS⁵ scores and high Stanford scores, the fact that my EVAAS scores showed no growth, it would look bad to the superintendent."

Another teacher reflected on when her observation scores were changed to match the SAS EVAAS[®], "I had high appraisals but low EVAAS, so they had to change appraisals to match lower EVAAS scores. I was actually put on a growth plan, but met all the requirements and was taken off." A veteran teacher explained her changed scores:

One year I received low performing [scores] on my evaluation...I knew the rating was due to her dislike for me. Upon the arrival of the [SAS EVAAS] scores my students did exemplary...The [assistant] principal changed the [evaluation] rating before I met with her to 'exceeds expectations.'

To look further at criterion-related evidence of validity, teachers were asked if they had received any awards, recommendations, student or parent feedback, or peer evaluations (again,

⁵ State Assessment of Knowledge and Skills (name changed)

assuming such indicators also describe teacher quality) which supported or contradicted their SAS EVAAS[®] scores. Out of the teachers ($n = 367/843$; 43.5%) who indicated they had received contradicting feedback, 286 teachers generated 263 substantive explanations. More than a third of these teachers ($n = 95/263$; 36.1%) reported that they had received or were nominated for awards by their colleagues and mentor teachers at the same time they had received low SAS EVAAS[®] scores. Several of these same teachers ($n = 24/263$; 9.1%) pointed out that they were master or lead teachers, department chairs, or development or academic coaches, having been appointed by peers or principals based on their expertise and skill in certain areas, yet they simultaneously demonstrated the “least growth” or had the “weakest” SAS EVAAS[®] scores in the same subject matter.

Other teachers ($n = 81/263$; 30.8%) described the positive feedback they received from parents and students, through letters, personal communication, and continued communication years after students had left their classrooms. Although some could argue these actions are the most subjective of all, for many of the teachers, this feedback served as a more solid indicator of their own effectiveness. As a third grade teacher simply put it, “Academic testing does not tell the whole story.” Another teacher shared:

Each year regardless of my EVAAS results, parent[s] request for their children to be in my class. I feel this is because they know I care about their children and that I am giving them my best each day. Each year my principal must tell parent[s] my class is full.

Formative Use

The potential for improving teacher quality stems from the utilization of value-added data for formative purposes. SAS claims to provide “easily understandable reporting” (SAS, 2012b, p. 1) that can be used by teachers to strengthen or adjust their teaching practices. Teachers were asked if they had received SAS EVAAS[®] reports for their students, and whether they used the information to inform their instruction. The majority ($n = 480/815$; 58.9%) indicated they do not use SAS EVAAS[®] reports to inform their instruction whatsoever. The teachers ($n = 335/815$; 41.1%) who indicated that they did use SAS EVAAS[®] data, were asked to explain how. Out of that group, 222 teachers provided 238 substantive explanations for how they used SAS EVAAS[®] data to inform their instruction.

The most common response was from teachers ($n = 53/238$; 22.3%) who responded that they knew they were “supposed to” look at their SAS EVAAS[®] reports, so they would look at the reports to get an overview on how the students performed; however, these same teachers called the reports “vague” and “unclear” and they were “not quite sure how to interpret” and use the data to inform instruction. As one sixth grade teacher explained, she looked at her SAS EVAAS[®] report “only to guess as to what to do for the next group in my class.” Another teacher responded:

[I] attempted [to use them] but the reports are not helpful at all. They are a mixture of Stanford and SAKS. I need to know what the anticipated SAKS and Stanford scores are so I can make goals for myself and [my] students; however, since part of EVAAS [is] comparing teachers at like schools, the goal is a moving target.

An advanced English teacher added that the timing of report distribution prevented her from using the SAS EVAAS[®] data, “By the time I get the scores the students are in another grade. I can look at the previous years’ scores, but [the reports] have to be pulled by individual students...This is too time consuming.”

Other teacher participants ($n = 45/238$; 18.9%) described how they used their SAS EVAAS[®] reports, stating that they used the reports for ability grouping to differentiate instruction, whereas, others ($n = 44/238$; 18.5%) reported that they looked at the SAS EVAAS[®] reports to plan for remedial instruction with future students. One fourth grade teacher explained, “If I’m low in one area, I try to maximize instruction in that area...I was low in [social studies] so I tried to incorporate more [social studies] activities into reading lessons.”

However, even among the teachers who indicated that they used SAS EVAAS[®] for ability grouping, differentiating instruction, and remedial education, almost no one actually articulated *how* the data were specifically used. This fourth grade teacher started describing how she used SAS EVAAS[®] reports to look at subgroups, but then revealed she was not quite sure what to do beyond that:

EVAAS is most helpful for me when looking at subgroups and their growth. For example, you can look at the growth of just boys, or girls in your class. You can also look at gifted versus non-gifted. I believe looking at how each subgroup performed is helpful. The only issue is that you're not 100% sure how this score is calculated, so it's not clear what part of your practice you should go back and change.

Another third grade teacher responded, “I do use them, but only to tell me what level these students are on and how much growth they need to make. It is not specific enough to tell me exactly the strength or weakness in each area.”

Other teachers ($n = 24/238$; 10.1%) indicated that they use SAS EVAAS[®] reports to identify the lowest performing kids to pull out for tutoring or remediation, and also the “bubble kids” whom they usually focused their teaching efforts on to try to maximize growth scores. Teachers identified bubble kids as students who performed just below average, with greater relative potential to demonstrate SAS EVAAS[®] growth. As one science teacher who used his SAS EVAAS[®] reports in this way explained, “It allowed me to focus on my bubble students early.” A handful of other teachers ($n = 15/238$; 6.3%) who indicated that they used SAS EVAAS[®] reports to inform their instruction responded that they actually used other data to inform their instruction instead, such as those derived via the Stanford and SAKS.

An additional question about formative use was included on the survey, but this time provided teachers with the opportunity to make multiple selections to describe which of the listed scenarios best described their SAS EVAAS[®] data usage (see Table 1).

Table 1
EVAAS Data Usage by Teachers

| Multiple Selection Scenarios for EVAAS Data Use | <i>N</i> | Percentage |
|--|----------|------------|
| You use other resources (not EVAAS data) to inform practices | 400/882 | 45.4% |
| You use EVAAS data to inform your classroom practices | 242/882 | 27.4% |
| You do not typically use EVAAS data to inform your practices | 220/882 | 24.9% |

Formative Use Support

Teachers were asked if their principal or supervisor typically discussed their SAS EVAAS[®] results with them. Slightly more teachers ($n = 422/868$; 48.6%) responded that their principals did

discuss their SAS EVAAS[®] results with them than those teachers ($n = 397/868$; 45.7%) who did not discuss their SAS EVAAS[®] results with a principal or supervisor. However, in analyzing the 277 substantive explanations provided by those teachers who had discussed SAS EVAAS[®] with their principals, it became clear that not all teachers had similar experiences or discussions.

The most common explanation of such circumstances came from the teacher respondents ($n = 85/277$; 30.7%) who indicated that their principals told or showed them their scores in a manner that was “vague,” “not in depth,” and “not discussed thoroughly.” Of these 85 teachers, 31 specifically indicated that they thought the “very basic discussions” were due to the fact that their principals did not understand the SAS EVAAS[®] reports either. One fifth grade teacher explained, “He looks at them [SAS EVAAS[®] scores], but is unable to explain them.” Another teacher stated that his principal “goes over the data, without much comprehension on how scores are derived. [The principal] cannot suggest improvements.” Another elementary teacher replied, “Our principal does not know how they get the score and has tried many times to get someone to come and explain it to us. No one can.”

Teacher respondents ($n = 56/277$; 20.2%) reported that their principals discussed their SAS EVAAS[®] reports with them at the end of the year during performance evaluations, but teachers did not provide much explanation for these discussions. One multi-grade teacher said their principal discussed SAS EVAAS[®] “during the last conference together at the end of the year. [I] would like better feedback/support in how to improve.” Another bilingual teacher said the SAS EVAAS[®] reports were discussed, “At the end of the year, before we can get hired again.”

A similar number of teachers ($n = 51/277$; 18.4%) indicated that their principals discussed their SAS EVAAS[®] scores in a group setting or team discussion, but not individually with each teacher. Other teachers ($n = 33/277$; 11.9%) reported that their principals discussed their SAS EVAAS[®] reports with them at the beginning of the year to set yearly goals. But out of all the responses, only 4.7% ($n = 13/277$) of the teachers reported that their principals were able to “explain what the scores mean” or tell teachers “how to use the data to improve scores.”

Intended Consequences and Claimed Benefits of SAS EVAAS[®]

The final section of the survey was designed with items meant to gather teacher participants’ perspectives on the intended uses (consequences) and claimed benefits of SAS EVAAS[®], as well as SAS EVAAS[®] statements generated to further capture teacher perception of the model and its use in the district. The Likert-type scale was used to capture teachers’ levels of agreement with the following statements with values, assigned as: Strongly Agree (SA) = 5, Agree (A) = 4, Neither Agree nor Disagree = 3, Disagree (D) = 2, Strongly Disagree (SD) = 1 (Gay, 1996, p. 155; see Table 2).

The descriptive statistics above illustrate that all mean values were between 1.76 and 2.27, which indicates that the average teacher disagreed more than they agreed with each of the SAS EVAAS[®] statements presented to them in this section of the survey instrument. In fact, more than 50% of the teachers disagreed or strongly disagreed with every single statement, and less than 20% of the teachers agreed or strongly agreed with every statement.

The teachers disagreed most with statement 13, “EVAAS will enhance working conditions” with 75.7% ($n = 637/842$) of the teachers disagreeing or strongly disagreeing with this assertion. Similarly, 72.9% ($n = 619/849$) of the teachers disagreed or strongly disagreed that “EVAAS will validly identify and help remove ineffective teachers,” and 72.5% ($n = 611/842$) of the teachers disagreed or strongly disagreed or agreed that “EVAAS will enhance the school environment.” A

table for all Likert-type items responses, including the number and proportion of respondents for each statement can be found in Appendix B.

Table 2

Items Capturing Respondents' Opinions About SAS EVAAS Statements

| | Statement | <i>N</i> | <i>M</i> | <i>SD</i> |
|----|---|----------|----------|-----------|
| 2 | EVAAS helps create professional goals | 870 | 2.27 | 1.25 |
| 3 | EVAAS helps improve instruction | 864 | 2.24 | 1.23 |
| 11 | EVAAS will provide incentives for good practices | 860 | 2.19 | 1.24 |
| 5 | EVAAS ensures growth opportunities for very low achieving students | 875 | 2.15 | 1.18 |
| 4 | EVAAS ensures growth opportunities for students | 873 | 2.14 | 1.16 |
| 7 | EVAAS helps increase student learning | 868 | 2.13 | 1.16 |
| 8 | EVAAS helps you become a more effective teacher | 869 | 2.12 | 1.21 |
| 15 | Overall, the EVAAS is beneficial to my school | 855 | 2.10 | 1.22 |
| 1 | EVAAS reports are simple to use | 866 | 2.09 | 1.14 |
| 14 | Overall, the EVAAS is beneficial to me as a teacher | 858 | 2.08 | 1.25 |
| 16 | Overall, the EVAAS is beneficial to the district | 847 | 2.08 | 1.23 |
| 6 | EVAAS ensures growth opportunities for very high achieving students | 870 | 2.06 | 1.14 |
| 10 | EVAAS will identify excellence in teaching or leadership | 849 | 2.00 | 1.15 |
| 9 | EVAAS will validly identify and help to remove ineffective teachers | 849 | 1.88 | 1.10 |
| 12 | EVAAS will enhance the school environment | 842 | 1.86 | 1.11 |
| 13 | EVAAS will enhance working conditions | 842 | 1.76 | 1.04 |

Note. Items are arranged by *M* in descending value.

As mentioned throughout this study, there was a high volume of teachers' union members ($n = 612/882$; 69.4%) represented in the sample. As such, chi-square analyses were calculated for each of the survey questions with categorical responses to examine whether the perceptions among the two mutually exclusive groups, union and non-union teachers, differed at statistically significant levels (Gay, 1996). Table 3 represents results from the chi-square analysis for Statement 1 below.

Table 3
Chi-square Analysis for Statement 1

| Statement and Chi-square result | Strongly disagree | Disagree | Neutral | Agree | Strongly agree | Total |
|--|-------------------|---------------|---------------|---------------|----------------|----------------------|
| 1: EVAAS reports are simple to use; $\chi^2 = (4, N = 866) = 1.96, p = .744$ | | | | | | |
| Non-union | 111 (12.8) | 74 (8.5) | 39 (4.5) | 39 (4.5) | 5 (0.6) | $n = 268$ (30.9%) |
| Union | 250 (28.9) | 147 (17.0) | 107 (12.4) | 83 (9.6) | 11 (1.3) | $n = 598$ (69.1%) |
| Total | 361 (41.7) | 221 (25.5) | 146 (16.9) | 122 (14.1) | 16 (1.8) | $n = 866$ (100%) |

Table 3 shows that the chi-square value of 1.96 with 4 degrees of freedom is not significant at conventional significance levels ($p = 0.744 > 0.05$). This signifies that there is no statistical difference between non-union and union teachers on their agreement (or disagreement) with the statement that SAS EVAAS reports are simple to use. In fact, none of the numerical statements included in the survey instrument yielded statistically significant differences between union and non-union members' responses ($p < 0.05$). For all related results, see Appendix C for chi-square tables for statements 2-16 above, and see Appendix D for all other chi-square tables pertaining to all other numerical items included in the survey instrument.

Unintended Consequences

Many unintended consequences were discovered via the aforementioned analyses that seem to be occurring as a result of SAS EVAAS[®] implementation in SSD. Such unintended consequences include: disincentives for teaching certain student groups; teacher mobility issues with teachers looping or teaching back-to-back grade levels and switching grade levels within the same content areas; gaming the system or teaching to the test as a result of accountability pressures; general distrust of the SAS EVAAS[®] model, competition and low morale among teachers, both of which are foreseen as perceived effects from SAS EVAAS[®] implementation.

Disincentives for Teaching Certain Groups of Students

Again, as evidenced throughout the study, teachers identified working with certain populations of students as problematic if they were to achieve high SAS EVAAS[®] scores. Specifically, high performing and gifted students who are inhibited by ceiling effects, transition students who are in their first year of English-only instruction, and teachers in classrooms with high proportions of special education and ELL students were of exceptional concern. As one teacher said, "it is extremely difficult to raise test scores for [gifted] students." A third grade teacher described her frustration with low SAS EVAAS[®] scores, stating that she is being "punished for teaching ELL and [gifted] students." If it were possible, the same teacher noted, "I would refuse to teach ELL and [gifted] students."

Teacher Mobility Issues

Though SAS EVAAS[®] claims to be able to account for teacher mobility (SAS, 2012c) results from this study also provide evidence indicating that in SSD it is pretty common to loop, or teach

the same content area in back-to-back grade levels. When these teachers have the same students within their classrooms for back-to-back years, teacher respondents expressed difficulty showing SAS EVAAS[®] growth two years in a row. One such teacher noted, “My scores have always varied from the absolute highest to the absolute lowest, even when I taught the same exact kids two years in a row.” A middle school teacher said, “I teach many of [the] same students in 7th and 8th [grades]. In 7th I show growth, then in 8th [I] suffer.”

Other teachers described the flip-flop effect (see Amrein-Beardsley & Collins, 2012), whereas a teacher rated as effective by SAS EVAAS[®] would switch grade levels with an ineffective teacher, and his/her SAS EVAAS[®] ratings would flip-flop from the mere move. A fifth grade teacher reported, “I taught social studies to my [5th] grade homeroom class and I was below district expectations. Previously, when I taught it to 4th graders I was considered above expectations.”

Gaming and Teaching to the Test

As a result of the pressures teachers are under to obtain high SAS EVAAS[®] scores, some respondents also indicated that there was evidence of gaming the system and other unprofessional and unethical behaviors occurring as principals and teachers tried to increase SAS EVAAS[®] scores. The various scenarios described by teachers spanned from befriending principals, to hand-selecting their own class rosters, to admitting to a drill and kill teaching approach to ensure students know the material for the high-stakes tests. One middle school teacher claimed that, “EVAAS is creating a very competitive setting. The teachers want to recruit the best profiles. There are conversations ‘during the summer’ to obtain the best rosters.” A fourth grade teacher described the opposite scenario with principals, “If they don't like you they stack [your roster with] the students with issues, give you no support and crucify you with EVAAS. It's a set up.”

As referenced by teachers throughout the findings of this study, teachers discussed “drill and kill” teaching approaches, “teaching to the test,” and reported knowing that “teachers cheat” to increase their SAS EVAAS[®] scores. One veteran teacher explained, “If [two] or more teachers can work together to cheat with each others’ students then they stand to profit \$7,000-\$10,000 per school year. That's upwards 3 times what could be made teaching one session of summer school.” An elementary teacher claimed, “You must be willing to teach strictly to the test, or be willing to cheat because that is the only way your [EVAAS] data will look good.” Yet another described, “To gain the highest EVAAS score, drill and kill and memorization yields the best results, as does teaching to the test.”

Numerous teachers reflected on their own questionable practices. As one English teacher said, “When I figured out how to teach to the test, the scores went up.” A fifth grade teacher added, “Anything based on a test can be ‘tricked.’ EVAAS leaves room for me to teach to the test and appear successful.”

However, teachers were also cognizant of the negative impacts that teaching to the test has on students. A veteran teacher claimed, “There is no real teaching anymore because the scores obsession is driving teachers to teach to the test. Students are learning to bubble an answer sheet instead of learning to think and reason.” A math teacher expanded on this idea:

As a result of the emphasis on EVAAS, we teach less math, not more. Too much drill and kill and too little understanding [for the] love of math. Students who have come up with this in place are remarkably less likely to think and inquire and innovate, and more likely to sit-and-get. Raising a generation of children under these circumstances seems best suited for a country of followers, not inventors, not world leaders.

Another teacher took these concerns one step further and discussed students' long-term well-being and success:

When they arrive at college, they are unprepared to write, read, take higher level assessments since the curriculum has been "dumbed down" to make sure that the students pass, and whatever the cost...

Distrust, Competition, and Low Morale

Lack of transparency surrounding the SAS EVAAS[®] model and data appears to have created a sense of distrust among teachers as well. An eighth grade teacher shared her sentiments:

Ultimately, there are no stated metrics and as such I don't trust that the people who assign this number are using this in my or my school's best interest. To use the lingo, the current system is not transparent. That makes me more resistant to data [or] a system that has the potential to be very useful for testing.

One math teacher acknowledged the sophistication of the SAS EVAAS[®] metrics, but added that he was skeptical of its usefulness, "I don't completely believe in it or trust that the calculations are valid. And even if the whole EVAAS operation is mathematically sound, I'm still not sure if it is all that important." Another added, "Since I don't find the reports consistent with my instruction, effort and quality of practice, I don't trust EVAAS reports."

A third grade teacher expanded on how distrust has impacted teacher collaboration and may be harming students:

Since the inception of the EVAAS system, teachers have become even more distrustful of each other because they are afraid that someone might steal a good teaching method or materials from them and in turn earn more bonus money. This is not conducive to having a good work environment, and it actually is detrimental to students because teachers are not willing to share ideas or materials that might help increase student learning and achievement.

Otherwise, it seems that the bonuses attached to SAS EVAAS[®] output have also lowered morale and created a sense of competition among teachers. This is also an unintended consequence, and a veteran teacher captured these teacher respondents' views best by noting, "It [EVAAS] trades 'it takes a village' for 'every man for himself.'" A fourth grade teacher mentioned, "This system undermines collaboration, a cooperative work place, and pits administration against the staff." Yet another teacher referenced the competition that has emerged as a result of using SAS EVAAS[®] scores for the merit-pay program by writing:

The incentive program is not an incentive. For something to be an incentive, you need to know what you have to do to get the incentive. All we know is that as a teacher you have to improve your scores more than the other teachers. You can make improvements each year, but if other teachers improve the same amount, you have made no gains according to the system. It is a constantly moving target. You don't know what you need to do to get the "prize" until after the "contest" is over.

An elementary teacher who had taught in the district for more than 21 years described her own weakened morale and how even non SAS EVAAS[®]-eligible teachers in SSD have been impacted. She wrote:

EVAAS and the bonuses attached to it are tearing down the morale of our school. Before, we worked as a team to get our kids where they needed to be...I see the same [negative] attitude with our lower grade teachers. They feel like they are chopped liver compared to the testing grades. We need tutors to help out

with our struggling kids in the testing grades, and usually we rely on our lower grade teachers to help out. This year, we can't beg, borrow, or steal anyone to stay after school or come in on Saturdays. Our upper grade teachers are barely running on steam, and our lower grade teachers feel unappreciated and disenfranchised, and say, "They're getting the big bucks, let them earn it." This is not a business, this is education. There is no formula or secret recipe that is fool-proof. These are kids...They don't even know themselves yet... and my career depends on how they do on a test that they take one day out of the 3 years that I teach them science?

Findings and Implications

Reliability

As discussed in related literature (Baker et al., 2010; Corcoran, 2010; EPI, 2010; Otterman, 2010; Schochet & Chiang, 2010) and preliminary studies in SSD (Amrein-Beardsley & Collins, 2012), it was evident that inconsistent SAS EVAAS[®] scores year-to-year were an issue of concern. According to teachers who participated in this study, reliability as measured by consistent SAS EVAAS[®] scores year-to-year was ironically, an inconsistent reality. About half of the responding teachers reported consistent data whereas the other half did not, just like one would expect with the flip of a coin (see also Amrein-Beardsley & Collins, 2012). Similarly, teachers reported split consistencies of SAS EVAAS[®] scores across grade levels and different subject areas taught (LeClaire, 2011), as well as given varied student characteristics (Hill et al., 2011; Newton, Darling-Hammond, Haertel, & Thomas, 2010; Rothstein, 2009). Teachers who reported SAS EVAAS[®] inconsistencies identified students as the main cause for the score fluctuations they observed. These teachers specifically mentioned the impact that motivation, behavior, prior academic preparation and demographic influences such as family support and home life all have on SAS EVAAS[®] scores, which one would think is obvious, but this is contrary to what SAS EVAAS[®] creators indicate can be “statistically controlled for” (see Sanders & Horn, 1998). The consensus among teachers was that gifted, transition, ELL and special education students were the most difficult student groups to demonstrate high levels of growth as measured by SAS EVAAS[®]. Even with the most sophisticated controls and blocks, it appears that SAS EVAAS[®] cannot control for the impact of extraneous variables such as home life, health, behavior, motivation, etc. on student achievement (see also Haertel, 2011; Harris, 2011; Rothstein, 2009).

Reliability Implications

Unless school districts could prevent teacher mobility and ensure equal, random student assignment, it appears that EVAAS is unable to produce reliable results, at least greater than 50% of the time. As such, it is highly inappropriate and invalid for SSD (and any other district) to use unreliable SAS EVAAS[®] results for anything since a teacher seemingly has the same probability of being rated “effective” or “ineffective” as (s)he would calling “heads” during a coin toss. If SAS EVAAS[®], the “most comprehensive and reliable” VAM available (SAS, 2012a), produces such unreliable results as reported by these SSD teachers, it would appear highly unlikely that any other VAM could further reduce the risk of misclassifying teachers. The discussion of “acceptable” reliability levels remains a highly debated topic, with studies using multi-year VAM scores producing correlations within the range of $0.3 \leq r \leq 0.4$ (McCaffrey, Sass, Lockwood, & Mihaly, 2009; Kane & Staiger, 2012; Lockwood & McCaffrey, 2009; Newton et al., 2010). Therefore, a statistical model used to evaluate teachers based on student test data, will likely never have an acceptable level of

reliability, and accordingly will likely always be inappropriate to use for classifying teachers or to inform consequential decisions.

Further, no matter how much more sophisticated the statistical model becomes, the reality is that human factors and life circumstances inherently impact a student's ability to learn, and cannot be "controlled for" or deduced from a one-size-fits-all equation. And as Kupermintz (2003) explained, statistical blocks or controls, such as those used in the SAS EVAAS[®] model, were intended to be used in controlled experiments that require random assignment of students for verification. Random student assignment is not assumed to be common practice in schools (Guarino, Reckase, & Wooldridge, 2012) with one recent state-wide study showing evidence that random assignment is not a common practice within schools (Paufler & Amrein-Beardsley, 2013).

Validity

The areas of concern came from the teacher participants who were primarily apprehensive with allocating instructional time among multiple teachers and whether student mobility, in and out of their classrooms, could indeed be controlled with fractional and proportional statistics (see also Corcoran, 2010; Ishii & Rivkin, 2009; Kane & Staiger, 2008; Kennedy, 2010; Nelson, 2011; Papay, 2010; Rothstein, 2009). Although SAS EVAAS[®] can purportedly account for team-teaching dynamics (Sanders & Horn, 1994), it is questionable whether the mathematical proportioning of instruction time without considering the interaction effects of multiple teachers is actually possible (see Amrein-Beardsley & Collins, 2012).

In terms of criterion-related evidence of validity, teachers described the relationship between their SAS EVAAS[®] scores and principal evaluation scores, both of which are considered the main measures of teacher quality in SSD. More than half of the teachers reported that the two evaluation scores did not typically match. The plurality of teachers indicated their observation scores were always higher than their SAS EVAAS[®] scores, and that their observation scores remained consistent year-to-year while their SAS EVAAS[®] scores fluctuated. Such findings could reflect the subjectivity of the more traditional principal evaluation method, which is believed to lack distinguishability, and largely overestimate the number of effective teachers (see Weisberg, Sexton, Mulhern, & Keeling, 2009).

Related, from the perspective of some of the teachers, the principals viewed SAS EVAAS[®] as the more objective evaluation score, and would accordingly adjust teacher observation scores (either higher or lower) to reflect their SAS EVAAS[®] scores. This perspective potentially confounds the criterion-related validity between both measures. Although researchers highly recommend that value-added output correlate with at least one other measure of teacher effectiveness to increase trustworthiness (AERA, APA, & NCME, 1999; Baker et al., 2010; Harris, 2011; Hill et al., 2011), such intentional adjustment of scores from one measurement to reflect those of the VAM would completely negate this rationale, yet there is evidence of this occurring elsewhere as well (Garland, 2012; Ravitch, 2012).

Validity Implications

Although SSD uses two different tools to evaluate teacher effectiveness: SAS EVAAS[®] and principal evaluations, and although researchers encourage the use of multiple measures to increase validity (AERA, APA, & NCME, 1999; Baker et al., 2010; Harris, 2011; Hill et al., 2011), having two measures that produce conflicting results approximately half of the time, reduces the validity of both measures and sends conflicting messages to teachers. Further, teachers can only truly assess their work when they have a clear understanding of the targets that their teaching practices are meant to achieve, and when two indicators of teacher quality produce conflicting results, the targets become even more blurred. This is an important issue to consider as states and districts try to follow

recommendations of incorporating multiple measures of teacher quality, recommendations that are currently most popular among academics and researchers (see, for example, Harris, 2011; Hill et al., 2011; Kane & Staiger, 2012; Sass & Harris, 2012); though not enough has been done to this point to determine what level of correlations among multiple measures are appropriate enough to indicate validity. Meanwhile, it seems even two measures of teacher quality cannot be trusted to determine whether a teacher is “effective” or “ineffective,” especially when one appears to influence or trump the value of the other.

Formative Use

Data alone cannot improve teacher quality; it is what teachers do with the data that has the potential to make a difference. Almost 60% of the teachers in this study reported that they do not use their SAS EVAAS[®] data for formative purposes whatsoever. Teachers who did report using SAS EVAAS[®] data referenced using other data resources in combination to inform their instructional practices, and many indicated that they used other data instead, not SAS EVAAS[®] output, to inform their practices. Of this roughly 40% of teachers who did report using SAS EVAAS[®] data, the majority called the reports “vague” and “unclear” and the teachers were “not quite sure how to interpret” or use the data to inform their instruction (see also Eckert & Dabrowski, 2010; Harris, 2011).

Teacher respondents reported relying on or looking to their principals for SAS EVAAS[®] information and explanations. Almost half of the teachers indicated that they typically discussed their SAS EVAAS[®] results with their principals, although the other half did not. Among those who did discuss their SAS EVAAS[®] reports with their principals, very few indicated that their principals were able to provide specific information on how they might use the data to improve instruction, however (see also Eckert & Dabrowski, 2010; Harris, 2011). Many teachers believed the “basic discussions” resulted from their principals not understanding SAS EVAAS[®] either, or definitely not understanding SAS EVAAS[®] well enough to explain it to their own teachers. Without principal understanding and buy-in, value-added data are essentially worthless (Kennedy, Peters, & Thomas, 2012), assuming, that the data have value in the first place.

Formative Use Implications

As a result of a cohesive district-wide plan, principals should be provided resources so that they become better equipped at understanding the SAS EVAAS[®] reports. Accordingly, principals might become more able to provide their teachers with specific actions and goals that incorporate the data, develop regular routines to discuss such data, plans and goals with the teachers (Kennedy, et al., 2012), and ensure that all teachers are aware of available training sessions provided by the district. This, however, follows the assumption that the SAS EVAAS[®] data are comprehensible and meaningful, which data from this study contradict.

Nonetheless, other districts and states looking to implement a VAM should realize that principals are fundamental in such plans, particularly as the instructional leaders of their schools. As such, they must not only be knowledgeable about the VAM, but informed of its fine intricacies and related literature base; that is, the academic literature and not just the literature base advanced by the VAM corporations sponsoring the VAM. Accordingly, principals must be supportive of teachers and encourage the use of these and other data to not only inform their practices, but also question, for example, when things do not make sense. This would increase teachers’ and administrators’ capacities to become critical consumers. Formative use is the culmination of VAMs, and many, including policymakers, assume that simply enacting legislation, which requires states and districts to use such models for summative purposes will simultaneously result in greater levels of formative data use.

Intended Consequences

The large majority of the teachers in this sample strongly disagreed with SAS EVAAS[®] marketing claims and statements. This provides solid evidence that the majority of the teacher respondents do not believe that the SAS EVAAS[®] works in the ways in which both Dr. Sanders and SAS have advertised, to not only SSD at the rate of \$500,000 per year (Amrein-Beardsley & Collins, 2012), but to many other states and districts across the country. Overwhelmingly, teacher respondents reported not believing that the SAS EVAAS[®] model has benefitted much of anything (see, again, each statement listed in Table 3 with levels of disagreement).

Intended Consequences Implications

This signifies that other districts and states need to be, again, critical consumers, and ask for preferentially peer-reviewed evidence to provide accurate, unbiased, and research-based insight into what VAMs look like in practice. It is one thing to judge a book by its cover, or to read the foreword written by an author's friend, but another completely to read the Consumer Reports, from those who have used the product. In this case that means looking beyond the proprietary company's literature and research on the VAM and gathering feedback from teachers—the real “consumers” of VAMs.

Unintended Consequences

Throughout teachers' reported experiences and perceptions about SAS EVAAS[®] within SSD, several unintended consequences were also uncovered. As mentioned, teachers repeatedly identified specific groups of students (e.g., gifted, ELL, transition, special education) that typically demonstrated little to no SAS EVAAS[®] growth. Other teachers described various teaching scenarios such as teaching back-to-back grade levels or switching grade levels which negatively impacted their SAS EVAAS[®] scores. Such reports contradict Dr. Sanders' claim that a teacher in one environment is equally as effective in another (LeClaire, 2011). Also a result of the pressure placed on SAS EVAAS[®] scores, teachers admitted that they “drill and kill,” teach to the test, or even cheat to effectively, although artificially (Haladyna, Nolen, & Hass, 1991), raise their SAS EVAAS[®] scores. This not only highlighted the fact that teachers believe the SAS EVAAS[®] model produces bias results, but it also demonstrated that teachers believed it can be manipulated or influenced by various criteria and characteristics of the students assigned to their classrooms (see also Braun, 2005; Hill et al., 2011; Kupermintz, 2003; Rothstein, 2010).

Likewise, teachers explained how SAS EVAAS[®] has created a sense of competition among teachers and has distorted collaboration, for example, when teachers realize that their efforts will go unrecognized and unrewarded, particularly if their actions may contribute to another's SAS EVAAS[®] scores. Researchers have implied such competition could occur when VAMs are used for high-stakes consequences, especially monetary compensation (Harris, 2011; Kennedy et al., 2012), but this remains relatively unexplored. Related, teachers reported that the overall focus on SAS EVAAS[®] scores has lowered morale in their schools as teachers feel overworked and underappreciated.

Unintended Consequences Implications

As the first study to examine what SAS EVAAS[®] looks like in practice from the perspectives and experiences of SSD teachers, many negative, unintended consequences were discovered as a result of SAS EVAAS[®] use, especially given the high-stake consequences attached to SAS EVAAS[®] output by the district. The evidence here should alarm district administrators, as SAS EVAAS[®] appears to be doing more harm than good, at least among these teachers, and is potentially preventing students from realizing a well-rounded education.

Conclusions

SAS EVAAS[®] and other VAMs, by themselves, are sophisticated statistical models that purportedly provide diagnostic information about student academic growth, and represent teachers' value-add. In other words, SAS EVAAS[®] and VAMs are tools. It is what teachers, schools, districts, and states *do* with this information that matters most. However, for the teachers in this study, even for those participating in training sessions on how to use the data, the SAS EVAAS[®] data alone were unclear and virtually unusable. For SSD, not only are teachers not using the “product” that costs the district half a million dollars per year, but teachers are aware that SAS EVAAS[®] inputs can be manipulated based on the student makeup of their classroom, and some teachers even confess to teaching to the test and cheating in attempt to increase their SAS EVAAS[®] scores.

The results from this study provide very important information of which not only SSD administrators should be aware, but also any other administrators from districts or states currently using or planning to use a VAM for teacher accountability. Although high-stakes use certainly exacerbates such findings, it is important to consider and understand that unintended consequences will accompany the intended consequences of implementing SAS EVAAS[®], or likely any other VAM. Reminiscent of Campbell's law⁶, the overreliance on value-added assessment data (assumed to have great significance) to make high-stakes decisions risks contamination of the entire educational process, for students, teachers and administrators (Nichols & Berliner, 2007).

Accordingly, these findings also strongly validate researchers' recommendations to not use value-added data for high-stakes consequences (Eckert & Dabrowski, 2010; EPI, 2010; Harris, 2011). While the SAS EVAAS[®] model's vulnerability as expressed by the SSD EVAAS[®]-eligible teachers is certainly compounded by the district's high-stakes use, the model's reliability and validity issues combined with teachers' feedback that the SAS EVAAS[®] reports do not provide sufficient information to allow for instructional modification or reflection, would make it seem inappropriate at this point to use value-added data for anything.

Yet the federal government continues to pressure and incentivize state and local governments to “race to the top,” while trying to convince the general public that these accountability models will reform education by eliminating ineffective teachers from the system, and consequently, albeit purportedly, lead to higher student achievement. The trend to adopt VAMs appears to be occurring via commands and promised federal dollars instead of implementing such policies in a holistic manner, which encourages and values the input and support of teachers, not to mention the growing research base surrounding such initiatives. The failure to consider and incorporate the perspectives and realities of teachers, specifically those who are experiencing these models first-hand in practice, will likely result in yet another one of education's “classic swing of the pendulum...the cycle of early enthusiasm, widespread dissemination, subsequent disappointment, and eventual decline” (Slavin, 1989, p. 752). The disappointment appears to be mounting among teachers, administrators, and many academic researchers, but the decline of these teacher evaluation policies, unfortunately, is not yet on the horizon.

⁶ In 1976, social scientist Donald T. Campbell wrote a paper describing the more that a quantitative indicator is relied on for decision-making, the more likely it is to corrupt and distort the social process it was intended to define or represent (see Campbell, 1976).

References

- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Amrein-Beardsley, A. (2008). Methodological concerns about the education value-added assessment system. *Educational Researcher*, 37, 65-75. <http://dx.doi.org/10.3102/0013189X08316420>
- Amrein-Beardsley, A., & Collins, C. (2012). The SAS Education Value-Added Assessment System (SAS® EVAAS®) in the Houston Independent School District (HISD): Intended and unintended consequences. *Education Policy Analysis Archives*, 20(12).
- Babbie, E. (1990). *Survey research methodology* (2nd ed.). Belmont, CA: Wadsworth Publishing Co.
- Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., ... Shepard, L. A. (2010). *Problems with the use of student test scores to evaluate teachers*. Washington, DC: Economic Policy Institute. Retrieved from <http://www.epi.org/publications/entry/bp278>
- Braun, H. I. (2005). *Using student progress to evaluate teachers: A primer on value-added models*. Princeton, NJ: Educational Testing Service.
- Campbell, D. T. (1976). Assessing the impact of planned social change. *Evaluation and program planning*, 2(1), 67-90. [http://dx.doi.org/10.1016/0149-7189\(79\)90048-X](http://dx.doi.org/10.1016/0149-7189(79)90048-X)
- Collins, C., & Amrein-Beardsley, A. (2014). Putting growth and value-added models on the map: A national overview. *Teachers College Record*, 16(1). Retrieved from <http://www.tcrecord.org/Content.asp?ContentId=17291>
- Corcoran, S. P. (2010). *Can teachers be evaluated by their students' test scores? Should they be? The use of value-added measures of teacher effectiveness in policy and practice*. Providence, RI: Annenberg Institute for School Reform. Retrieved from <http://www.annenberginstitute.org/products/Corcoran.php>
- Creative Research Systems. (n.d.). *The survey system*. Retrieved from <http://www.surveysystem.com/sscalc.htm>
- Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: Sage.
- Eckert, J., & Dabrowski, J. (2010). Should value-added measures be used for performance pay? *Phi Delta Kappan*, 91, 88-92. <http://dx.doi.org/10.1177/003172171009100821>
- Economic Policy Institute. (2010). *Problems with the use of student test scores to evaluate teachers*. Washington, D.C.: Baker, Barton, Darling-Hammond, Haertel, Ladd, Linn, Ravitch, Rothstein, Shavelson, Shepard.
- Garland, S. (2012). Tennessee teacher evaluation systems have rough road ahead. *The Huffington Post*. Retrieved from http://www.huffingtonpost.com/2012/02/07/tennesseeteacher-evaluat_n_1260790.html?page=1
- Gay, L. R. (1996). *Educational research: Competencies for analysis and application* (5th ed.). Upper Saddle River, NJ: Prentice-Hall, Inc.
- Glaser, B. G., & Strauss, A. L. (1967). *The Discovery of grounded theory: Strategies for qualitative research*. Chicago, IL: Aldine Publishing Company.
- Greene, J. C. (2008). Is mixed methods social inquiry a distinctive methodology? *Journal of Mixed Methods Research*, 2, 7-22. <http://dx.doi.org/10.1177/1558689807309969>
- Guarino, C. M., Reckase, M. D., & Wooldridge, J. M. (2012). *Can value-added measure of teacher education performance be trusted?* East Lansing, MI: Education Policy Center. Retrieved from http://education.msu.edu/epc/library/documents/WP18Guarino-Reckase-Wooldridge-2012-Can-Value-Added-Measures-of-Teacher-Performance-Be-T_000.pdf

- Haertel, E. (2011). *Using student test scores to distinguish good teachers from bad*. Paper presented at the Annual Conference of the American Educational Research Association (AERA), New Orleans, LA.
- Haladyna, T., Nolen, S. B., & Haas, N. S. (1991). Raising standardized achievement test scores and the origins of test score pollution. *Educational Researcher*, 20, 2-7.
<http://dx.doi.org/10.3102/0013189X020005002>
- Harris, D. N. (2011). *Value-added measures in education: What every educator needs to know*. Cambridge, MA: Harvard Education Press.
- Hill, H. C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, 48, 794-831.
<http://dx.doi.org/10.3102/0002831210387916>
- Holloway-Libell, J., & Collins, C. (2014). VAM-based teacher evaluation policies: Ideological foundations, policy mechanisms, and implications. *InterActions: UCLA Journal of Education and Information Studies*, 10(1). Retrieved from <http://escholarship.org/uc/item/1pm161v8>
- Horsburgh, D. (2003). Evaluation of qualitative research. *Journal of Clinical Nursing*, 12, 307-312.
<http://dx.doi.org/10.1046/j.1365-2702.2003.00683.x>
- Ishii, J., & Rivkin, S. G. (2009). Impediments to the estimation of teacher value added. *Education Finance and Policy*, 4, 520-536. <http://dx.doi.org/10.1162/edfp.2009.4.4.520>
- Kane, T. J., & Staiger, D. O. (2008). *Are teacher-level value-added estimates biased? An experimental validation of non-experimental estimates*. Paper presented at the National Conference on Value-Added Modeling, Madison, WI.
- Kane, T., & Staiger, D. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Seattle, WA: Bill & Melinda Gates Foundation. Retrieved from http://www.metproject.org/downloads/Preliminary_Findings_Research_Paper.pdf
- Kennedy, M. M. (2010). Attribution error and the quest for teacher quality. *Educational Researcher*, 39, 591-598. <http://dx.doi.org/10.3102/0013189X10390804>
- Kennedy, K., Peters, M., & Thomas, M. (2012). *How to use value-added analysis to improve student learning: A field guide for school and district leaders*. Thousand Oaks, CA: Sage.
- Kupermintz, H. (2003). Teacher effects and teacher effectiveness: A validity investigation of the Tennessee value-added assessment system. *Educational Evaluation and Policy Analysis*, 25, 287-298. <http://dx.doi.org/10.3102/01623737025003287>
- LeClaire, B. (2011). Will EVAAS[®] make Wake schools better? *Raleigh Public Record*. Retrieved from <http://www.raleighpublicrecord.org/news/2011/06/01/will-evaas-make-wake-schools-better-part-ii/>
- Lincoln Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage.
- Lockwood, J. R., & McCaffrey, D. F. (2009). Exploring student-teacher interactions in longitudinal achievement data. *Education Finance and Policy*, 4(4), p. 439-467.
<http://dx.doi.org/10.1162/edfp.2009.4.4.439>
- Lowrey, A. (2012). Big study links good teachers to lasting gain. *The New York Times*. Retrieved from <http://www.nytimes.com/2012/01/06/education/big-study-links-good-teachers-to-lasting-gain.html?emc=eta1-commentsh>
- Maxwell, J. A. (1941). *Qualitative research design: An interactive approach* (2nd ed.). Thousand Oaks, CA: Sage.
- McCaffrey, D. F., Sass, T. R., Lockwood, J. R., & Mihaly, K. (2009). The intertemporal variability of teacher effect estimates. *Education Finance and Policy*, 4(4), 572-606.
<http://dx.doi.org/10.1162/edfp.2009.4.4.572>

- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 5-11. <http://dx.doi.org/10.3102/0013189X018002005>
- National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U.S. GPO.
- Nelson, F. H. (2011). *A guide for developing growth models for teacher development and evaluation*. Paper presented at the Annual Conference of the American Educational Research Association (AERA), New Orleans, LA.
- Newton, X. A., Darling-Hammond, L., Haertel, E., & Thomas, E. (2010). Value-added modeling of teacher effectiveness: An exploration of stability across models and contexts. *Education policy analysis archives*, 18. Retrieved from <http://epaa.asu.edu/ojs/article/view/810http>
- Nichols, S. L., & Berliner, D. C. (2007). *Collateral damage: How high stakes testing corrupts America's schools*. Cambridge, MA: Harvard Education Press.
- Otterman, S. (2010). Hurdles emerge in rising effort to rate teachers. *New York Times*. Retrieved from <http://www.nytimes.com/2010/12/27/nyregion/27teachers.html>
- Papay, J. P. (2010). Different tests, different answers: The stability of teacher value-added estimates across outcome measures. *American Educational Research Journal*, 48(1), 163-193. <http://dx.doi.org/10.3102/0002831210362589>
- Paufler, N. A., & Amrein-Beardsley, A. (2013). The random assignment of students into elementary classrooms: Implications for value-added analyses and interpretations. *American Educational Research Journal*, 51(2), 328-362. <http://dx.doi.org/10.3102/0002831213508299>
- Philips, R. H. (2012, Sep. 10). No child left behind waiver sought by Alabama, six other states. *Press Register*. Retrieved from http://blog.al.com/live/2012/09/no_child_left_behind_waiver_so.html
- Popay, J., Rogers, A., & Williams, G. (1998). Rationale and standards for the systemic review of qualitative literature in health services research. *Qualitative Health Research*, 8(3), 341-351. <http://dx.doi.org/10.1177/104973239800800305>
- Ravitch, D. (2012). No student left untested. *The New York Review of Books*. Retrieved from <http://www.nybooks.com/blogs/nyrblog/2012/feb/21/no-student-left-untested/>
- Rothstein, J. (2009). *Student sorting and bias in value-added estimation: Selection on observables and unobservables*. Cambridge, MA: The National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w14607>
- Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *Quarterly Journal of Economics*, 125, 175-214. <http://dx.doi.org/10.1162/qjec.2010.125.1.175>
- Sanders, W. L., & Horn, S. P. (1994). The Tennessee value-added assessment system (TVAAS): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education* 8, 299-311. <http://dx.doi.org/10.1007/BF00973726>
- Sanders, W. L., & Horn, S. P. (1998). Research findings from the Tennessee value-added assessment system (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education*, 12, 247-256. <http://dx.doi.org/10.1023/A:1008067210518>
- SAS. (2012a). SAS® EVAAS® for K-12: Assess and predict student performance with precision and reliability. Retrieved from <http://www.sas.com/govedu/edu/k12/evaas/index.html>
- SAS. (2012b). *SAS EVAAS for K-12 Fact Sheet*. Retrieved from <http://www.sas.com/resources/factsheet/education-evaas-factsheet.pdf>
- SAS. (2012c). *SAS EVAAS for K-12 Overview Brochure*. Retrieved from http://www.sas.com/resources/product-brief/SAS_EVAAS_for_K-12.pdf

- Sass, T., & Harris, D. (2012). Skills, productivity and the evaluation of teacher performance. W. J. Usery Workplace Research Group Paper No. 2012-3-1. Retrieved from <http://ssrn.com/abstract=2020717> or <http://dx.doi.org/10.2139/ssrn.2020717>
- Schochet, P. Z., & Chiang, H. S. (2010). *Error rates in measuring teacher and school performance based on student test score gains*. U.S. Department of Education. Retrieved from <http://ies.ed.gov/ncee/pubs/20104004/>
- Slavin, R. E. (1989). PET and the pendulum: Faddism in education and how to stop it. *Phi Delta Kappan*, 70, 752-758.
- Stake, R.E., & Trumbull, D. (1982). Naturalistic generalizations. *Review Journal of Philosophy and Social Science*, 7, 1-12.
- Strauss, A. L., & Corbin, J. (1998). *Basics of qualitative research: Grounded theory procedures and techniques* (2nd ed.). Thousand Oaks, CA: Sage.
- Sparks, S. D. (2011). Value-added formulas strain collaboration. *Education Week*. Retrieved from <http://edweek.org/ew/articles/2011/11/16/12collab-changes.h21.html?tkn=OVMFb8PQXxQi4wN6vpelNlr7+NhOFCbi71mI&intc=es>
- U.S. Department of Education. (2009). *Race to the Top program: Executive summary*. Washington, DC: Retrieved from <http://www2.ed.gov/programs/racetothetop/executive-summary.pdf>
- U.S. Department of Education (2012a). *Awards*. Washington, DC: Retrieved from <http://www2.ed.gov/programs/racetothetop/awards.html>
- U.S. Department of Education (2012b). *Education department announces 16 winners of Race to the Top – district competition*. Washington, DC: Retrieved from <http://www.ed.gov/news/press-releases/education-department-announces-16-winners-race-top-district-competition>
- U.S. Department of Education (2013). *Awards—race to the top district*. Washington, DC: Retrieved from <http://www2.ed.gov/programs/racetothetop-district/awards.html>
- Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. Brooklyn, NY: The New Teacher Project.

Appendix A

Survey Protocol

Verification Questions

1. Are you currently employed by the Southwest School District (SSD)?
 - a. Yes
 - b. No
2. How many years have you taught in SSD?
 - c. This is my first year teaching
 - d. 1-3
 - e. 4-5
 - f. 6-10
 - g. 11-15
 - h. 16-20
 - i. 21+
3. How many years have you taught in total?
 - j. This is my first year teaching
 - k. 1-3
 - l. 4-5
 - m. 6-10
 - n. 11-15
 - o. 16-20
 - p. 21+
4. How many years have you received individual Education Value-Added Assessment System (EVAAS) scores? (not school/campus-wide scores)
 - q. 0
 - r. 1
 - s. 2
 - t. 3
 - u. 4
 - v. 5
5. From what type of institution did you receive your teaching certification?
 - w. Public university – in state
 - x. Public university – out of state
 - y. Private university – in state
 - z. Private university – out of state
 - aa. SSD certification program
 - bb. Alternative certification program
 - cc. Teach for America
 - dd. State Teaching Fellows
 - ee. Other – please specify
6. Including this year, what grade levels have you taught in SSD? (select all that apply)
 - ff. Pre-K
 - gg. Kindergarten
 - hh. 1
 - ii. 2
 - jj. 3
 - kk. 4
 - ll. 5
 - mm. 6
 - nn. 7
 - oo. 8
 - pp. 9
 - qq. 10
 - rr. 11
 - ss. 12
 - tt. Multi-grade

7. Including this year, what subject areas have you taught in SSD? (select all that apply)
- uu. Mathematics
 - vv. Social Studies/History
 - ww. Reading/English, Language Arts
 - xx. Science
 - yy. Music
 - zz. Art
 - aaa. ESL/Bilingual Education
 - bbb. Special Education
 - ccc. Test Preparation
 - ddd. Physical Education
 - eee. Other – please specify
- 8a. How would you classify the socioeconomic status of students you typically teach in SSD in terms of their needs?
- a. Very high needs
 - b. High needs
 - c. Average
 - d. Low needs
 - e. Very low needs
 - f. Not applicable
- 8b. How would you classify the academic status of the students you typically teach in SSD in terms of their needs?
- a. Very high needs
 - b. High needs
 - c. Average
 - d. Low needs
 - e. Very low needs
 - f. Not applicable
9. Please list your employer organizations (select all that apply):
- a. Congress of Southwest Teachers
 - b. ASPE (Assoc. of State Professional Educators)
 - c. SFT (Southwest Federation of Teachers)
 - d. SCTA (State Classroom Teachers Assoc.)
 - e. STA (State Teachers Assoc.)
 - f. Other – please specify
10. What is your gender?
- a. Male
 - b. Female
11. What is your identified race?
- a. African American/Black
 - b. Asian
 - c. Hispanic/Latino(a)
 - d. Native American/Indian
 - e. Caucasian/White
 - f. Two or more races
 - g. Other
12. In what year were you born?
(Dropdown menu)
- Reliability Questions
13. If you have received more than one year of EVAAS scores, have your scores been consistent over time?
- a. Yes
 - b. No, please explain
 - c. Not applicable
14. If you currently teach or have taught more than one grade level, have your scores been consistent across grade levels?
- a. Yes
 - b. No, please explain
 - c. Not applicable
15. If you currently teach or have taught more than one subject level, have your scores been consistent across subject areas?
- a. Yes
 - b. No, please explain
 - c. Not applicable

16. If you currently teach or have taught different types of students (i.e., varied proportions of ELL, gifted, special ed., low/high income), have your scores been consistent regardless of the students you have taught?
- Yes
 - No, please explain
 - Not applicable
17. If there is anything else you would like to add regarding the questions above, please do so here:

Validity Questions

18. Have you ever been evaluated using EVAAS for a grade level for which you were not the teacher of record?
- Yes, please explain
 - No
 - Not applicable
19. Have you ever been evaluated using EVAAS for a subject area for which you were not the teacher of record?
- Yes, please explain
 - No
 - Not applicable
20. Have you ever been evaluated using EVAAS for a group of students for which you were not the teacher of record?
- Yes, please explain
 - No
 - Not applicable
21. Do your EVAAS scores typically match your principal/supervisor observation/evaluation scores?
- Yes
 - No, please explain
22. Are there any recommendations, awards, student/parent feedback, peer or mentor evaluations that support or contradict your EVAAS scores?
- Yes, please explain
 - No
23. If there is anything else you would like to add regarding the questions above, please do so here:

Formative Uses & Consequences

24. When do you typically receive EVAAS reports for the students you teach?
- Prior to them entering your classroom
 - Summer – after students have left your classroom
 - Fall – when students are in the next grade level
 - You do not typically receive the EVAAS scores for your students
 - You have never received the EVAAS scores for your students
 - Other, please specify
- 25a. If you have received EVAAS reports for your students, have you used their EVAAS reports to inform your instruction?
- Yes, please explain
 - No
- 25b. With regard to EVAAS data usage, which of the following scenarios describe your situation (check all that apply):
- You use EVAAS data to inform your classroom practices
 - You do not typically use EVAAS data to inform practices
 - You use other resources (not EVAAS data) to inform practices
26. Are you aware of EVAAS training sessions that are available to help you understand the model and reports?
- Yes
 - No
27. Are EVAAS trainings mandatory or optional?
- Mandatory
 - Optional
 - You are not aware of such trainings
28. How many in-person sessions have you attended to better understand EVAAS, your EVAAS scores, how to use your EVAAS scores, etc.?
- 1
 - 2
 - 3
 - 4
 - 5 or more

29. How many online trainings have you attended to better understand EVAAS, your EVAAS scores, how to use your EVAAS scores, etc.?

- a. 1
- b. 2
- c. 3
- d. 4
- e. 5 or more

30. Did you find the EVAAS training sessions helpful?

- a. Yes
- b. No
- c. Not applicable

31. Does your principal/supervisor typically reflect on your EVAAS report to improve your instruction?

- a. Yes
- b. No
- c. Not applicable

32. To what extent do you typically reflect on your EVAAS report to improve your instruction? Please explain

33. If there is anything else you would like to add regarding the questions above, please do so here:

Overall Questions

To what extent do you agree with the following statements:

- a. Strongly agree
- b. Agree
- c. Neither agree or disagree
- d. Disagree
- e. Strongly disagree
- f. Not applicable

- 1. EVAAS reports are simple to use
- 2. EVAAS helps create professional goals
- 3. EVAAS helps improve instruction
- 4. EVAAS ensures growth opportunities for students
- 5. EVAAS ensures growth opportunities for very low achieving students
- 6. EVAAS ensures growth opportunities for very high achieving students
- 7. EVAAS helps increase student learning
- 8. EVAAS helps you become a more effective teacher
- 9. EVAAS will validly identify and help to remove ineffective teachers
- 10. EVAAS will identify excellence in teaching or leadership
- 11. EVAAS will provide incentives for good practices
- 12. EVAAS will enhance the school environment
- 13. EVAAS will enhance working conditions
- 14. Overall, the EVAAS is beneficial to me as a teacher
- 15. Overall, the EVAAS is beneficial to my school
- 16. Overall, the EVAAS is beneficial to the district

34. If there is anything else you would like to add regarding the questions above, please do so here:

35. If there is anything else you would like to add overall, please do so here:

Appendix B

Table B1

Likert-Scale Table With Participant Response Per Item

| Statement | Strongly disagree | Disagree | Neutral | Agree | Strongly agree | Total |
|-----------|-------------------|------------|------------|------------|----------------|-------------------|
| 1 | 361 (41.7) | 221 (25.5) | 146 (16.9) | 122 (14.1) | 16 (1.8) | N = 866 (100%) |
| 2 | 346 (39.8) | 165 (19.0) | 162 (18.6) | 173 (19.9) | 24 (2.8) | N = 870 (100%) |
| 3 | 343 (39.7) | 173 (20.0) | 174 (20.1) | 148 (17.1) | 26 (3.0) | N = 864 (100%) |
| 4 | 357 (40.9) | 195 (22.3) | 182 (20.8) | 121 (13.9) | 18 (2.1) | N = 873 (100%) |
| 5 | 362 (41.4) | 188 (21.5) | 181 (20.7) | 122 (13.9) | 22 (2.5) | N = 875 (100%) |
| 6 | 383 (44.0) | 185 (21.3) | 187 (21.5) | 96 (11.0) | 19 (2.2) | N = 870 (100%) |
| 7 | 359 (41.4) | 191 (22.0) | 184 (21.2) | 114 (13.1) | 20 (2.3) | N = 868 (100%) |
| 8 | 383 (44.1) | 180 (20.7) | 156 (18.0) | 123 (14.2) | 27 (3.1) | N = 869 (100%) |
| 9 | 435 (51.2) | 184 (21.7) | 151 (17.8) | 52 (6.1) | 27 (3.2) | N = 849 (100%) |
| 10 | 399 (47.0) | 181 (21.3) | 166 (19.6) | 74 (8.7) | 29 (3.4) | N = 849 (100%) |
| 11 | 365 (42.4) | 165 (19.2) | 174 (20.2) | 117 (13.6) | 39 (4.5) | N = 860 (100%) |
| 12 | 455 (54.0) | 156 (18.5) | 147 (17.5) | 62 (7.4) | 22 (2.6) | N = 842 (100%) |
| 13 | 484 (57.5) | 153 (18.2) | 143 (17.0) | 47 (5.6) | 15 (1.8) | N = 842 (100%) |
| 14 | 415 (48.4) | 148 (17.2) | 139 (16.2) | 124 (14.5) | 32 (3.7) | N = 858 (100%) |
| 15 | 399 (46.7) | 144 (16.8) | 165 (19.3) | 121 (14.2) | 26 (3.0) | N = 855 (100%) |
| 16 | 407 (48.1) | 130 (15.3) | 172 (20.3) | 109 (12.9) | 29 (3.4) | N = 847 (100%) |

Note. Responses are presented as raw numbers, with respective valid proportions of the total in parentheses.

Appendix C

Table C1

Chi-Square Analyses Results for Likert-Scale Items

| Statement/Chi-square result | Strongly disagree | Disagree | Neutral | Agree | Strongly agree | Total |
|--|-------------------|---------------|---------------|---------------|----------------|---------------------------|
| 2: EVAAS helps create professional goals; $\chi^2 = (4, N = 870) = 2.11, p = .715$ | | | | | | |
| Non-union | 102 (11.7) | 52 (6.0) | 45 (5.2) | 58 (6.7) | 9 (1.0) | <i>n</i> = 266 (30.6%) |
| Union | 244 (28.0) | 113 (13.0) | 117 (13.4) | 115 (13.2) | 15 (1.7) | <i>n</i> = 604 (69.5%) |
| Total | 346 (39.8) | 165 (19.0) | 162 (18.6) | 173 (19.9) | 24 (2.8) | <i>N</i> = 870 (100%) |
| 3: EVAAS helps improve instruction; $\chi^2 = (4, N = 864) = 2.35, p = .672$ | | | | | | |
| Non-union | 104 (12.0) | 49 (5.7) | 52 (6.0) | 53 (6.1) | 8 (1.0) | <i>n</i> = 266 (30.8%) |
| Union | 239 (27.7) | 124 (14.4) | 122 (14.1) | 95 (11.0) | 18 (2.1) | <i>n</i> = 598 (69.2%) |
| Total | 346 (39.7) | 173 (20.0) | 174 (20.1) | 148 (17.1) | 26 (3.0) | <i>N</i> = 864 (100%) |
| 4: EVAAS ensures growth opportunities for students; $\chi^2 = (4, N = 873) = 3.11, p = .539$ | | | | | | |
| Non-union | 104 (11.9) | 60 (6.9) | 58 (6.6) | 42 (4.8) | 3 (0.3) | <i>n</i> = 267 (30.6%) |
| Union | 253 (29.0) | 135 (15.5) | 124 (14.2) | 79 (9.0) | 15 (1.7) | <i>n</i> = 606 (69.4%) |
| Total | 357 (40.9) | 195 (22.3) | 182 (20.8) | 121 (13.9) | 18 (2.1) | <i>N</i> = 873 (100%) |
| 5: EVAAS ensures growth opportunities for very low achieving students; $\chi^2 = (4, N = 875) = 7.20, p = .126$ | | | | | | |
| Non-union | 103 (11.8) | 52 (5.9) | 62 (7.1) | 46 (5.3) | 4 (0.5) | <i>n</i> = 267 (30.5%) |
| Union | 259 (29.6) | 136 (15.5) | 119 (13.6) | 76 (8.7) | 18 (2.1) | <i>n</i> = 608 (69.5%) |
| Total | 362 (41.1) | 188 (21.5) | 181 (20.7) | 122 (13.9) | 22 (2.5) | <i>N</i> = 875 (100%) |
| 6: EVAAS ensures growth opportunities for very high achieving students; $\chi^2 = (4, N = 870) = 3.96, p = .412$ | | | | | | |
| Non-union | 122 (14.0) | 48 (5.5) | 61 (7.0) | 33 (3.8) | 4 (0.5) | <i>n</i> = 268 (30.8%) |
| Union | 261 (30.0) | 137 (15.7) | 126 (14.5) | 63 (7.2) | 15 (1.7) | <i>n</i> = 602 (69.2%) |
| Total | 383 (44.0) | 185 (21.3) | 187 (21.5) | 96 (11.0) | 19 (2.2) | <i>N</i> = 870 (100%) |
| 7: EVAAS helps increase student learning; $\chi^2 = (4, N = 868) = 7.16, p = .128$ | | | | | | |
| Non-union | 105 (12.1) | 51 (5.9) | 59 (6.8) | 45 (5.2) | 4 (0.5) | <i>n</i> = 264 (30.4%) |
| Union | 254 (29.3) | 140 (16.1) | 125 (14.4) | 69 (7.9) | 16 (1.8) | <i>n</i> = 604 (69.6%) |
| Total | 359 (41.1) | 191 (22.0) | 184 (21.2) | 114 (13.1) | 20 (2.3) | <i>N</i> = 868 (100%) |
| 8: EVAAS helps you become a more effective teacher; $\chi^2 = (4, N = 869) = 3.70, p = .448$ | | | | | | |
| Non-union | 108 (12.4) | 58 (6.7) | 46 (5.3) | 45 (5.2) | 7 (0.8) | <i>n</i> = 264 (30.4%) |
| Union | 275 (31.6) | 122 (14.0) | 110 (12.7) | 78 (9.0) | 20 (2.3) | <i>n</i> = 605 (69.6%) |
| Total | 383 (44.1) | 180 (20.7) | 156 (18.0) | 123 (14.2) | 27 (3.1) | <i>N</i> = 869 (100%) |

Table C1 (cont.d)
Chi-Square Analyses Results for Likert-Scale Items

| Statement/Chi-square result | SD | D | N | A | SA | Total |
|--|---------------|---------------|---------------|---------------|-------------|---------------------------|
| 9: EVAAS will validly identify and help to remove ineffective teachers; $\chi^2 = (4, N = 849) = 4.10, p = .393$ | | | | | | |
| Non-union | 124 (14.6) | 66 (7.8) | 46 (5.4) | 19 (2.2) | 8 (0.9) | <i>n</i> = 263 (31.0%) |
| Union | 311 (36.6) | 118 (13.9) | 105 (12.4) | 33 (3.9) | 19 (2.2) | <i>n</i> = 586 (69.0%) |
| Total | 435 (51.2) | 184 (21.7) | 151 (17.8) | 52 (6.1) | 27 (3.2) | <i>N</i> = 849 (100%) |
| 10: EVAAS will identify excellence in teaching or leadership; $\chi^2 = (4, N = 849) = 2.58, p = .631$ | | | | | | |
| Non-union | 118 (13.9) | 64 (7.5) | 49 (5.8) | 25 (2.9) | 8 (0.9) | <i>n</i> = 264 (31.1%) |
| Union | 281 (33.1) | 117 (13.8) | 117 (13.8) | 49 (5.8) | 21 (2.5) | <i>n</i> = 585 (68.9%) |
| Total | 399 (47.0) | 181 (21.3) | 166 (19.6) | 74 (8.7) | 29 (3.4) | <i>N</i> = 849 (100%) |
| 11: EVAAS will provide incentives for good practice; $\chi^2 = (4, N = 860) = 3.23, p = .520$ | | | | | | |
| Non-union | 105 (12.2) | 54 (6.3) | 60 (7.0) | 31 (3.6) | 13 (1.5) | <i>n</i> = 263 (30.6%) |
| Union | 260 (30.2) | 111 (12.9) | 114 (13.3) | 86 (10.0) | 26 (3.0) | <i>n</i> = 597 (69.4%) |
| Total | 365 (42.4) | 165 (19.2) | 174 (20.2) | 117 (13.6) | 39 (4.5) | <i>N</i> = 860 (100%) |
| 12: EVAAS will enhance the school environment; $\chi^2 = (4, N = 842) = 0.84, p = .933$ | | | | | | |
| Non-union | 138 (16.4) | 49 (5.8) | 46 (5.5) | 22 (2.6) | 6 (0.7) | <i>n</i> = 261 (31.0%) |
| Union | 317 (37.6) | 107 (12.7) | 101 (12.0) | 40 (4.8) | 16 (1.9) | <i>n</i> = 581 (69.0%) |
| Total | 455 (54.0) | 156 (18.5) | 147 (17.5) | 62 (7.4) | 22 (2.6) | <i>N</i> = 842 (100%) |
| 13: EVAAS will enhance working conditions; $\chi^2 = (4, N = 842) = 2.57, p = .632$ | | | | | | |
| Non-union | 145 (17.2) | 47 (5.6) | 51 (6.1) | 15 (1.8) | 3 (0.4) | <i>n</i> = 261 (31.0%) |
| Union | 339 (40.3) | 106 (12.5) | 92 (10.9) | 32 (3.8) | 12 (1.4) | <i>n</i> = 581 (69.0%) |
| Total | 484 (57.5) | 153 (18.2) | 143 (17.0) | 47 (5.6) | 15 (1.8) | <i>N</i> = 842 (100%) |
| 14: Overall, the EVAAS is beneficial to my school; $\chi^2 = (4, N = 858) = 7.93, p = .094$ | | | | | | |
| Non-union | 123 (14.3) | 37 (4.3) | 52 (6.1) | 44 (5.1) | 7 (0.8) | <i>n</i> = 263 (30.7%) |
| Union | 292 (34.0) | 111 (12.9) | 87 (10.1) | 80 (9.3) | 25 (2.9) | <i>n</i> = 595 (69.3%) |
| Total | 415 (48.4) | 148 (17.2) | 139 (16.2) | 124 (14.5) | 32 (3.7) | <i>N</i> = 858 (100%) |
| 15: Overall, the EVAAS is beneficial to my school; $\chi^2 = (4, N = 855) = 5.162, p = .271$ | | | | | | |
| Non-union | 112 (13.1) | 46 (5.4) | 50 (5.8) | 46 (5.4) | 6 (0.7) | <i>n</i> = 260 (30.4%) |
| Union | 287 (33.6) | 98 (11.5) | 115 (13.5) | 75 (8.8) | 20 (2.3) | <i>n</i> = 595 (69.6%) |
| Total | 399 (46.7) | 144 (16.8) | 165 (19.3) | 121 (14.2) | 26 (3.0) | <i>N</i> = 855 (100%) |
| 16: Overall, the EVAAS is beneficial to the district; $\chi^2 = (4, N = 847) = 5.96, p = .202$ | | | | | | |
| Non-union | 115 (13.6) | 39 (4.6) | 56 (6.6) | 42 (5.0) | 6 (0.7) | <i>n</i> = 258 (30.5%) |
| Union | 292 (34.5) | 91 (10.7) | 116 (13.7) | 67 (7.9) | 23 (2.7) | <i>n</i> = 589 (69.5%) |
| Total | 407 (48.1) | 130 (15.3) | 172 (20.3) | 109 (12.9) | 29 (3.4) | <i>N</i> = 847 (100%) |

Appendix D

Table D1

Chi-Square for All Other Categorical Items

| Statement and Chi-square result | N/A | Yes | No | Total |
|---|---------------|---------------|---------------|----------------|
| 13: If you have received more than one year of EVAAS scores, have your scores been consistent over time? $\chi^2 = (2, N = 874) = 3.589, p = .166$ | | | | |
| Non-union | 33 (3.8) | 101 (11.6) | 134 (15.3) | 268 (30.7%) |
| Union | 66 (7.6) | 270 (30.9) | 270 (30.9) | 606 (69.3%) |
| Total | 99 (11.3) | 371 (42.4) | 404 (46.2) | 874 (100%) |
| 14: If you currently teacher or have taught more than one grade level, have your EVAAS scores been consistent across grade levels? $\chi^2 = (2, N = 873) = 2.818, p = .244$ | | | | |
| Non-union | 107 (12.3) | 82 (9.4) | 78 (8.9) | 267 (30.6%) |
| Union | 207 (23.7) | 205 (23.5) | 194 (22.2) | 606 (69.4%) |
| Total | 314 (36.0) | 287 (32.9) | 272 (31.2) | 873 (100%) |
| 15: If you currently teach or have taught more than one subject area, have your EVAAS scores been consistent across subject areas? $\chi^2 = (2, N = 867) = 4.251, p = .119$ | | | | |
| Non-union | 98 (11.3) | 76 (8.8) | 90 (10.4) | 264 (30.4%) |
| Union | 192 (22.1) | 215 (24.8) | 196 (22.6) | 603 (69.6%) |
| Total | 290 (33.4) | 291 (33.6) | 286 (33.0) | 867 (100%) |
| 16: If you currently teach or have taught different types of students (i.e., varied proportions of ELL, gifted, special ed, low/high income), have your EVAAS scores been consistent regardless of the students you taught? $\chi^2 = (2, N = 877) = 1.448, p = .485$ | | | | |
| Non-union | 55 (6.3) | 95 (10.8) | 117 (13.3) | 267 (30.4%) |
| Union | 112 (12.8) | 242 (27.6) | 256 (29.2) | 610 (69.6%) |
| Total | 167 (19.0) | 337 (38.4) | 373 (42.5) | 877 (100%) |
| 18: Have you ever been evaluated using the EVAAS for a grade level for which you were not the teacher of record? $\chi^2 = (2, N = 875) = 0.840, p = .657$ | | | | |
| Non-union | 10 (1.1) | 21 (2.4) | 237 (27.1) | 268 (30.6%) |
| Union | 24 (2.7) | 59 (6.7) | 524 (59.9) | 607 (69.4%) |
| Total | 34 (3.9) | 80 (9.1) | 761 (87.0) | 875 (100%) |
| 19: Have you ever been evaluated using EVAAS for a subject area for which you were not the teacher of record? $\chi^2 = (2, N = 874) = 0.218, p = .897$ | | | | |
| Non-union | 12 (1.4) | 24 (2.7) | 233 (26.7) | 269 (30.8%) |
| Union | 26 (3.0) | 60 (6.9) | 519 (59.4) | 605 (69.2%) |
| Total | 38 (4.3) | 84 (9.6) | 752 (86.0) | 877 (100%) |

Table D1 (cont.d)

Chi-Square for All Other Categorical Items

| Statement and Chi-square result | N/A | Yes | No | Total |
|---|---------------|---------------|---------------|----------------|
| 20: Have you ever been evaluated using EVAAS for a group of students for which you were not the teacher of record? $\chi^2 = (2, N = 871) = 2.067, p = .356$ | | | | |
| Non-union | 9 (1.0) | 40 (4.6) | 218 (25.0) | 267 (30.7%) |
| Union | 25 (2.9) | 112 (12.9) | 467 (53.6) | 604 (69.3%) |
| Total | 34 (3.9) | 152 (17.5) | 685 (78.6) | 871 (100%) |
| 21: Do your EVAAS scores typically match your principal/ supervisor observation/ evaluation scores? $\chi^2 = (1, N = 863) = 3.007, p = .083$ | | | | |
| Non-union | | 124 (14.4) | 141 (16.3) | 265 (30.7%) |
| Union | | 242 (28.0) | 356 (41.3) | 598 (69.3%) |
| Total | | 366 (42.4) | 497 (57.6) | 863 (100%) |
| 22: Are there any recommendations, awards, student/ parent feedback, peer mentor evaluations that contradict your EVAAS scores? $\chi^2 = (1, N = 843) = 0.028, p = .866$ | | | | |
| Non-union | | 113 (13.4) | 144 (17.1) | 257 (30.5%) |
| Union | | 254 (30.1) | 332 (39.4) | 586 (69.5%) |
| Total | | 367 (43.5) | 476 (56.5) | 843 (100%) |
| 25: If you have received EVAAS reports for your students, have you used their EVAAS reports to inform your instruction? $\chi^2 = (1, N = 815) = 0.027, p = .868$ | | | | |
| Non-union | | 98 (12.0) | 143 (17.5) | 241 (29.6%) |
| Union | | 237 (29.1) | 337 (41.3) | 574 (70.4%) |
| Total | | 335 (41.1) | 480 (58.9) | 815 (100%) |
| 26: Are you aware of EVAAS training sessions that are available to help you understand the model and reports? $\chi^2 = (1, N = 870) = 1.373, p = .241$ | | | | |
| Non-union | | 174 (20.0) | 91 (10.5) | 265 (30.5%) |
| Union | | 372 (42.8) | 233 (26.8) | 605 (69.5%) |
| Total | | 546 (62.8) | 324 (37.2) | 870 (100%) |
| 27: Are EVAAS trainings mandatory or optional? $\chi^2 = (2, N = 863) = 1.360, p = .507$ | | | | |
| | Mandatory | Optional | Not aware | |
| Non-union | 68 (7.3) | 128 (14.8) | 72 (8.3) | 263 (30.5%) |
| Union | 136 (15.8) | 276 (32.0) | 188 (21.8) | 600 (69.5%) |
| Total | 199 (23.1) | 404 (46.8) | 260 (30.1) | 863 (100%) |
| 30: Did you find the EVAAS trainings helpful? $\chi^2 = (2, N = 864) = 2.259, p = .323$ | | | | |
| Non-union | 64 (7.4) | 75 (8.7) | 122 (14.1) | 261 (30.2%) |
| Union | 178 (20.6) | 161 (18.6) | 264 (30.6) | 603 (69.8%) |
| Total | 242 (28.0) | 236 (27.3) | 386 (44.7) | 864 (100%) |

Table D1 (cont.d)

Chi-Square for All Other Categorical Items

| Statement and Chi-square result | N/A | Yes | No | Total |
|--|-------------|---------------|---------------|----------------|
| 31: Does your principal/ supervisor typically discuss your EVAAS results with you? $\chi^2 = (2, N = 868) = 0.212, p = .899$ | | | | |
| Non-union | 16 (1.8) | 127 (14.6) | 124 (14.3) | 267 (30.8%) |
| Union | 33 (3.8) | 295 (34.0) | 273 (31.5) | 601 (69.2%) |
| Total | 49 (5.6) | 422 (48.6) | 397 (45.7) | 868 (100%) |

About the Author

Clarín Collins, Ph.D.

Private foundation, Phoenix, Arizona

clarin.collins@asu.edu

Dr. Collins is a graduate from the Educational Policy and Evaluation program at the Mary Lou Fulton Teachers College at Arizona State University. Her research interests include national and local policy implementation at the school and classroom level, teacher influences on policymaking and implementation, and education evaluation and accountability systems. Clarín currently works as a research and evaluation officer at a private foundation in Phoenix.

education policy analysis archives

Volume 22 Number 98 October 27th, 2014 ISSN 1068-2341



Readers are free to copy, display, and distribute this article, as long as the work is attributed to the author(s) and **Education Policy Analysis Archives**, it is distributed for non-commercial purposes only, and no alteration or transformation is made in the work. More details of this Creative Commons license are available at <http://creativecommons.org/licenses/by-nc-sa/3.0/>. All other uses must be approved by the author(s) or **EPAA**. **EPAA** is published by the Mary Lou Fulton Institute and Graduate School of Education at Arizona State University. Articles are indexed in CIRC (Clasificación Integrada de Revistas Científicas, Spain), DIALNET (Spain), [Directory of Open Access Journals](#), EBSCO Education Research Complete, ERIC, Education Full Text (H.W. Wilson), QUALIS A2 (Brazil), SCImago Journal Rank; SCOPUS, SOCOLAR (China).

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/EPAAAPE> and **Twitter feed** @epaa_aape.

education policy analysis archives
editorial board

Editor **Gustavo E. Fischman** (Arizona State University)

Associate Editors: **Audrey Amrein-Beardsley** (Arizona State University) **Rick Mintrop**, (University of California, Berkeley) **Jeanne M. Powers** (Arizona State University)

Jessica Allen University of Colorado, Boulder

Gary Anderson New York University

Michael W. Apple University of Wisconsin, Madison

Angela Arzubiaga Arizona State University

David C. Berliner Arizona State University

Robert Bickel Marshall University

Henry Braun Boston College

Eric Camburn University of Wisconsin, Madison

Wendy C. Chi University of Colorado, Boulder

Casey Cobb University of Connecticut

Arnold Danzig Arizona State University

Antonia Darder University of Illinois, Urbana-Champaign

Linda Darling-Hammond Stanford University

Chad d'Entremont Strategies for Children

John Diamond Harvard University

Tara Donahue Learning Point Associates

Sherman Dorn University of South Florida

Christopher Joseph Frey Bowling Green State University

Melissa Lynn Freeman Adams State College

Amy Garrett Dikkers University of Minnesota

Gene V Glass Arizona State University

Ronald Glass University of California, Santa Cruz

Harvey Goldstein Bristol University

Jacob P. K. Gross Indiana University

Eric M. Haas WestEd

Kimberly Joy Howard* University of Southern California

Aimee Howley Ohio University

Craig Howley Ohio University

Steve Klees University of Maryland

Jackyung Lee SUNY Buffalo

Christopher Lubienski University of Illinois, Urbana-Champaign

Sarah Lubienski University of Illinois, Urbana-Champaign

Samuel R. Lucas University of California, Berkeley

Maria Martinez-Coslo University of Texas, Arlington

William Mathis University of Colorado, Boulder

Tristan McCowan Institute of Education, London

Heinrich Mintrop University of California, Berkeley

Michele S. Moses University of Colorado, Boulder

Julianne Moss University of Melbourne

Sharon Nichols University of Texas, San Antonio

Noga O'Connor University of Iowa

João Paraskveva University of Massachusetts, Dartmouth

Laurence Parker University of Illinois, Urbana-Champaign

Susan L. Robertson Bristol University

John Rogers University of California, Los Angeles

A. G. Rud Purdue University

Felicia C. Sanders The Pennsylvania State University

Janelle Scott University of California, Berkeley

Kimberly Scott Arizona State University

Dorothy Shipps Baruch College/CUNY

Maria Teresa Tatto Michigan State University

Larisa Warhol University of Connecticut

Cally Waite Social Science Research Council

John Weathers University of Colorado, Colorado Springs

Kevin Welner University of Colorado, Boulder

Ed Wiley University of Colorado, Boulder

Terrence G. Wiley Arizona State University

John Willinsky Stanford University

Kyo Yamashiro University of California, Los Angeles

archivos analíticos de políticas educativas
consejo editorial

Editor: **Gustavo E. Fischman** (Arizona State University)

Editores. Asociados **Jason Beech** (Universidad de San Andrés) **Alejandro Canales** (UNAM) y **Jesús Romero Morante** (Universidad de Cantabria)

Armando Alcántara Santuario Instituto de Investigaciones sobre la Universidad y la Educación, UNAM México

Claudio Almonacid Universidad Metropolitana de Ciencias de la Educación, Chile

Pilar Arnaiz Sánchez Universidad de Murcia, España

Xavier Besalú Costa Universitat de Girona, España

Jose Joaquín Brunner Universidad Diego Portales, Chile

Damián Canales Sánchez Instituto Nacional para la Evaluación de la Educación, México

María Caridad García Universidad Católica del Norte, Chile

Raimundo Cuesta Fernández IES Fray Luis de León, España

Marco Antonio Delgado Fuentes Universidad Iberoamericana, México

Inés Dussel DIE, Mexico

Rafael Feito Alonso Universidad Complutense de Madrid, España

Pedro Flores Crespo Universidad Iberoamericana, México

Verónica García Martínez Universidad Juárez Autónoma de Tabasco, México

Francisco F. García Pérez Universidad de Sevilla, España

Edna Luna Serrano Universidad Autónoma de Baja California, México

Alma Maldonado Departamento de Investigaciones Educativas, Centro de Investigación y de Estudios Avanzados, México

Alejandro Márquez Jiménez Instituto de Investigaciones sobre la Universidad y la Educación, UNAM México

José Felipe Martínez Fernández University of California Los Angeles, USA

Fanni Muñoz Pontificia Universidad Católica de Perú

Imanol Ordorika Instituto de Investigaciones Economicas – UNAM, México

Maria Cristina Parra Sandoval Universidad de Zulia, Venezuela

Miguel A. Pereyra Universidad de Granada, España

Monica Pini Universidad Nacional de San Martín, Argentina

Paula Razquin Universidad de San Andrés

Ignacio Rivas Flores Universidad de Málaga, España

Daniel Schugurensky Arizona State University

Orlando Pulido Chaves Universidad Pedagógica Nacional, Colombia

José Gregorio Rodríguez Universidad Nacional de Colombia

Miriam Rodríguez Vargas Universidad Autónoma de Tamaulipas, México

Mario Rueda Beltrán Instituto de Investigaciones sobre la Universidad y la Educación, UNAM México

José Luis San Fabián Maroto Universidad de Oviedo, España

Yengny Marisol Silva Laya Universidad Iberoamericana, México

Aida Terrón Bañuelos Universidad de Oviedo, España

Jurjo Torres Santomé Universidad de la Coruña, España

Antoni Verger Planells University of Amsterdam, Holanda

Mario Yapu Universidad Para la Investigación Estratégica, Bolivia

arquivos analíticos de políticas educativas
conselho editorial

Editor: **Gustavo E. Fischman** (Arizona State University)
Editores Associados: **Rosa Maria Bueno Fisher** e **Luis A. Gandin**
(Universidade Federal do Rio Grande do Sul)

Dalila Andrade de Oliveira Universidade Federal de Minas Gerais, Brasil
Paulo Carrano Universidade Federal Fluminense, Brasil
Alicia Maria Catalano de Bonamino Pontifícia Universidade Católica-Rio, Brasil
Fabiana de Amorim Marcello Universidade Luterana do Brasil, Canoas, Brasil
Alexandre Fernandez Vaz Universidade Federal de Santa Catarina, Brasil
Gaudêncio Frigotto Universidade do Estado do Rio de Janeiro, Brasil
Alfredo M Gomes Universidade Federal de Pernambuco, Brasil
Petronilha Beatriz Gonçalves e Silva Universidade Federal de São Carlos, Brasil
Nadja Herman Pontifícia Universidade Católica –Rio Grande do Sul, Brasil
José Machado Pais Instituto de Ciências Sociais da Universidade de Lisboa, Portugal
Wenceslao Machado de Oliveira Jr. Universidade Estadual de Campinas, Brasil

Jefferson Mainardes Universidade Estadual de Ponta Grossa, Brasil
Luciano Mendes de Faria Filho Universidade Federal de Minas Gerais, Brasil
Lia Raquel Moreira Oliveira Universidade do Minho, Portugal
Belmira Oliveira Bueno Universidade de São Paulo, Brasil
Antônio Teodoro Universidade Lusófona, Portugal
Pia L. Wong California State University Sacramento, U.S.A
Sandra Regina Sales Universidade Federal Rural do Rio de Janeiro, Brasil
Elba Siqueira Sá Barreto Fundação Carlos Chagas, Brasil
Manuela Terrasêca Universidade do Porto, Portugal
Robert Verhine Universidade Federal da Bahia, Brasil
Antônio A. S. Zuin Universidade Federal de São Carlos, Brasil