



Education Policy Analysis  
Archives/Archivos Analíticos de Políticas  
Educativas

ISSN: 1068-2341

epaa@alperin.ca

Arizona State University  
Estados Unidos

Kappler Hewitt, Kimberly  
Educator Evaluation Policy that Incorporates EVAAS Value-Added Measures:  
Undermined Intentions and Exacerbated Inequities  
Education Policy Analysis Archives/Archivos Analíticos de Políticas Educativas, vol. 23,  
2015, pp. 1-46  
Arizona State University  
Arizona, Estados Unidos

Available in: <http://www.redalyc.org/articulo.oa?id=275041389046>

- How to cite
- Complete issue
- More information about this article
- Journal's homepage in redalyc.org

redalyc.org

Scientific Information System

Network of Scientific Journals from Latin America, the Caribbean, Spain and Portugal

Non-profit academic project, developed under the open access initiative

## Educator Evaluation Policy that Incorporates EVAAS Value-Added Measures: Undermined Intentions and Exacerbated Inequities

*Kimberly Kappler Hewitt*

University of North Carolina Greensboro  
United States

**Citation:** Hewitt, K. K. (2015). Educator evaluation policy that incorporates EVAAS value-added measures: Undermined intentions and exacerbated inequities. *Education Policy Analysis Archives*, 23(76). <http://dx.doi.org/10.14507/epaa.v23.1968>

**Abstract:** In the United States, policies in forty states and D.C. incorporate student growth measures – estimates of student progress attributed to educators – into educator evaluation. The federal government positions such policies as levers for ensuring that more students are taught by effective teachers and that effective educators are more equitably distributed amongst schools. Because these policies are new, little is known about how educators respond to them. Mixed methods survey data from a large, diverse district in North Carolina, a state that incorporates value-added data into teacher evaluations, indicate that substantive, unintended effects may undermine the purposes for which these policies were developed. Results indicate that educators evaluated by value-added are generally opposed to its use. Those who have previously been evaluated by value-added have significantly more negative perceptions about the fairness and accuracy of value-added, are more opposed to its use in educator evaluation, and are more likely to perceive that it will not result in more equitable distribution of good educators across schools and that educators will avoid working with certain students because of value-added. Respondents perceived effects of the use of value-added for teacher accountability that fall within five themes: 1) Educators increasingly game the system and teach to the test, 2) Teachers increasingly leave the field, 3) Some educators seek to avoid working with certain students and at certain schools, 4) Educators feel an increase in stress,

pressure, and anxiety, 5) Educator collaboration is decreasing, and competition is increasing. Based on findings, the author recommends five mid-course policy corrections.

**Keywords:** United States of America; value-added; student growth measures; teacher evaluation; teacher accountability; educational policy; survey research; mixed methods

### **Política de Evaluación del Educadores que Incorporan Medidas EVAAS de Valor Agregado: Intenciones Socavadas y Desigualdades Exacerbadas**

**Resumen:** En los Estados Unidos, políticas en cuarenta estados y DC incorporan medidas de estimación de crecimiento del estudiante – progreso del estudiante atribuidos a los educadores – en la evaluación-educador. El gobierno federal clasifica esas políticas como instrumentos para asegurar que más estudiantes sean enseñados por maestros eficaces y que educadores más efectivos se distribuyan de manera más equitativa entre las escuelas. Debido a que estas políticas son nuevas, se sabe poco sobre cómo los educadores responden a ellas. Con datos de una encuesta de métodos mixtos de un distrito grande en Carolina del Norte, un estado que incorpora datos de valor agregado en la evaluación de maestros, indican que, efectos no intencionales sustantivos pueden socavar los fines para los cuales se desarrollaron estas políticas. Los resultados indican que los educadores evaluados por modelos de valor añadido en general se oponen a su uso. Los que han sido previamente evaluados por modelos de valor agregado tienen percepciones significativamente más negativas sobre la equidad y la exactitud de valor añadido, son más opuestos a su uso en la evaluación docente, y son más propensos a percibir que no dará lugar a una distribución más equitativa de buenos educadores a través de las escuelas y que los educadores evitarán trabajar con ciertos estudiantes debido al modelo de valor agregado. Los encuestados perciben los efectos de la utilización de valor agregado para la rendición de cuentas dentro de cinco temas: 1) Los educadores juegan cada vez más el sistema y enseñan para aprobar los exámenes. 2) Cada vez mas profesores dejan la profesión. 3) Algunos educadores tratan de evitar trabajar con ciertos estudiantes y en ciertas escuelas. 4) Los educadores sienten un aumento del estrés, presión, y ansiedad. 5) la colaboración entre educadores está disminuyendo, y la competencia es cada vez mayor. Con base en los hallazgos, la autora recomienda cinco correcciones de políticas.

**Palabras clave:** Estados Unidos de América; valor añadido; medidas de crecimiento de estudiantes; evaluación docente; responsabilidad docente; política educativa; encuestas; métodos mixtos

### **Política de Avaliação do Educador que Incorpora Medidas EVAAS de Valor Agregado: Intenções Debilitadas e Desigualdades Exacerbadas**

**Resumo:** Nos Estados Unidos, as políticas em quarenta estados e DC incluem medidas de crescimento do aluno -estimativa de progresso atribuídas a educadores- em a avaliação dos educadores. O governo federal classifica essas políticas como instrumentos para garantir que mais estudantes sejam ensinados por professores efetivos e que os educadores mais eficazes sejam distribuídos de forma mais equitativa entre as escolas. Porque estas políticas são novas, pouco se sabe sobre como os educadores respondem a elas. Usando dados de uma pesquisa de métodos mistos de um grande distrito na Carolina do Norte, um estado que incorpora dados de valor agregado na avaliação de professores, indicam que efeitos involuntários substanciais podem minar os fins para os quais estas políticas foram desenvolvidas. Os resultados indicam que os educadores avaliados por modelos de valor agregado geralmente se opõem a sua utilização. Aqueles que foram previamente avaliados por modelos de valor agregado têm percepções significativamente mais negativas sobre a equidade e a precisão do modelo de valor adicionado, se opõem ao seu uso na avaliação de professores, e são mais propensos a perceber que não levará a uma distribuição de bons professores mais equitativa entre as escolas e os educadores evitaram trabalhar com alguns alunos

devido ao modelo de valor agregado. Os entrevistados percebem os efeitos da utilização de um valor agregado para a prestação de contas dentro de cinco temas: 1) Os educadores cada vez mais procuram manipular o sistema para passar nos exames. 2) Cada vez mais professores deixaram a profissão. 3) Alguns educadores tentaram evitar trabalhar com alguns alunos e em algumas escolas. 4) Os educadores sentem o aumento do estresse, pressão e ansiedade. 5) a colaboração entre professores está em declínio, e a concorrência está aumentando. Com base nas conclusões, a autora recomenda cinco correções dessas políticas.

**Palavras-chave:** Estados Unidos da América; valor adicionado; medidas de crescimento estudante; avaliação de professores; responsabilidade de professores; política educacional; questionários; métodos mistos

## Purpose

One of the US Department of Education's (DoE) FY2014-2015 priority goals is to ensure that "more students have effective teachers and leaders" (US Department of Education, nd, p. 2) and that effective teachers and leaders are more equitably distributed across schools. The DoE is leveraging "teacher and principal evaluation and support systems that consider multiple measures of effectiveness, with student growth as a significant factor" (p. 2) as a policy mechanism to support this goal.

Currently, forty states and the District of Columbia require objective measures of student learning to be included in educator evaluations – a sea change from just five years ago (Doherty & Jacobs/National Council on Teacher Quality, 2013). These changes are, in part, predicated upon the recognition that teachers are the most crucial school-related factor in student learning (Rivkin, Hanushek, & Kain, 2005; Rockoff, 2004) and that educator effectiveness varies considerably across classrooms (Chetty, Friedman, & Rockoff, 2013) and has a host of important, long-term effects on students, including life-time earnings, matriculation to college, and likelihood of having a child as a teenager (Chetty, Friedman, & Rockoff, 2014).

Amongst the most common student growth measures are value-added models (VAM) – statistical models that measure student progress or achievement test-score change over time (Ehlert, Koedel, Parsons, & Podgursky, 2014). Harris and Herrington (2015) argue that the "use of teacher value-added measures could have a greater influence on classroom instruction than perhaps any single reform in decades – for good and for ill" (p. 71). Ultimately, the effects of using VAM for "high-stakes purposes will depend on the way in which teachers and prospective teachers react, their "behavior responses" (Goldhaber, 2015, p. 88). Yet little is known about how educators perceive and respond to the use of VAM for educator evaluation (Corcoran & Goldhaber, 2013; Harris, 2011; Jiang, Sporte, & Luppescu, 2015), and whether their behavioral responses will lead to increased teacher effectiveness and the more equitable distribution of teachers and leaders. Jiang et al. (2015) argue that "studying teacher perceptions will provide insight to both researchers and practitioners on the successes and challenges of these new evaluation systems" (p. 106).

There are two main purposes of this study: 1) To examine educators' perceptions of the use of the Education Value Added Assessment System (EVAAS) – a type of value-added model – for educator evaluation, in particular what effects educators predict these systems will have on teaching and learning and what, if any, consequences of implementing these systems they have observed in their own schools; and 2) To determine how perceptions vary by educator familiarity and experience with the use of value-added for educator evaluation. Findings can inform the DoE's initiative for increased teacher effectiveness and more equitable distribution of teachers and leaders. Findings can also be used as leading indicators about un/intended and un/anticipated impacts of new-generation teacher evaluation systems and can inform mid-course policy corrections.

## Conceptual Framework and Examination of the Literature

Teacher accountability is contested terrain and has received much attention from researchers, policymakers, practitioners, and the mainstream media. Drawing upon Hewitt (2013), this study utilizes a framework that includes five broad areas of consideration to the use of VAM for educator evaluation: technical and validity considerations; test considerations; policy considerations; considerations regarding practice; and equity and social justice considerations (see Figure 1). Although this study focuses on two elements of the framework – considerations regarding practice and equity and social justice considerations – the elements interact in important ways, as described later in this section. As such, this section attends to all five elements of the framework. While technical and validity considerations have received the most attention by scholars, increasingly, empirical and simulation studies are speaking to all five elements. Scholarship that addresses these considerations can maximize the benefits of value-added in teacher accountability, increase its credibility, and reduce unintended effects.

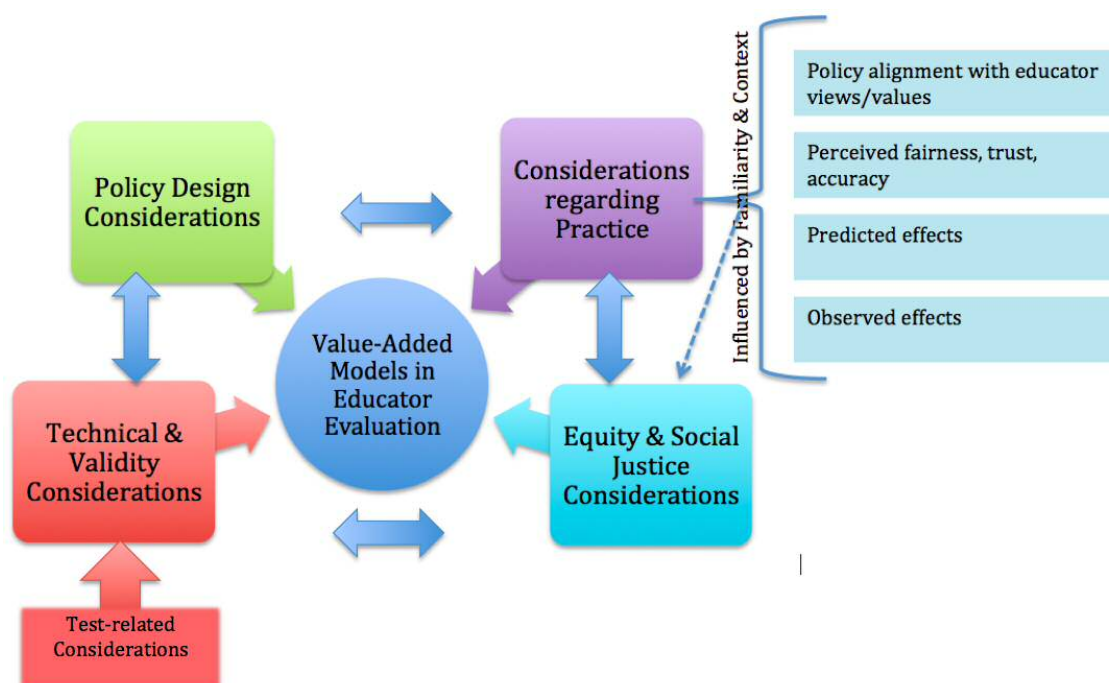


Figure 1. Considerations for the use of value-added in educator evaluation systems.

Note: Adapted from “The Use of Value-Added for Accountability and to Inform Leadership” by Hewitt (2013) in K. K. Hewitt, C. Childers-McKee, E. M. Hodge, & R. C. Schuhler (Eds.), *Postcards from the schoolhouse: Practitioner scholars examine contemporary issues in instructional leadership* (pp. 198-223). Ypsilanti, MI: NCEA Press.

### Technical and Validity Considerations

To date, the technical aspects of VAM have gotten the lion’s share of researcher attention (Johnson, 2015), and experts have hotly debated the validity, reliability, and appropriateness of their use. While some experts support the use of VAM for educator evaluation (e.g. Chetty et al., 2013, 2014; Goldhaber in Corcoran & Goldhaber, 2013; Hanushek & Rivkin, 2010), others have called into question its use (e.g., American Statistical Association, 2014; Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012; Haertel, 2013).

An increasingly prodigious body of scholarship on the technical and validity elements of VAM includes attention to model selection, since different VAMs tend to yield different estimates

of teacher effectiveness (Darling-Hammond et al., 2012; Newton, Darling-Hammond, Haertel, & Thomas, 2010; Timmermans, Doolaard, & de Wolf, 2011), and scholars have compared various models to one another (e.g., Sanders, 2006) and argued for which is the most appropriate model to use (e.g., Ehlert, Koedel, Parsons, & Podgursky, 2014). Another technical consideration is spillage, the influence of other content area educators on a teacher's effectiveness rating in a tested area, which can contaminate value-added estimates of teacher effectiveness (Corcoran, 2010; Koedel, 2009; Yuan, 2015). Another area of debate for scholars is the degree to which issues of persistence and decay are important and how best to account for them statistically. Persistence refers to a teacher's influence on student learning beyond the period of time she is assigned a student, and decay refers to the declining influence of a teacher on former students over time. Scholars argue that persistence is nontrivial (e.g., Konstantopoulos & Chung, 2011) and does decay over time (Briggs & Weeks, 2011; McCaffrey et al., 2004; Mariano, McCaffrey, & Lockwood, 2010), yet there is no definitive answer as to how best to account for persistence and decay statistically in VAM.

An additional technical consideration is whether and how to account statistically for non-teacher influences in value-added measures. For example, there is some evidence that classroom composition (Hill, Kapitula, & Umland, 2011) and school characteristics (McCaffrey et al., 2004) can influence value-added scores, including factors such as strong principal leadership and having more effective colleagues (Corcoran, 2010). Another thorny issue is sorting bias. Students and teachers are nonrandomly assigned to schools and classrooms, and this sorting bias can distort value-added measures (Braun, 2005). Rothstein (2010) dramatically illustrated sorting bias in a study using North Carolina data when he found that a student's fifth grade teacher was a better predictor of the student's fourth grade growth than was the student's fourth grade teacher. In research using a quasi-experimental design, Chetty et al. (2014) subsequently concluded that value-added estimates of teacher effectiveness are unbiased by student sorting. Shortly thereafter, Rothstein (2014) replicated their study with a different sample and found that teacher switching is associated with differences in student preparation, which resulted in moderate sorting bias. Koedel and Betts (2011) found that while some VAMs are markedly biased by nonrandom sorting, a value-added model that incorporates teacher data from multiple years can largely resolve sorting bias. This debate exemplifies the contested terrain of VAM.

Another thorny issue for VAMs is instability of value-added estimates. Relationships between teachers' year-to-year value-added estimates are modest, and teacher's value-added scores tend to be unstable from year-to-year (Braun, 2015; Corcoran, 2010; Goldhaber & Hansen, 2008; Morgan, Hodge, Trepinski, & Anderson, 2014). Additionally, a teacher's value-added estimates tend to be unstable from content area to content area and from one class period to another (e.g., Darling-Hammond et al., 2012). Value-added estimates for a teacher also vary across different tests within the same content area (e.g., Darling-Hammond et al., 2012; Papay, 2011).

### **Test Considerations**

Some scholars argue that grade level standardized tests used to calculate value-added measures do not have sufficient stretch – range of difficulty of items – to accurately identify students' performance (Amrein-Beardsley, 2008; Carey & Manwaring, 2011; Darling-Hammond, 2015). Some scholars also argue that tests need to be on a vertical scale and measure the same skills/content over time so that construct-shift does not distort value-added measures (Martineau, 2006; Schmidt, Houang, & McKnight, 2005). Polikoff and Porter (2014) conjecture that state tests are “not particularly able to detect differences in the content or quality of classroom instruction . . . [and] may not be up to the task of differentiating effective from ineffective (or aligned from misaligned) teaching” (p. 16). In contrast, Wright, White, Sanders, and Rivers (2010) argue that

almost all commercial and state accountability tests meet specifications for use with value-added models, thus suggesting that test issues are not a substantive concern.

### **Policy Design Considerations**

Not only does VAM model selection and test selection matter; it also matters how VAM is incorporated into teacher accountability policy. For example, Winters and Cowen (2015) demonstrate through simulation that the policy decision of whether to base a teacher's dismissal on two consecutive years of low value-added scores versus a two-year average has big implications for how many teachers are identified for dismissal, with the two-year average approach identifying a larger set of teachers for dismissal, and the two-year consecutive approach doing a better job of identifying teachers who tend to be less effective. Even if specifications are set such that the two approaches yield the same number of teachers for dismissal, the two approaches often identify different teachers for dismissal. Additionally, Winters and Cowen (year) found that unless the cutoff percentile for dismissal is set quite high (e.g., 27<sup>th</sup> percentile for the two-year consecutive approach), then "policymakers should limit their expectations for the effectiveness of such a policy on overall student achievement because it will tend to remove few teachers and many ineffective teachers will remain unidentified" (p. 336). Additional considerations for policy design include the nature of the teacher labor market and the role of natural attrition of teachers. Simulation modeling by Winters and Cowen (2013) demonstrates that effects of dismissal policies based on value-added could be substantially influenced by the size and nature of the labor market, such that in limited labor markets, the potential for positive effects of dismissal policies could be markedly reduced. Also, Winters and Cowen (2013) found that when natural attrition of teachers is incorporated into simulation modeling, the potential for positive effects of dismissal policies are reduced, due to the fact that less effective teachers are more likely to leave the profession. Cowen and Winters (2015) conclude that "the quality and number of teachers dismissed under value-added policies depends heavily on policy design" (p. 331).

In a study that speaks to the effects of policy design on teacher quality, Dee and Wyckoff (2013) examined the Washington, D.C. IMPACT program, which incorporates multiple-measure teacher evaluations, including the use of value-added data, with high contrast incentives, including immediate dismissal for an ineffective rating in contrast to large one-time bonuses (up to \$25,000) for a highly effective rating and base pay increases of up to \$27,000 for teachers with two consecutive years of highly effective ratings. It is important to note that only about 17% of D.C. teachers in the study had individual value-added data as part of their evaluations (math and reading teachers in grades 4-8), and for those who did, the data accounted for 50% of their evaluation. Dee and Wyckoff concluded that IMPACT improved the effectiveness of D.C. teachers in two ways: the voluntary attrition of low-performing teachers increased, and the performance of remaining teachers improved. Additionally, teachers entering the district outperformed teachers who had left it. This study suggests that evaluation designs that pair multi-measure evaluations with high contrast incentives may be a powerful way to increase teacher effectiveness.

### **Considerations for Equity**

Extant scholarship on equity considerations for teacher accountability is limited; what literature exists suggests that there could be important equity considerations for teachers and students. Transient students, who often have missing test score data (Corcoran, 2010), could be marginalized if teachers invest in them less because their data will not contribute towards teachers' effectiveness scores. Moreover, some literature (e.g., Baker et al., 2010; Darling-Hammond et al., 2012; Jackson, 2012; Kupermintz, 2003; McCaffrey & Buzick, 2014; Newton et al., 2010) suggests that VAM estimates of teacher effectiveness can be biased (Braun, 2015) against educators whose

teaching assignments include substantial numbers of students with disabilities, impoverished students, English Language Learners, and gifted students. However, Ballou, Sanders, and Wright (2004) found that student background factors and characteristics have negligible influence on value-added estimates, which suggests that concerns about equity might be overblown. If such bias exists, VAM would be inequitable for educators who teach these students. Additionally, perceptions of such bias, found in research by Collins (2014) on teachers in a large Southwestern urban district, could be a perverse incentive for teachers to avoid working with these students. This could create further inequity for marginalized students by relegating them to novice teachers who tend to be less effective than their more experienced counterparts (e.g., Hanushek & Kane, Rockoff, & Staiger, 2005). Interestingly, the simulation modeling of Winters and Cowen (2013) suggests that using value-added to dismiss low performing teachers would have minimal effects (either ameliorative or exacerbating) on well-documented existing inequities (e.g., Kalogrides & Loeb, 2013; Lankford, Loeb, & Wyckoff, 2002) in the distribution of quality teachers.

### **Considerations for Practice**

Considerations for practice involve educators' responses to teacher accountability policies. Scholars point out that the effects of using value-added for teacher accountability will largely be determined by how educators react to such policies – their behavioral responses (Goldhaber, 2015; Harris, 2011), and Harris and Herrington (2015) point out that “policies rarely affect practice as intended” (p. 72). There is limited research on educators' perceptions of and responses to the use of VAMs for high stakes purposes. In a three-year study of pay-for-performance in Nashville based on a value-added model, Springer et al. (2010) found that two-thirds of teachers involved in the study perceived that the value-added model could not accurately discriminate between effective and ineffective teaching, reflecting perceived validity issues. Amrein-Beardsley and Collins (2012), in a study of the use of SAS EVAAS (a type of value-added model) in Houston Independent School District (HISD), found that teachers were adverse to the use of VAM for their evaluation and bonus system and that “teachers do not seem to understand why they are rewarded, especially because they profess that they do nothing differently from year to year as their SAS EVAAS rankings ‘jump around’” (p. 4). Additionally, teachers in HISD who did not earn merit pay perceived that the type of students they taught negatively biased their scores. These findings also reflect perceived validity issues as well as misalignment with educator views and values.

Collins' (2014) study of educators' perceptions of and experiences with EVAAS in a large, urban district that uses EVAAS for high-stake personnel decisions found that educators' scores fluctuated substantially from year to year and showed little consistency with observation-based measures of their teaching. Additionally, educators perceived systematic bias in EVAAS data against teachers who serve gifted students, English language learners, and students with disabilities. Respondents reported increased pressure and competition with colleagues and decreased collaboration and morale. Educators also felt that high stakes use of EVAAS data encouraged educators to cheat and to game the system by teaching to the test and drilling students. These reports by participants of perceived effects of the evaluation system suggest unintended policy effects. Collins concluded that the high stakes use of EVAAS “appears to be doing more harm than good” (p. 25).

Research (Jiang, Sporte, & Luppescu, 2015) on Chicago's REACH multi-measure evaluation system, which includes a value-added component, found that teachers are overall positive about REACH, but they have concerns about the value-added component, including a lack of clarity in how the component was calculated and incorporated into their evaluations; concerns over the value-added component weighing too heavily into their overall evaluation; and concerns about fairness, based in part on a sense that value-added data was influenced by things beyond their control.



Additionally, respondents reported increase in stress as a function of REACH and concern that the effort involved in REACH overstretched its benefits. These findings suggest that educators' perceptions could be influenced by un/familiarity with value-added. Additionally, Jiang et al. also found that contextual factors – including teachers' experience and teaching assignment (elementary versus secondary and special education versus general education) are related to their perceptions.

Balch and Koedel (2014) identified four key issues that teachers have with value-added: 1) Differentiated students: How can the model account for differences in the types of students a teacher serves (e.g., students of poverty, students with disabilities, etc.)? 2) Student attendance: How can the model account for students with problematic attendance? 3) Outside events and policies: How can the model account for major events, such as excessive snow days and policy changes, such as the move to Common Core? 4) Ex ante expectations: Why do teachers not have access to students' predicted scores in advance? Balch and Koedel argue that addressing teacher questions and concerns "has the potential to increase teacher engagement and help promote the sustainability of evaluation systems that can be useful for improving instruction" (p. 10). This argument supports efforts to examine teacher perceptions and sense-making of value-added – and their responses to it – to make new generation evaluation systems more successful and, by extension, to ensure that more students have effective teachers and that those teachers are more equitably distributed across schools.

The arrows in Figure 1 represent the notion that these various areas of consideration do not exist in isolation. Rather, they interact with one another in potentially powerful ways. For example, Harris and Herrington (2015) point out that educators' responses to the use of value-added in teacher accountability systems depend in important ways on the design of those systems, as the work of Dee and Wyckoff (2013; 2015) suggests. Additionally, technical elements of a VAM can intersect with policy in tricky ways. For example, EVAAS models use successive data to refine previous teacher value-added estimates from prior years, which is highly problematic for policy design, given that hiring decisions and dismissals would need to be made prior to the receipt of value-added score adjustments, which could call those decisions into question (Ballou & Springer, 2015). Additionally, the process of linking teachers to students for value-added score purposes incorporates technical and policy elements – such as whether to allow fractional linkages as in New York State – as well as considerations regarding teacher practice, given that teachers could potentially game the system through the linkage process (Ballou & Springer, 2015). Thus, research on teacher accountability needs to recognize that these five considerations interact in potentially complex and profound ways.

Using the conceptual framework from Figure 1, this study examines considerations regarding practice related to the use of EVAAS for educator evaluation. Specifically, it examines 1) the alignment of policy with educator views/values; 2) educators' perceptions of validity, including fairness, trust, and accuracy of value-added; 3) educators' predictions of the effects of the use of value-added for educator evaluation; 4) educators' reported observations of the effects (i.e. perceived effects) of value-added for educator evaluation; and, 5) whether educators more familiar and experienced with the use of value-added vary in their perceptions compared to educators less familiar and experienced. These five aspects of considerations regarding practice are potentially influenced by the context in which educators are situated, in terms of the students they teach, how long they have been teaching, and the characteristics of the schools they serve. Additionally, some of these issues of practice may intersect with equity and social justice considerations.

This study builds upon and extends the current literature by a) focusing on predicted and perceived effects of the use of value-added for educator evaluation; and b) by examining differences in perceptions based on respondents' familiarity and degree of experience with value-added.

Additionally, this study aims specifically to speak to the U.S. Department of Education's initiative to increase equity of educator effectiveness across schools and to inform midcourse accountability policy corrections.

## Method

### Study Site

The study site, Abrams County Schools (pseudonym) in North Carolina, serves approximately 22,000 students in 41 schools. The district spans a large geographic area that includes rural areas, suburban areas, and one large, urban area. Approximately 22% of students are Hispanic/Latino, 21% are African American, 51% are white, and 5% fall into another category. Approximately 56% of district students receive free/reduced lunch.

### North Carolina Educator Evaluation System

Under North Carolina's new educator evaluation system, which went into effect with the 2011-2012 school year, part of a teacher's evaluation is based on student growth. This is known as Standard 6. Standard 6 ratings are initially based on the best two years of scores in a three-year period from a set of value-added models known as the Education Value Added Assessment System (EVAAS); after this initial status score, a three-year rolling average of scores will be used. An educator who "does not meet expected growth" based on Standard 6 will automatically be designated as *in need of improvement* and placed on an improvement plan (NC Department of Public Instruction, 2012). Teachers who do not improve under the plan can be subject of termination. Because educators must have three years of value-added data before a status is assigned, no teachers in North Carolina to date have been subject to improvement plans or termination due to Standard 6; the first status designations will be assigned when 2014-2015 Standard 6 data is received in fall of 2015. The number of educators whose evaluations are informed by EVAAS data has been increasing each year as additional assessments are operationalized.<sup>1</sup> In 2013-2014, the school year in which these data were collected, the following subgroups of educators were to receive evaluations that incorporated individual EVAAS data: K-8 teachers; high school English, math, science, and social studies teachers; career-technical teachers; teachers of gifted students; teachers of English Language Learners; and teachers of students with disabilities.

### Instrumentation

Data come from a Web-based, self-administered, anonymous survey that contained 32 items including demographic, attitudinal, and open-ended items. Survey development involved three phases: 1) An initial survey was informed by a modest qualitative (interview) study of Abrams educators ( $n = 9$ ) in spring/summer of 2012. Interviews and a review of the literature led to the identification of certain constructs around which items were developed: knowledge/familiarity with EVAAS/Standard 6; attitudes towards teacher accountability; perceptions of validity, including consequential validity (Messick, 1998); and predicted effects of teacher evaluation policy. 2) The initial survey was piloted (Litwin, 2003) with a different set of educators in fall of 2012 ( $n = 16$ ). The pilot led to the revision of several items and the elimination of one. 3) In fall, 2012, Abrams educators took the Year 1 version of the survey, the purpose of which was to serve as a baseline to examine changes in educators' perceptions over time. Based on the results of this administration, further revisions were made to the survey.<sup>2</sup> Data for this study come from the Year 2 administration of the survey in fall of 2013.

Two sets of psychometric analyses were conducted on the instrument. Tests of internal reliability using Cronbachs alpha were conducted on the following subscales: respondents' familiarity/knowledge of EVAAS/Standard 6 (6 items;  $\alpha = .89$ ); attitudes towards use of EVAAS in educator evaluation (2 items;  $\alpha = .77$ ); and perceptions of validity (6 items;  $\alpha = .70$ ). Additionally, items regarding respondents' predictions of the effects of EVAAS/Standard 6 were examined using principle components factor analysis using varimax orthogonal rotation with Kaiser normalization, which identified four factors with eigenvalues greater than one: predicted effects on collegiality (2 items); predicted effects on students (3 items); predicted effects on teachers (3 items); and predicted effects on education quality (7 items).<sup>3</sup> All items had primary loadings over .6. The four factors explain 72.7% of the variance. The factor loading matrix for the final solution is presented in Appendix A.

### Sample

In fall, 2013, all Abrams educators received a link to the Web-based survey. Of the approximately 1600 Abrams teachers, about 1105 met inclusion criteria – those to be evaluated in 2013-2014 by Standard 6. A total of 206 inclusion-eligible people responded to the survey (18.6% response rate). While this response rate appears low, it is within the typical range for large-scale (> 1000 recipients), Web-based surveys (e.g., Hardigan, Succas, & Fleisher, 2012; Sinclair & O'Toole, 2012). However, response rates varied considerably amongst items, with some items hovering around 150 responses. There appears, though, to be no consistent internal non-response pattern that would indicate an issue of representativeness, beyond lack of familiarity/knowledge of aspects of the evaluation system. This is discussed in the findings section, where relevant.

Additionally, testing for nonresponse bias is considered a more appropriate measure of representativeness than response rate (Davern, 2013). A sample/population comparison (Davern, 2013) non-response bias test was conducted (Chi Square goodness-of-fit test) and found no statistically significant differences between the sample and population in terms of race, gender, and years experience, suggesting that general non-response bias was not an issue. An additional sample/population comparison non-response bias test was conducted comparing the sample to the population in terms of teaching assignment (K-3; 4-5; 6-8; 9-12; Career Technical Education; and Special Populations, including students with disabilities, gifted students, and English Language Learners). Chi Square goodness-of-fit test identified a significant difference  $X^2 (5, N = 144) = 36.17$ ,  $p < .001$ . The sample underrepresents K-3 teachers and over-represents 4-5 teachers. The implications of this are discussed in the findings section, where relevant. Additionally, sample/population comparisons based on demographics do not necessarily indicate the degree to which the sample is representative of the population in unobservable ways germane to the specific perceptions being measured, in this case teachers' perceptions of the use of EVAAS data as a component of their evaluations.

### Limitations and Delimitations

The key delimitations are that this study focuses on educators in one district in one state that uses one particular type of VAM. Limitations include overall low response rate and possible non-response bias in terms of grade level assignments of respondents (underrepresentation of K-2 teachers and overrepresentation of 4-5 teachers). Additionally, while 206 inclusion-eligible teachers responded to the survey, some items had substantially lower response rates (hovering around 152 responses). Demographic items tended to have the highest response rates, and items assuming knowledge/familiarity tended to have the lowest response rates. Thus lack of familiarity/knowledge

may interfere with the ability to accurately examine educators' views. Based on these delimitations and limitations, generalizability is limited, and the study should be considered exploratory.

### Analysis

Analysis attended to two elements of the conceptual framework: considerations regarding equity and social justice and considerations regarding practice, including policy alignment with educator views and values; perceived validity; predicted effects; reported (perceived) effects; and ways in which the aforementioned may be influenced by familiarity and context. Descriptive, inferential (independent t-test, ANOVA), and correlational analyses were conducted on the quantitative data. Qualitative data were analyzed using an iterative process (Glesne, 2015) that involved line-by-line coding (micro-analysis; Stringer, 2009) using *a priori* codes drawn from the conceptual framework and literature (e.g., fairness, trust, accuracy, collegiality) as well as open coding (e.g., lack of control, pressure/stress/anxiety).

## Findings

### Knowledge/Familiarity

One of the most striking things about the findings is participants' lack of familiarity with value-added/EVAAS and Standard 6 of their evaluation system. On a scale of 0 (not at all) to 10 (extremely), participants were asked to rate their familiarity with EVAAS/value-added on a number of elements (see Table 1). Findings indicate that respondents are weakly to moderately familiar with EVAAS/value-added and that they are most familiar with its limitations/weaknesses and least familiar with research about the use of EVAAS/value-added to evaluate educators. Perhaps even more troubling is that substantial percentages of educators were not sure whether they received an EVAAS rating in 2012 (28%) and were not sure whether they were to receive one in 2013 (43%). Additionally, of those who knew they received EVAAS ratings in 2012, 13% indicated that they did not go online to look at their data.

Table 1

*Respondent Familiarity with Value-Added/EVAAS*

Item	Min Value	Max Value	Mean	Standard Deviation	N
How familiar are you with how school-level and teacher-level EVAAS/ value-added is calculated?	0.00	10.00	3.99	2.61	149
How familiar are you with the benefits/ strengths of using EVAAS/ value- added to evaluate educators?	0.00	10.00	3.68	2.56	152
How familiar are you with the limitations/ weaknesses of using EVAAS/ value-added to evaluate educators?	0.00	10.00	4.87	2.96	150
How confident are you that you can accurately read and interpret teacher-level EVAAS/ value-added data?	0.00	10.00	4.25	2.96	151
How familiar are you with research about the use of value EVAAS/ value-added to evaluate educators?	0.00	10.00	3.14	2.50	145
How knowledgeable are you about how EVAAS will be used for Standard 6 on teacher/principal evaluations?	0.00	10.00	4.45	2.91	147

While overall respondents have limited familiarity with EVAAS/value-added, those who knew they had received EVAAS scores/Standard 6 ratings the previous year, compared to those

who did not or were unsure, were significantly more familiar with and knowledgeable of EVAAS/value-added, across all six aforementioned items except familiarity with research (see Table 2). Effect sizes (Cohen's  $d$ ; see Table 2) for all five significant findings fall within the moderate range, suggesting that previous receipt of value-added scores has a moderate association with perceptions of knowledge about EVAAS/value-added. Given the underrepresentation in the sample of K-3 teachers, who had not previously received EVAAS, it is possible that these data overestimate the familiarity/knowledge of the population. Regardless, these data suggest that experience with EVAAS/value-added scores is associated with increased familiarity with and knowledge of EVAAS/value-added. It is important to emphasize that none of the six items regarding familiarity/knowledge had a mean above 6.0 on a ten-point scale, indicating that even those who had previously received scores had only moderate familiarity/knowledge of EVAAS/value-added. Additionally, it is possible that lack of knowledge/familiarity is influencing non-response on some survey items. This question will be taken up in the section that follows.

Those who knew they had received EVAAS scores/Standard 6 ratings the previous year also had significantly more positive perceptions of the sufficiency of the professional development they had received on EVAAS/value-added (see Table 3). The effect size ( $d = 0.35$ ) indicates that the association of previous receipt of EVAAS scores to perceptions of the sufficiency of professional development is modest to moderate. Overall, though, only 22% of respondents felt that the professional development they received on EVAAS/value-added was fairly (19%) or completely (3%) sufficient, and 13% claimed they had received no professional development. One respondent wrote, "Most of the professional development on EVAAS has included administrators saying, 'This is what I heard, but I don't know anything else . . . I don't really know.' They've also provided us with incorrect information." Another wrote, "I think someone from central office came to talk to us once, 3 years ago." Another respondent shared, "We were given a brief overview, but I have not had the chance to really look at it to ensure that I understand how to use/read it."

Table 2

*Results of t-tests and Descriptive Statistics, EVAAS Familiarity by Previous Receipt of EVAAS/ Value-added Data*

	Respondents who Received EVAAS/ Value-Added Data the Previous Year			Respondents who did not Receive EVAAS/ Value-Added Data the Previous Year or were Unsure			95% CI for Mean Difference	<i>t</i>	<i>df</i>	Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>				
How familiar are you with how school-level and teacher-level EVAAS/ value-added is calculated?	4.61	2.70	55	3.58	2.46	93	0.18, 1.90	2.39*	146	0.40
How familiar are you with the benefits/ strengths of using EVAAS/ value-added to evaluate educators?	4.56	2.54	55	3.13	2.41	96	0.62, 2.56	3.47**	149	0.57
How familiar are you with the limitations/ weaknesses of using EVAAS/ value-added to evaluate educators?	5.73	2.67	56	4.29	2.97	93	0.49, 2.40	2.98**	147	0.49
How confident are you that you can accurately read and interpret teacher-level EVAAS/ value-added data?	4.91	3.20	58	3.80	2.73	92	0.36, 2.15	2.19* <sup>a</sup>	148	0.37
How familiar are you with research about the use of value EVAAS/ value-added to evaluate educators?	3.58	2.70	55	2.81	2.30	89	0.16, 1.76	1.83	142	0.31
How knowledgeable are you about how EVAAS will be used for Standard 6 on teacher/principal evaluations?	5.09	2.79	57	3.98	2.88	89	0.32, 2.10	2.30*	144	0.38

Note. \*  $p < .05$ ; \*\*  $p < .01$ <sup>a</sup> Due to violation of the assumption of equal variances identified by Levene's Test ( $F = 4.00, p < .05$ ), equal variances are not assumed.

Table 3

*Results of t-tests and Descriptive Statistics, Sufficiency of Professional Development by Previous Receipt of EVAAS/ Value-added Data*

	Respondents who Received EVAAS/ Value-Added Data the Previous Year			Respondents who did not Receive EVAAS/ Value-Added Data the Previous Year or were Unsure			95% CI for Mean Difference	<i>t</i>	<i>df</i>	Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>				
Sufficiency of professional development received (1-6; 1 = no professional development; 6 = completely sufficient).	3.49	1.30	57	2.99	1.42	104	0.05, 0.95	2.20 *	15 9	0.35

*Note.* \*  $p < .05$

### Policy Alignment with Educator Views/Values

Respondents overwhelmingly (74%) agreed/strongly agreed that educators should be held accountable for student learning, although only 32% felt that data on student growth should be incorporated into educator evaluations, and even fewer (8%) felt that value-added, specifically, should be part of educator evaluations (see Table 4). Interestingly, regarding views about being held accountable for student learning and incorporating growth into educator evaluations, there were no significant differences in the views of educators who had previously received EVAAS/value-added scores compared to those who did not or were unsure; however, those who had previously received EVAAS/value-added scores were significantly more likely to disagree with the use of value-added for educator evaluation (see Table 5), although the practical significance (effect size) is modest ( $d = 0.34$ ). In other words, when it comes to views specifically about the use of value-added (as opposed to more general sentiments about teacher accountability and use of growth data) those who have received value-added data are significantly more likely to disagree with its use for educator evaluation. This suggests that educators who have experience with value-added are more skeptical about VAM specifically. On another item that asked respondents about the degree to which they support or oppose the use of value-added for educator evaluation, respondents who had previously received EVAAS/value-added scores for Standard 6 were significantly more opposed to the use of value-added for educator evaluation than those who had not previously received value-added scores or were unsure whether they had (see Table 6), although, again, the practical significance is modest ( $d = 0.34$ ). Thus while overall support for the use of value-added for educator evaluation is low, it is significantly lower amongst those who had experienced its use in their evaluation. Interestingly, knowledge/familiarity of EVAAS/value-added is not directly correlated with attitudes towards teacher accountability ( $r$ -values range from  $-.083$  to  $.029$ , with no significant findings), so it is possible that lack of knowledge/familiarity – while possibly influencing item non-response – is not substantially distorting findings.

Table 4  
*Views Regarding Educator Evaluation*

Item ( $n = 151$ )	Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree	Mean	Standard Deviation
Educators (teachers and principals) should be held accountable for student learning.	2.6% (4)	8.6% (13)	14.6% (22)	58.9% (89)	15.2% (23)	3.75	0.91
Data on student growth should be incorporated into educator evaluations.	15.2% (23)	27.2% (41)	25.8% (39)	27.2% (41)	4.6% (7)	2.79	1.14
Value-added should be part of educator evaluations.	24.5% (37)	29.8% (45)	37.7% (57)	7.3% (11)	0.7% (1)	2.30	0.94



Table 5

*Results of t-tests and Descriptive Statistics, Perceptions Regarding Value-Added by Previous Receipt of EVAAS/ Value-added Data*

	Respondents who Received EVAAS/ Value-Added Data the Previous Year			Respondents who did not Receive EVAAS/ Value-Added Data the Previous Year or were Unsure			95% CI for Mean Difference	<i>t</i>	<i>df</i>	Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>				
Educators (teachers and principals) should be held accountable for student learning.	3.75	0.79	57	3.75	0.99	93	-0.30, 0.31	0.01	148	0.00
Data on student growth should be incorporated into educator evaluations.	2.75	1.27	57	2.83	1.05	93	-0.47, 0.32	-0.38 <sup>a</sup>	148	0.06
Value-added should be part of educator evaluations.	2.11	0.99	57	2.43	0.89	93	-0.63, -0.02	-2.08*	148	0.34

Note. \*  $p < .05$ <sup>a</sup> Due to violation of the assumption of equal variances identified by Levene's Test ( $F = 5.09, p < 0.05$ ), equal variances are not assumed.

Table 6

*Results of t-tests and Descriptive Statistics, EVAAS Familiarity by Previous Receipt of EVAAS/ Value-added Data*

	Respondents who Received EVAAS/ Value-Added Data the Previous Year			Respondents who did not Receive EVAAS/ Value-Added Data the Previous Year or were Unsure			95% CI for Mean Difference	<i>t</i>	<i>df</i>	Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>				
Opposition/Support for the use of value-added for educator evaluation (1-5; 1 is most opposed; 5 is most in favor).	1.65	0.86	57	1.97	0.95	92	-.62, -.01	-2.06*	147	0.34

*Note.* \*  $p < .05$

Respondents who wrote that they support teacher accountability qualified their support with stipulations about the nature of the accountability system and the tests upon which it is based. For example, one respondent wrote: “I don't have a problem with accountability, but I do have an issue with being held accountable for factors way outside the scope of my influence,” and “YES, teachers should be held accountable, but that needs to be done on a school and district level, by multiple observations, not by evaluating teachers using student data.”

In summary, respondents generally support being held accountable for student learning but are more skeptical about the use of student test data – and more specifically value-added data – being incorporated into their evaluations. Those who have previously received value-added scores as part of their evaluations are significantly more opposed to the practice.

### **Perceived Validity: Fairness, Trust, and Accuracy**

Validity involves accuracy of findings, and consequential validity (Messick, 1998) is concerned with issues of fairness, transparency, utility, and credibility (Admiraal, Hoeksma, van de Kamp, & van Duin, 2011). Respondents generally felt that value-added is neither a fair nor accurate way to evaluate educators, and they question the credibility of the measure (see Table 7). Only 6% of respondents agreed/strongly agreed that value-added is a fair way to evaluate educators, and 7% felt that it is an accurate way to evaluate educators. Respondents who had previously received value-added scores were significantly more skeptical about the fairness and accuracy of value-added (see Table 8), and the practical significance is moderate ( $d = 0.40$  and  $0.42$ , respectively; see Table 8). Additionally, a strong majority of teachers believe that educators who work with certain students (79%) or who work at certain schools (79%) will get better value-added scores, regardless of whether they are better teachers. Additionally, the majority of respondents (57%) felt that EVAAS ratings have little to no credibility, and only 13% feel that EVAAS ratings are “pretty” or “very” credible.

Table 7

*Perceived Fairness and Accuracy*

Item ( <i>n</i> = 152, unless otherwise noted)	Strongly Disagree	Disagree	Neither Agree Nor Disagree	Agree	Strongly Agree	Mean	Standard Deviation
Value-added is a fair way to evaluate educators. ( <i>n</i> = 151)	27.2% (42)	32.5% (49)	33.8% (51)	5.3% (8)	0.7% (1)	2.19	0.93
Value-added is an accurate way to evaluate educators.	30.3% (46)	36.2% (55)	26.3% (40)	5.9% (9)	1.3% (2)	2.12	0.96
Value-added cannot capture the breadth and depth of what I do as an educator.	2.0% (3)	3.3% (5)	10.5% (16)	25.7% (39)	58.6% (89)	4.36	0.94
Value-added is difficult to understand. ( <i>n</i> = 151)	1.3% (2)	9.9% (15)	33.8% (51)	35.8% (54)	19.2% (29)	3.62	0.95
Educators who work with certain students will get better value-added data, regardless of whether they are better educators.	2.0% (3)	5.3% (8)	13.8% (21)	32.2% (49)	46.7% (71)	4.16	0.97
Educators who work at certain schools will get better value-added data, regardless of whether they are better educators.	0.7% (1)	3.9% (6)	16.4% (25)	32.9% (50)	46.1% (70)	4.20	0.90

Table 8

*Results of t-tests and Descriptive Statistics, Perceptions of Fairness and Accuracy by Previous Receipt of EVAAS/ Value-added Data*

	Respondents who Received EVAAS/ Value-Added Data the Previous Year			Respondents who did not Receive EVAAS/ Value-Added Data the Previous Year or were Unsure			95% CI for Mean Difference	<i>t</i>	<i>df</i>	Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>				
Value-added is a fair way to evaluate educators.	1.96	0.89	57	2.33	0.93	93	-0.67, -0.07	-2.41*	148	0.40
Value-added is an accurate way to evaluate educators.	1.88	0.91	57	2.28	0.96	94	-.071, -0.09	-2.54*	149	0.42
Value-added cannot capture the breath and depth of what I do as an educator.	4.42	0.94	57	4.31	0.94	94	-0.20, 0.43	0.71	149	0.11
Value-added is difficult to understand.	3.65	1.14	57	3.60	0.83	93	-0.27, 0.37	0.27	148	0.04
Educators who work with certain students will get better value-added data, regardless of whether they are better educators.	4.32	0.99	57	4.11	0.93	94	-0.11, -0.53	1.31	149	0.21
Educators who work at certain schools will get better value-added data, regardless of whether they are better educators.	4.32	0.91	57	4.16	0.83	94	-0.13, -0.44	1.08	149	0.18

*Note.* \*  $p < .05$

Across the qualitative data, there were several themes regarding the perceived unfairness and inaccuracy of value-added scores as a way to measure teacher effectiveness:

**Unaccounted for variables influence value-added scores.** One respondent stated, “There are just too many unmeasurable factors. It is not an accurate view of what teachers do in the classroom. It is based on one brief test.” Another lamented, “We do not teach in a vacuum, and this standard puts all accountability on the teacher.” Another respondent described the myriad influences on students’ performance, a number of which she feels little control over:

I teach at a school full of at risk students. This year so many of our students are preoccupied with other things, home life, food, warmth, and do not see the value that education has to offer. More than ever before, discipline and dedication/ apathy of the student has become the priority in class. Lack of parenting is HUGE, so classroom management has become the priority.

For this educator, things that she cannot directly influence – such as a student’s home life – have ramifications for what she must address in class – apathy and discipline. These things foreground content instruction. Another educator explained:

Standards 6 & 8 [for principal evaluation] discount the home environment, medical needs of students, and the myriad of other factors that combine to make a student either successful or unsuccessful. I do believe that teachers/administrators need to be held accountable for student achievement; however, I believe this particular system to be riddled with flaws.

Notably, 33% of respondents indicated that one third or more of students in their school are facing significant health, emotional, and/or academic challenges. Some respondents believe these factors influence value-added scores. Another responded simply stated, “There are far too many factors not taken into consideration regarding student growth that cannot be measured by one 30 question test.” Another explained how performance on the test can reflect other factors:

If they have a bad day, it looks like they made no growth and that I did nothing to help them as a teacher. Too bad if the day they test happens to be a day that their parents get divorced or they’re fighting a cold and don’t test well. All we get is that one score.

**Value-added cannot capture the complexity of teacher work.** Many respondents feel that because Standard 6 is based on brief tests given on one day, Standard 6 cannot capture the complexity and “full breadth” of teacher work. One respondent wrote, “I find it defeating that the entire year of teaching comes down to kids taking a test on one day.” Another explained, “I want to be rewarded for strong teaching but am unsure as to whether or not one test will show the true results of my teaching.” Another emphasized, “There are so many other aspects of teaching that are not part of teaching content.” Another stated, “You can’t measure the social skills that I teach my students, or the character building I do.” A single test of the formal curriculum cannot, in respondents’ views, reflect the complexity of teacher work.

**Value-added scores reflect, to some degree, the students whom one has been assigned to teach.** Many respondents feel that value-added scores reflect the students one teaches, as exemplified by this response: “I’ve seen people be concerned about which students they were working with because of the data and the reflection on the teacher.” One respondent wrote, “Our students are so far behind when they enter our building (only 29% of a recent freshman class could read on grade level) that we struggle to teach them high school material and have them be successful.” Another explained, “It is very difficult to grow Honors students who are already at the top of their achievement levels. EC students [exceptional children/students with disabilities] have a much higher possibility of growing. I was a highly effective teacher when teaching inclusion and a

neutrally effective teacher when teaching the upper-level students.” Another respondent explained, “As a teacher of gifted students, it’s unfair to judge my students based on growth when they have come to me at the 99<sup>th</sup> percentile. On the other hand, if you judge me based on scores [achievement] and not growth I have an unfair advantage.”

Beyond perceived bias in the value added-model, a number of respondents perceive that some students are easier to teach, and some schools are easier to teach at: “Every class is different and every student is different and some classes are easier and some students are easier, and some schools are easier to teach at and it is very difficult to compare teachers based on test scores accurately.” Further, some respondents believe that great teachers – because of the students they are assigned to teach – may be unfairly judged by value-added: “Usually the great teachers are the ones that are assigned to teach the low performing students because they are better teachers, which does not help that teacher’s evaluation. Growth can show some improvement, but in struggling schools with struggling students, this does not capture the entire picture.” An elementary teacher illustrated:

The class I taught last year was similar to this one. When there are multiple students who are functioning on a KINDERGARTEN level coming into a fourth-grade class, they could make two years growth and still fail the [state standardized test] miserably because they are being tested on the fourth grade level, and they still aren’t anywhere close yet. I have five such students in my class this year and had that many last year. That is 25% of my class. Also, when all the EC [exceptional children/students with disabilities] students are concentrated in one class, it is not accurate because some of these students have different goals. i.e. a child with autism functioning on a kindergarten level is working mainly on social skills in the room with me and working mainly on academic goals ON HIS LEVEL with the EC teacher and in small groups with me. His IEP [individual education plan] does not have as a goal for him to suddenly be on grade level, so why does the state deem me "not proficient" if he doesn't get there, but DOES meet his IEP goals?

For this teacher, the growth of students who are well below grade is unlikely to be accurately captured by a grade level test. Additionally, her work with students with disabilities is judged by a grade level test and not the degree to which she helped students meet their IEP goals.

In summary, many respondents feel that value-added scores reflect the students one is assigned to teach. This is particularly the case with students who are multiple years below grade level, students with disabilities, and gifted and high performing students.

**Contextual factors influence value-added scores.** Respondents perceive that personal, classroom, school, and district contexts can influence student performance. Student mobility/transience is perceived by one respondent as disruptive to classroom culture: “Our school serves the lowest economic area in our community, so we have a lot of new students added during the school year which disrupts the flow of the class.” Approximately 22% of respondents indicated that their school experiences high student mobility/transience. Class size is also perceived as influencing growth: “With class sizes greater than 32 in many cases, our students will have a difficult time producing a years growth in the year.” School and district leadership is also a perceived influence on performance:

I have taught at four different schools for 16 different administrators. I learned that the leadership in a school has a profound effect on the success of the school. I’ve also worked for two different systems, and have found that some systems offer no support, and some systems restrict teachers with their own pacing guides and programs.

Additionally, a teacher’s personal context is perceived to influence score meaningfulness:

My value was calculated without essential information. That year I had been out 76 days on maternity leave and with my husband having heart surgery. That was not taken into consideration. Also, there were students on my roster whom I didn't teach because the EC [exceptional children/students with disabilities] teacher pulled them instead of leaving them in inclusion. That was not considered either. Another thing is that I taught two subjects, but was only assessed on one. This year's [value-added] will be based on a new test on a new curriculum, and I hope that is somehow figured into my score.

Because contextual factors such as student transience, class size, leadership, and personal situation are not accounted for in the value-added model, many respondents see the model as unfair and inaccurate.

**The tests used to calculate value-added are problematic.** A number of respondents communicated skepticism about the tests used to calculate value-added. One explained, "As a science teacher, where students do not take a state assessment each year, I think that the rating does not really measure my effectiveness very accurately." Some educators believe that grade level tests currently used to calculate value-added in North Carolina have little stretch (few questions above and below grade level), making them problematic for measuring growth:

[. . .] with the current testing system, you would not be able to see the growth from students of poverty who are significantly behind their peers. Unfortunately, at my school, over 75% of fourth-grade students are not on grade level, and over half of those are significantly below grade level. This is not a situation that has to be dealt with at non-title I schools.

According to this line of thinking, if tests cannot accurately capture a student's achievement level, then they cannot accurately be used to measure growth over time. Another teacher identified other test-related issues: "Until there are better measures for assessing student growth, and assessments that can be compared year-to-year (assessments keep changing), it is not fair to compare growth of a student on different assessment measures and say this teacher made these students grow."

Beyond general test shortcomings, some respondents felt that current tests for students with disabilities are particularly problematic:

I am a 3-5 EC [exceptional children/students with disabilities] teacher. My classes made up of students with severe disabilities, yet they are expected to take an end of grade test that in no way measures their ability level. I am talking about students who are non-ambulatory, nonverbal and rely on someone for everything from feeding to bathrooming. We work on the most basic skills, yet at the end of the year they are given a test, even though supposedly modified, they are still expected to read, add, multiply, and find the perimeter as well as other academic problems so far above their cognitive level ... yet because they are [grades] 3-5 I will be evaluated as well on their scores. They won't show growth on the test because it doesn't test them on their ability level.

Respondents indicate concerns about a ceiling effect for high achieving students and a floor effect for low achieving students and students with disabilities. One respondent explained, "One day when there is a test that can test students more accurately on their level and see if they have grown, then the scores might accurately reflect what goes on in the classroom."

In summary, respondents generally feel that the use of value-added for educator evaluation is neither fair nor accurate, and many take issue with the evaluation system because they feel that unaccounted for variables influence value-added scores; value-added cannot capture the complexity of teacher work; value-added scores reflect, to some degree, the students whom one has been



assigned to teach; personal, classroom, school, and district contextual factors influence value-added scores; and the tests used to calculate value-added are problematic.

### **Educator Predicted Effects**

Respondents were generally pessimistic about how they feel Standard 6 of the educator evaluation system will impact education (see Table 9). Most striking were perceptions that the use of Standard 6 for educator evaluation will *not* result in more equitable distribution of good educators across schools (73%); rather, participants believe educators will avoid working with certain students (74%) and will leave certain schools (70%) because of Standard 6. Those who had previously received value-added scores were significantly more skeptical that Standard 6 will result in more equitable distribution of effective educators across schools and significantly more likely to agree or strongly agree that educators will avoid working with certain students because of Standard 6 (see Table 10), and the practical significance is moderate ( $d = 0.47$  and  $0.36$ , respectively).

Additionally, 76% of respondents believed that it will be hard to recruit people into the teaching profession because of the use of Standard 6 for educator evaluation. Additionally, a majority of respondents perceived that the use of Standard 6 for educator evaluation will not lead to better teaching (64%), better student learning (64%), or even higher achievement test scores (56%). Rather, the majority of respondents perceived that Standard 6 will not improve the quality of educators (67%), will not make education a stronger profession (69%), and will ultimately harm students (57%). Additionally, 50% of respondents feel that EVAAS/value-added increases competition amongst educators, and 46% believe that it will decrease collaboration.

Table 9

*Predicted Effects of the Use of Value-Added for Educator Evaluation*

	<b>Strongly Disagree</b>	<b>Disagree</b>	<b>Neither Agree Nor Disagree</b>	<b>Agree</b>	<b>Strongly Agree</b>	<b>Mean</b>	<b>Standard Deviation</b>
EVAAS/ value-added increases competition amongst educators. ( <i>n</i> = 152)	6.6% (10)	13.8% (21)	29.6% (45)	28.3% (43)	21.7% (33)	3.45	1.17
Standard 6 will result in a more equitable distribution of good educators across schools. ( <i>n</i> = 151)	44.4% (67)	28.5% (43)	21.2% (32)	4.0% (6)	2.0% (3)	1.91	1.00
Educators will leave certain schools because of Standard 6. ( <i>n</i> = 152)	2.0% (3)	3.9% (6)	23.7% (36)	34.2% (52)	36.2% (55)	3.99	0.97
Educators will avoid working with certain students because of Standard 6. ( <i>n</i> = 152)	3.3% (5)	3.3% (5)	19.1% (29)	32.9% (50)	41.4% (63)	4.06	1.02
Standard 6 will decrease teacher collaboration. ( <i>n</i> = 151)	2.6% (4)	16.6% (25)	34.4% (52)	25.8% (39)	20.5% (31)	3.45	1.08
Standard 6 will make it harder to recruit people into the teaching profession. ( <i>n</i> = 151)	0.0% (0)	1.3% (2)	23.2% (35)	31.8% (48)	43.7% (66)	4.18	0.83
Standard 6 will ultimately harm students. ( <i>n</i> = 152)	0.7% (1)	4.6% (7)	38.2% (58)	26.3% (40)	30.3% (46)	3.81	0.95
Standard 6 will improve the quality of educators in K-12. ( <i>n</i> = 151)	34.4% (52)	32.5% (49)	24.5% (37)	6.6% (10)	2.0% (3)	2.09	1.02
Standard 6 makes education a stronger profession. ( <i>n</i> = 152)	39.5% (60)	29.6% (45)	25.7% (39)	3.9% (6)	1.3% (2)	1.98	0.97
Standard 6 will lead to better teaching. ( <i>n</i> = 152)	32.2% (49)	32.2% (49)	28.9% (44)	5.9% (9)	0.7% (1)	2.11	0.95
Standard 6 will lead to better student learning. ( <i>n</i> = 151)	31.8% (48)	31.8% (48)	31.8% (48)	4.0% (6)	0.7% (1)	2.10	0.92
Standard 6 will lead to higher achievement test scores. ( <i>n</i> = 152)	28.3% (43)	27.6% (42)	32.9% (50)	10.5% (16)	0.7% (1)	2.28	1.01

Table 10

*Results of t-tests and Descriptive Statistics, Perceptions Regarding Predicted Effects by Previous Receipt of EVAAS/ Value-added Data*

	Respondents who Received EVAAS/ Value-Added Data the Previous Year			Respondents who did not Receive EVAAS/ Value-Added Data the Previous Year or were Unsure			95% CI for Mean Difference	<i>t</i>	<i>df</i>	Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>				
EVAAS/ value-added increases competition amongst educators.	3.54	1.20	57	3.41	1.13	94	-0.25, 0.51	0.67	149	0.11
Standard 6 will result in a more equitable distribution of good educators across schools.	1.65	0.77	57	2.08	1.09	93	-0.75, 0.10	-2.81 <sup>a*</sup>	145	0.47
Educators will leave certain schools because of Standard 6.	4.09	0.91	57	3.91	1.00	94	-0.15, 0.49	1.06	149	0.17
Educators will avoid working with certain students because of Standard 6.	4.28	0.82	57	3.91	1.10	94	-0.03, 0.70	2.17 <sup>b*</sup>	143	0.36
Standard 6 will decrease teacher collaboration.	3.58	1.15	57	3.38	1.03	93	-0.16, -0.56	1.12	148	0.18
Standard 6 will make it harder to recruit people into the teaching profession.	4.25	0.76	57	4.14	0.88	93	-0.18, 0.384	0.75	148	0.12
Standard 6 will ultimately harm students.	3.81	0.93	57	3.81	0.97	94	-0.32, -0.32	-0.01	149	0.00
Standard 6 will improve the quality of educators in K-12.	2.07	1.11	56	2.11	0.97	94	-0.38, 0.31	-0.20	148	0.00
Standard 6 makes education a stronger profession.	1.81	1.01	57	2.09	0.94	94	-0.60, 0.04	-1.72	149	0.28
Standard 6 will lead to better teaching.	1.98	0.95	57	2.18	0.95	94	-0.51, 0.12	-1.24	149	0.20
Standard 6 will lead to better student learning.	2.00	0.93	57	2.16	0.92	93	-0.47, 0.15	-1.04	148	0.17
Standard 6 will lead to higher achievement test scores.	2.23	1.07	57	2.31	0.98	94	-0.41, 0.26	-0.47	149	0.08

Note. \*  $p < .05$ <sup>a</sup> Due to violation of the assumption of equal variances identified by Levene's Test ( $F = 6.65, p < 0.05$ ), equal variances are not assumed.<sup>b</sup> Due to violation of the assumption of equal variances identified by Levene's Test ( $F = 4.95, p < 0.05$ ), equal variances are not assumed.

### **Educator Reported Observed Effects (i.e., Perceived Effects)**

In order to determine whether respondents who completed open-ended items were representative of all respondents in terms of perceptions of support/opposition to the use of value-added for educator evaluation, an independent samples t-test was conducted to determine if there were differences between the two groups (those who responded to open-ended items versus those who did not). Of all 11 open-ended items, the only item that indicated a significant difference ( $t(148) = 2.03; p < .05$ ) asked participants to report their observations of effects of the use of value-added for educator evaluation. Respondents to this item were more likely to be opposed to the use of value-added for educator evaluation. As such, the results shared in this section should be interpreted with this in mind. Of 63 responses to this open-ended item asking what effects, if any, of Standard 6 respondents had seen or experienced, a handful of respondents indicated that they had not observed any effects of the use of value-added to evaluate educators, and zero respondents reported any positive effects. The large majority of respondents reported observing negative effects, including teaching to the test, “flocking away from the profession,” becoming “more selective in where they teach and with whom they teach,” and an environment in which teachers are more stressed, more anxious, and “more competitive and less trustful of their peers.” Several respondents reported dehumanizing effects: “Students have become data points rather than people.” Effects reported by the following respondent reflect common themes in the data:

Teachers who are reluctant to continue to teach students who are significantly below grade level. Teachers who want to leave our school to teach at a “better” school.

Teachers who want to leave teaching altogether in order to avoid being labeled failures after pouring their hearts and souls into their students. It is a super discouraging time to be a teacher, especially at a high-poverty school.

Five perceived effects of the use of value-added data for educator evaluation emerged as themes: Gaming the system and teaching to the test; teacher retention issues; avoiding certain students and schools; increased stress, pressure, and anxiety; and decreased collaboration and increased competition. Each of these is explored in the following sections.

**Gaming the system and teaching to the test.** A number of respondents report pressure to “teach to the test rather than helping students develop necessary skills.” Also, because of pressure to teach to the test, some respondents report a narrowing of the curriculum: “In fact, the majority of the Language Arts curriculum is not tested, so therefore, many teachers do not even teach these standards.” Another respondent reported “teaching to the test and less creativity and passion in education.” One respondent admitted:

It makes me want to give up. I had great ratings last year and all of a sudden they dropped. Did I become a bad teacher in one year?? It makes me want to only teach the test and forget about emotional and educational needs of my students.

Another respondent felt that teachers’ practices of teaching to the test gamed the system, such that less effective teachers received accolades for value-added data that did not reflect real student learning:

I have seen teachers of mediocre ability put on a pedestal, but their students cannot write complete sentences or struggle with application and synthesizing knowledge.

The students also have no global or 21st century skills whatsoever, but because the students scored well on one high-stakes test, the parents are given the wrong picture and the teachers are looked at through a flawed vision.

**Teacher retention issues.** A number of respondents report issues with teacher retention that are – at least in part – due to Standard 6. Respondents reported “teacher despair, fear, and leaving the profession;” it “has forced some teachers who love teaching to leave the profession;”

and “panic and talk about leaving the field. These are good teachers.” One educator reported, “I have seen good, experienced teachers have the wind taken out of their sails by a number. Some have chosen to pursue other careers rather than fight the system. This is sad news for our future students.” Another respondent echoes the sentiment that the current accountability system is dissuading “effective” teachers from remaining: “I am looking to get out of teaching, and I have had positive growth for EVAAS. There’s so much more accountability on teachers.” Another wrote, “Teachers want out because of all the extra work with no pay increases! Education is a horrible profession to enter these days! I’m strongly considering teaching in a private school because of the way it’s headed!” Standard 6 in the new evaluation system is just one of a number of factors that interact in complex ways to influence educator decisions to leave the profession:

I think that the new standards [6 and 8] just contribute to teachers being stressed out even more about their job. Standards seem like they are just adding to the job responsibilities that we have. A lot of teachers are already feeling burned out because of added testing, lower pay, and the new standards just make teachers want to leave the profession.

**Avoiding certain students and schools.** A respondent stated, “Bickering has begun about how unfair it is that some teachers have to teach children more likely to show little growth.” One respondent reported witnessing “teachers asking for students with IEPs [individualized education plans] to be removed from their classrooms and taught in a more restrictive setting than what the child needed and citing this standard as a reason.” Especially troubling are numerous observations about educators trying to avoid working with students with disabilities because of fear that doing so will depress their value-added scores. One respondent reported witnessing “teachers argue over who has to have the inclusion classes or the EC students in their classrooms since their jobs are on the line if their students are not making growth. Their scores are not going to be as good if the EC kids are in their classes, so they don’t want them.” Another recounted:

I have experienced Regular Education Teachers not wanting to have EC students [exceptional children/students with disabilities] in their classrooms since their jobs are on the line if their students are not making growth. They are less inclined to want to do inclusion since if the student is pulled out from the reg. ed. [general education] classroom they are not responsible for the time the student is gone.

In an inclusion classroom, a general education teacher and special education teacher work collaboratively to teach students – those on IEPs and those who are not. In North Carolina, guidelines are vague about how to determine responsibility for these students with regards to linking students to teachers for value-added calculations. When students with disabilities are pulled from the general education classroom, they do not contribute towards the general education teacher’s value-added scores. One respondent concluded, “EC students are falling through the cracks.”

Teachers are avoiding working not only with EC students but also with students who are multiple grade levels behind. One participant admitted:

I personally am requesting to NOT teach an intensive class [for students whose achievement is multiple years below grade level] next year and will likely try to move to another school because I am scared of what taking on the lowest students in my grade, and really the lowest students in my district, will do to my evaluation. I have the heart to teach these kids, but they are not the kids that show the scores on EVAAS, and I am scared if I stay too long then I will get trapped and will not be able to find another job . . . We have created "intensive" classes basically to separate the low kids from the high kids, all with the goal of improving test scores. What it does is create segregation.

These are compelling claims that the use of value-added for educator evaluation might be exacerbating educational inequities. Because some teachers are leaving for schools at which they believe they are more likely to show growth, one respondent predicted, “Weaker, less proficient teachers will be hired in the schools where the students are more needy.”

These data are troubling for multiple reasons. First, teachers are seeing students as potential score increasers or scored compressors. This dehumanizes students. Additionally, there is evidence of social justice and equity issues. Students with disabilities and students who are multiple years below grade level are being avoided. In these ways, effects of the current evaluation system in North Carolina run contrary to the federal government’s initiative to more equitably distribute teachers within and across schools and in fact may be exacerbating equity issues.

**Increased stress, pressure, and anxiety.** One of the most common and vehemently expressed effects of the new evaluation system involves educators’ feelings of stress, pressure, and anxiety. The following comment is typical of many: “Teachers are worried about the effects on their jobs, the work environment is tense and stressful, good teachers are leaving the profession so that they can have some control over their professional lives.” Another respondent wrote, “educators and administrators are scared to death about the scores.” A career-technical education (CTE) teacher shared:

My principal told me that my scores need to improve so that I will not be penalized in the future. This is due to the proficiency and mastery requirements in my courses. I still show huge growth, but how does that rate against proficiency and mastery of a standardized test in classes where products of work and learning should matter more than multiple choice answers.

For this CTE teacher, a standardized, multiple-choice test did not hold much meaning compared to student work products. Yet the test determined his Standard 6 rating.

These feelings of stress, pressure, and anxiety, according to some respondents, have a direct and negative impact on morale: “I’ve seen a drastic dip in teacher morale this year in particular. I think that, overall, legislative mandates have created an atmosphere of teachers feeling undervalued and overworked.” Another respondent wrote, “Teachers are upset and discouraged that their performance will be measured on these assessments.” Another respondent reported that she has witnessed “vastly decreased morale.”

**Decreased collaboration and increased competition.** In North Carolina, value-added is essentially a normative measure, as the progress of a teacher’s students is compared to that of students across the state to establish a typical “year’s growth.” One teacher confided, “I am already comparing myself to other teachers based on our ratings.” A number of respondents reported an increase in competition and a decrease in collaboration amongst teachers:

For me, what I have noticed this year is the lack of desire to collaborate with colleagues, even though we had done so in the past. Teachers who have been collaborative are becoming more competitive and less trustful of their peers. Teachers don’t want to release their students to the care of other professionals because they are being held accountable for the learning of those students.

## Context

The conceptual framework in Figure 1 suggests that context may influence educators’ perceptions about the use of value-added for educator evaluation. To investigate context, inferential statistical tests (independent samples t-tests and ANOVAs) and correlation analysis were conducted to examine differences in perceptions by various teacher, student, and school characteristics.

**Teacher characteristics.** Inferential and correlational statistics were calculated to examine for differences in teacher support/opposition to the use of value-added for educator evaluation by

respondent race, sex, age, and years of teaching experience (see Appendix B, Tables 11-14). There were no significant differences along any of the teacher demographics. Respondent race accounted for 6% of variance; sex accounted for 2%; and age accounted for 5%.

**Student demographics.** Independent samples t-tests were conducted to examine for differences in teacher support/opposition to the use of value-added for educator evaluation by the student demographics of their schools (see Appendix C, Table 15). All demographic data were self-reported by respondents. Specifically, the perceptions of respondents who teach at schools with the following characteristics were compared to the perceptions of respondents who teach at schools that do not have these characteristics: a) more than 50% of students are minority; b) more than 50% of students receive Free/Reduced lunch; c) more than 15% of students are English Language Learners; d) high student mobility; and e) more than 1/3 of students have significant health, emotional, and/or academic needs. There were no significant differences in respondent perceptions across any of these variables, and effect sizes ranged from very weak to weak. This suggests that teachers' support/opposition for the use of value-added for teacher accountability is not meaningfully influenced by the type of students who attend a respondent's school.

**School characteristics.** Two school-related elements of context were examined using inferential statistics: parent involvement and school setting (see Appendix D, Tables 16-17). There were no significant differences in the perceptions of respondents regarding their support/opposition for the use of value-added for teacher accountability based on whether or not they taught at a school with perceived weak parent involvement, and the effect size was weak. Regarding school setting, ANOVA indicated that there is a small and statistically significant difference in respondents' support/opposition for the use of value-added for teacher accountability based on school setting – rural ( $n = 64$ ;  $m = 2.09$ ), suburban ( $n = 75$ ;  $m = 1.68$ ), urban ( $n = 40$ ;  $m = 1.70$ ). Rural teachers are least opposed to the use of value-added, and suburban teachers are most opposed. Due to unequal group sizes, a Welch statistic was calculated ( $f[2, 149] = 3.254$ ,  $p < .05$ ,  $\eta_p^2 = 4.34$ ), indicating a significant difference. However, post hoc analyses (Bonferroni and Tukey, as well as Games-Howell for two items that signaled unequal variances) indicated no significant inter-group differences; this is due to the mean differences not being sufficiently large enough relative to the standard errors, which are influenced by sample size and variance. Thus, care should be taken not to over interpret differences in perception by school setting, which accounted for about 4% of variance.

## Discussion

According to these data, educators lack knowledge about value-added and its use in educator evaluation in North Carolina. While overall knowledge is low, it is significantly higher for those educators who had previously received value-added scores. Additionally, only 23% of respondents believe that they have received sufficient professional development about value-added and its use in educator evaluation; however, educators who had previously received value-added scores felt their professional development was significantly more sufficient. Other research has similarly identified educator issues with the opacity of VAM: Jiang et al. (2015) report teachers' misconceptions and confusion about how VAM was used in their evaluations, and Goldring et al. (2015) illustrate administrators' uncertainty about how VAM is calculated. Balch and Koedel (2014) argue that addressing educators' questions and concerns can promote buy-in, which encourages the persistence and effectiveness of teacher evaluation systems.

Interestingly, though, despite low reports of knowledge and familiarity with value-added, respondents identified some of the same challenges with VAM that scholars have pondered, including challenges regarding roster verification, the process for determining who is responsible for

which students' scores (Ballou & Springer, 2015); limitations in the ability of grade level state tests to adequately measure the learning of students who are well above and below grade level (e.g., Darling-Hammond, 2015); and the difficulty of accounting for all of the factors that influence student test scores: "No statistical manipulation can assure fair comparisons of teachers working in very different schools, with very different students, under very different conditions" (Haertel, 2013, p. 24).

While educators in this study generally felt that they should be held accountable for student learning, few believed student growth data should be incorporated into educator evaluation, and even fewer felt that value-added specifically should be used in their evaluations. Additionally, educators who had previously received value-added scores were significantly more opposed to their use in evaluations. These findings suggest policy misalignment with educator views and values, which is documented in the work of Collins (2014), Jiang et al. (2015), and in a recent Gallup poll (Lyons, 2014) that found that 89% of teachers oppose the linking of student test scores to teacher evaluations.

North Carolina educators in this study generally felt that the use of value-added in their evaluations was unfair, inaccurate, and influenced by the students whom they were assigned to teach and the schools at which they worked. Specifically, educators felt that unaccounted for variables influenced their scores; that value-added cannot capture the complexity of teacher work; that personal, classroom, school, and district contextual factors influence scores; and that the tests used to calculate value-added scores are problematic. Educators who have previously received value-added scores were significantly more likely to feel that the use of those scores in evaluations is unfair and inaccurate. These findings raise questions about whether opposition to the use of value-added in educator evaluations is more likely to diminish or flare as teachers become more experienced with these models and as evaluation consequences (improvement plans and dismissals) take effect in fall of 2015. Teachers' skepticism about the use of VAM for educator evaluation might not be quick to dissipate as teachers become more accustomed to it. Social network analysis may help to examine the degree to which "behavioral contagion" (Valente, Palinkas, Czaja, Chu, & Brown, 2015, p. 13) may occur between educators who are more experienced with the use of value-added for their evaluations and those who are less experienced. Interestingly, while educator opposition was related to neither teacher experience level nor age in this study, Jiang et al. (2015) found that beginning teachers were more positive about the use of value-added in their evaluations than were more experienced teachers. These studies raise the question whether opposition will recede as older, more experienced teachers retire and are replaced by newer teachers.

Interestingly, context, in terms school setting (rural, suburban, and urban) seems to have little to no association with overall opposition to the use of value-added for educator evaluation, nor does context in terms of student demographics based on race, poverty, English Language Learner status, mobility, low parent involvement, and high health/emotional/academic challenges facing students. Additional analyses should be conducted to determine if there are significant differences by context with other constructs (e.g., knowledge/familiarity; perceptions of validity; predicted effects) and also to examine whether contextual variables interact in complex ways with these constructs. Attention should be paid to additional elements of context that include educators' perceptions of school supportiveness (e.g., professional development, collegiality; Johnson, 2015), as well as other factors, including perceptions of leadership support for policy (Valente et al., 2015) on the use of value-added for educator evaluation.

Educators in this study predict that the use of value-added in educator evaluation will not increase the equitable distribution of effective teachers. On the contrary, they predict that educators will avoid working with certain students and avoid teaching in certain schools because of perceptions about the influence of teaching assignments and personal, classroom, school, and



district contexts on scores. Those who have previously received value-added scores are significantly more likely to believe that the practice will not increase the equitable distribution of effective educators and that educators will avoid working with certain students. Additionally, educators predict increases in competition amongst educators and decreases in educator collaboration. Other researchers have pointed out that VAM is normative in nature; teachers' scores are relative to one another's, as opposed to an absolute standard, such that teachers are essentially competing against one another for the most growth (e.g., Darling-Hammond, 2015; Goldhaber, 2015; Winters & Cowen, 2015), which may undermine collective efficacy (Raudebush, 2015, p. 140).

Educators' reports regarding perceived effects of the use of value-added for educator evaluation fall within five themes. First, educators are increasingly gaming the system and teaching to the test. Collins' (2014) study of a large, urban district in the Southwest, which like North Carolina uses an EVAAS value-added measure of educator effectiveness, documents teacher perceptions that the use of value-added for educator evaluation encourages teaching to the test as well as cheating. The potential for gaming the system is echoed by some researchers (e.g., Ballou & Springer, 2015) and may amplify a "credibility gap" (Harris & Herrington, 2015, p. 72). Additionally, these findings may reflect Campbell's Law, the notion that the greater the stakes linked to a measure, the "more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor" (Campbell, 1976, p. 49).

Second, respondents perceive that teachers are increasingly leaving the field due to teacher accountability. The North Carolina Department of Public Instruction (2014) reported a 2013-2014 statewide turnover rate of 14.1%, which ranges from 6.0% to 34.4% across districts. The county in which this study took place had a 15.5% turnover rate. The state's turnover rate in 2013-2014 was down from 14.3% in 2012-2013, which was an increase over the 2011-2012 rate of 12.1%. These statistics do not seem alarming, although they represent an increase in turnover of 16.5% over a two-year period. On the pipeline/recruitment side, the North Carolina university system has seen a 30.4% drop in educator majors over the last four years (UNC General Administration, 2015), and some North Carolina districts have decried recruitment challenges, including the state's two largest districts, Wake County (Hui, 2014) and Charlotte-Mecklenburg (Rhew, 2015). However, it is important to note that teacher evaluation is just one of several major education-related areas of recent legislative action in North Carolina. Wake County (Hui, 2014) credited the elimination of tenure (recently overturned by the North Carolina Supreme Court) and elimination of additional pay for masters' degrees as key policies related to the recruitment problem. Thus, it is difficult to determine to what extent any pipeline/recruitment and retention challenges are related to educator evaluation policy.

Third, some educators are seeking to avoid working with certain students and at certain schools. This is perhaps the most disconcerting finding of this study. Particularly problematic are respondent reports regarding the way in which educators are increasingly conceptualizing students with disabilities, students of poverty, and students who are multiple grade levels behind as score depressors. Such discourse dehumanizes students and reflects a deficit mentality that pathologizes these student groups. This discourse is toxic and must be disrupted in order for all students to be treated justly and with dignity. These findings suggest that the use of value-added for educator evaluation may have a Matthew Effect – the notion that those who are advantaged get increasingly so, and those who are disadvantaged become more disadvantaged, leading to an increasing gap between the groups over time (Kerckhoff & Glennie, 1999).

Fourth, educators are feeling an increase in stress, pressure, and anxiety. This notion of teacher stress is reflected in research on Chicago's REACH evaluation program (which also includes a component based on value-added scores), in which 79% of respondents felt increased stress and

anxiety because of the evaluation system (Jiang et al., 2015, p. 113). Similarly, Collins (2014) found increased stress and decreased morale amongst teachers.

Fifth, educator collaboration is decreasing and competition amongst educators is increasing. This may reinforce the “egg crate school” (Johnson, 2015), the notion that teachers are isolated from one another and work independently. Furthermore, Johnson (2015) makes a compelling argument that “teachers are not inherently effective or ineffective but (sic) their development may be stunted when they work alone, without the benefit of ongoing collegial influence” and that “successful school-wide improvement increases norms of shared responsibility among teachers and creates structures and opportunities for learning that promote interdependence – rather than independence – among them” (p. 119). In other words, decreased collegiality and increased competition may have an overall deleterious effect on teacher effectiveness, suggesting that the very evaluation policies established to increase effectiveness could potentially have the opposite effect.

## **Conclusion and Significance**

### **Policy**

Given the DoE’s priority to leverage teacher accountability to more equitably distribute effective teachers and to increase the number of students served by effective teachers, these findings are particularly troubling. As leading indicators, they point to several unintended and unanticipated consequences of educator evaluation policies that incorporate student growth measures, specifically value-added. First, educators’ familiarity with value-added is limited, and increased educator experience with value-added is significantly associated with deeper skepticism and more negative views of about it. Second, findings suggest that evaluation policy that incorporates value-added might exacerbate existing (Odden & Kelly, 2008) educator recruitment and retention issues, particularly in schools serving high populations of traditionally marginalized students. Policymakers must track and analyze rates at which people enter teacher preparation programs and enter the field, as well as educator turnover data, with special attention to recruitment and turnover rates at schools serving high populations of traditionally marginalized students. Within schools, findings suggest that students with special needs and those who are experiencing significant academic struggles are being segregated and are increasingly seen as score suppressors, distorting and corrupting the educational process and resulting in the “abandonment of an ethic of caring” (Nichols & Berliner, 2005, p. 166). Thus equity issues among and within schools may be exacerbated, undermining the DoE’s priority to more equitably distribute effective teachers.

These findings suggest the need for urgent and substantive midcourse policy corrections, including 1) increase the sufficiency, in terms of quantity and quality, of professional development about the use of value-added in educator evaluation, which may increase teacher buy-in and support the longevity and effectiveness of such evaluation systems (Balch & Koedel, 2014). 2) initiate a temporary moratorium on the use of student test score data for educator evaluations, a sentiment echoed by the American Statistical Association, ASCD, and the Gates Foundation (Hewitt, 2015); 3) use value-added not as a calculable component of an educator’s evaluation but as a screener to flag educators who may need further scrutiny or support, a recommendation made by a number of value-added experts (e.g., Baker et al., 2010; Hill et al., 2011; IES, 2010; Linn, 2008); 4) while recognizing that no value-added model can adequately account for all of the ways in which educators’ circumstances differ (Haertel, 2013), shift to a value-added model or other student growth measure that can address nonrandom sorting of students (Koedel & Betts, 2009) and systematic bias due to test design (Darling-Hammond, 2015); account for students who are multiple years below grade level, exceptional students (e.g., gifted students and students with disabilities), and

English Language Learners; and further examine the best way to account for transient students; and 5) implement incentives to draw and retain teachers to the most challenging settings (Donaldson, 2013).

Additionally, given that educators, according to these findings, do not anticipate increases to student learning, advances to the field of education, and improvements to the teaching profession as a function of new educator accountability policy, the theory of action that unpins such policy – namely that “teacher accountability will motivate teachers to work harder and smarter and help attract and retain only those who are successful” (Harris & Herrington, 2015, p. 72), should be revisited. Murphy, Hallinger, and Heck (2013) argue that “if school improvement is the goal, school leaders would be advised to spend their time and energy in areas other than teacher evaluation” (p. 352). Policy attention should shift away from the teacher as the unit of focus to ways in which teachers collectively and interdependently can improve their effectiveness (Johnson, 2015; Raudenbush, 2015). Data on educator performance is best used formatively and integrated with the most efficacious elements of professional learning communities into a thoughtful system of job-embedded professional development (Woodland & Mazur, 2015).

## Research

More research on educators’ perceptions of and responses to the use of value-added for teacher accountability is needed; such research should include larger, more representative samples from states using these systems. Research also needs to consider how the design of teacher accountability policies influences teachers’ perceptions of and responses to teacher accountability. Additionally, longitudinal research is needed to examine how educators’ perceptions of and responses to the use of value-added in educator evaluations change over time. Such research can inform policy corrections and evaluate the extent to which these policies are achieving their intended effects, as well causing as any unanticipated and unintended consequences.

Revisiting the conceptual framework that guided this study (Figure 1), it is clear from these findings that elements of the framework interact in complex and important ways, and research needs to examine these complex interconnections. For example, teachers’ perceptions of limitations of tests used to calculate value-added influence their perceptions of the fairness, accuracy, and credibility of teacher accountability data. Those perceptions, in turn, may influence teacher practice in profound ways, including the potential increased segregation of students with disabilities, which is an equity and social justice issue. Research is needed on how policy issues, including, for example, how students are linked to teachers in value-added models and the weight given to value-added measures as part of an educator’s overall evaluation, influence issues of practice. In other words, research is needed to examine differential impact on perceptions and educator behaviors based on important differences in the structure of evaluation policies. Empirical research is needed to speak to how these policy differences play out in actuality from state to state and district to district, and to what the most salient components of these systems are.

Value-added is the student growth measure (SGM) that has received the lion’s share of attention by scholars and the media. Policy impact research is needed on other SGMs, including student growth percentiles, which are the most commonly used SGM (Amrein-Beardsley, 2014) and a variety of non-standardized test based measures, alternatively known as student growth objectives (e.g. New Jersey), measures of student learning (e.g., New York City), student learning objectives (e.g., New York State), and analysis of student work (e.g., North Carolina), that are increasingly used to quantify teacher contributions to student learning in non-tested grades/courses. In some states, more teachers are evaluated by these SGMs than by value-added and student growth percentiles. As argued by Braun (2015), the intersection of teacher accountability and school improvement needs

examination. For example, to what extent and in what ways do new educator evaluation policies that incorporate SGMs interact with school-wide improvement efforts? Additionally, beyond the recent work of Goldering et al. (2015), little attention has been paid to how administrators make meaning of and use teacher accountability data, and there has been virtually no attention paid to evaluation systems for principals and assistant principals that incorporate SGMs.

## Notes

<sup>1</sup> In 2011-2012, teachers who taught courses tested by End of Grade (4-8 reading and math; 5 and 8 science) and End of Course tests (Math I, English II, and Biology) as part of the state's Accountability Model were to be evaluated under Standard 6. Later legislative action delayed the use of these test data for educator evaluation for one year. In 2012-2013, Final Exams were introduced as part of the Educator Effectiveness Model to produce EVAAS measures for Standard 6 in grades 4-12 in English/language arts, math, science, and social studies that were not already tested through the state's Accountability Model (through End of Grade and End of Course exams). A complete list of North Carolina final exams that are used to evaluate teachers under Standard 6 is available at <http://www.ncpublicschools.org/docs/accountability/common-exams/ncfematlist14.pdf>. Also in 2012-2013, Career and Technical Education State Assessments began to be used to evaluate Career Technical Education teachers. In 2013-2014, teachers of grades K-2 began to be evaluated using data from the mClass: Reading 3D program, and teachers of grade 3 began to be evaluated using the Beginning of Grade 3 English Language Arts/Reading Test (along with the existing End of Grade 3 English Language Arts/Reading Test). In addition to the use of EVAAS to calculate teacher effectiveness using the aforementioned tests, in 2014-2015 the North Carolina Department of Public Instruction is implementing the Analysis of Student Work process to evaluate for Standard 6 the effectiveness of arts, world languages, healthful living, Advanced Placement, and International Baccalaureate teachers. For more information on the Analysis of Student Work process, please see <http://ncasw.ncdpi.wikispaces.net/>.

<sup>2</sup> Two sets of items were eliminated. These items were designed to examine respondents' perceptions of national discourse around educator evaluation through questions that gauged respondents' perceptions of the views of people who support and people who oppose the use of value-added for educator evaluation. A small subset of respondents found these items objectionable because they believed they were polarizing.

<sup>3</sup> One item in this factor, "Standard 6 will benefit me as an educator," does not seem conceptually related to the others. As such, it has been removed from analysis for purposes of this study.

## References

- Admiraal, W., Hoeksma, M., van de Kamp, M.-T., & van Duin, G. (2011). Assessment of teacher competence using video portfolios: Reliability, construct validity, and consequential validity. *Teaching and Teacher Education*, 27(6), 1019–1028. doi: 10.1016/j.tate.2011.04.002
- American Statistical Association. (2014). *ASA statement on using value-added models for educational assessment*. Alexandria, VA: Author.
- Amrein-Beardsley, A. (2008). Methodological concerns about the education value-added assessment system. *Educational Researcher*, 37(2), 65–75. doi: 10.3102/0013189X08316420
- Amrein-Beardsley, A. (2014). *Rethinking value-added models in education: Critical perspectives on tests and assessment-based accountability*. New York, NY: Routledge.
- 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), 1–31. doi: 10.14507/epaa.v20n12.2012
- Baker, A. 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, D.C.: EPI.
- Balch, R., & Koedel, C. (2014). Anticipating and incorporating stakeholder feedback when developing value-added models. *Education Policy Analysis Archives*, 22(97), 1–17.
- Ballou, D., Sanders, W., & Wright, P. (2004). Controlling for student background in value-added assessment of teachers. *Journal of Educational and Behavioral Statistics*, 29(1), 37–66. doi: 10.3102/10769986029001037
- Ballou, D., & Springer, M. G. (2015). Using student test scores to measure teacher performance: Some problems in the design and implementation of evaluation systems. *Educational Researcher*, 44(2), 77–86. doi: 10.3102/0013189X15574904
- Braun, H. (2005). Value-added modeling: What does due diligence require? In Lissitz, R. (Ed.), *Value added models in education: Theory and applications* (pp. 19–39). Maple Grove, MN: JAM Press.
- Braun, H. (2015). The value in value added depends on the ecology. *Educational Researcher*, 44(2), 127–131. doi: 10.3102/0013189X15576341
- Briggs, D. C., & Weeks, J. P. (2011). The persistence of school-level value-added. *Journal of Educational and Behavioral Statistics*, 36(5), 616–637. doi: 10.3102/1076998610396887
- Campbell, D. (1976). *Assessing the impact of planned social change*. Hanover, NH: Dartmouth College. Retrieved from <https://www.globalhivmeinfo.org/CapacityBuilding/Occasional%20Papers/08%20Assessing%20the%20Impact%20of%20Planned%20Social%20Change.pdf>
- Carey, K., & Manwaring, R. (2011). *Growth Models and Accountability: A Recipe for Remaking ESEA*. *Education Sector Reports*. Washington, D. C.: Education Sector.
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2013). *Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood* (Working Paper No. 19424). Cambridge, MA: National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w19424>. doi: 10.3386/w19424
- Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9), 2593–2632. doi: 10.1257/aer.104.9.2593
- 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), 1–42.

- 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, R.I.: Annenberg Institute for School Reform.
- Corcoran, S. P., & Goldhaber, D., D. (2013). Value-added and its uses: Where you stand depends on where you sit. *Education Finance and Policy*, 8(2), 418–434. doi: 10.1162/EDFP\_a\_00104
- Darling-Hammond, L. (2015). Can value added add value to teacher evaluation? *Educational Researcher*, 44(2), 132–137. doi: 10.3102/0013189X15575346
- Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *The Phi Delta Kappan*, 93(6), 8–15.
- Davern, M. (2013). Nonresponse rates are a problematic indicator of nonresponse bias in survey research. *Health Services Research*, 48(3), 905–912. <http://doi.org/10.1111/1475-6773.12070>
- Dee, T., & Wyckoff, J. (2013). *Incentives, selection, and teacher performance: Evidence from IMPACT* (No. Working Paper No. 19529). Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w19529
- Doherty, K. M., Jacobs, S., & National Council on Teacher Quality. (2013). *State of the states 2013: Connect the dots: Using evaluation of teacher effective to inform policy and practice*. Washington, D. C.: National Council on Teacher Quality.
- Donaldson, M. L. (2013). Principals' approaches to cultivating teacher effectiveness: Constraints and opportunities in hiring, assigning, evaluating, and developing teachers. *Educational Administration Quarterly*, 49(5), 838–882. doi: 10.1177/0013161X13485961
- Ehlert, M., Koedel, C., Parsons, E., & Podgursky, M. (2014). Choosing the right growth measure. *Education Next*, 14(2), 1–7.
- Goldhaber, D. (2015). Exploring the potential of value-added performance measures to affect the quality of the teacher workforce. *Educational Researcher*, 44(2), 87–95. doi: 10.3102/0013189X15574905
- Goldhaber, D., & Hansen, M. (2008). *Is it just a bad class? Assessing the stability of measured teacher performance*. Seattle, WA: University of Washington/Center on Reinventing Public Education.
- Goldring, E., Grissom, J., Rubin, M., Neumerski, C. M., Cannata, M., Drake, T., & Schuermann, P. (2015). Make room value added: Principals' human capital decisions and the emergence of teacher observation data. *Educational Researcher*, 44(2), 96–104. doi: 10.3102/0013189X15575031
- Haertel, E. H. (2013, March). *Reliability and validity of inferences about teachers based on student test scores*. Paper presentation presented at the 14th William H. Angoff Memorial Lecture, Washington, D.C.
- Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review: Papers and Proceedings*, 100(May), 267–271. doi: 10.1257/aer.100.2.267
- Hardigan, P. C., Succar, C. T., & Fleisher, J. M. (2012). An analysis of response rate and economic costs between mail and web-based surveys among practicing dentists: a randomized trial. *Journal of Community Health*, 37(2), 383–394. doi: 10.1007/s10900-011-9455-6
- Harris, D. N. (2011). *Value-added measures in education: What every educator needs to know*. Cambridge, Mass: Harvard Education Press.
- Harris, D. N., & Herrington, C. D. (2015). The use of teacher value-added measures in schools: New evidence, unanswered questions, and future prospects. *Educational Researcher*, 44(2), 71–76. doi: 10.3102/0013189X15576142
- Hewitt, K. K. (2013). The use of value added for accountability and to inform leadership. In K. K. Hewitt, Childers-McKee, C., Hodge, E., & Schuhler, R. (Eds.), *Postcards from the schoolhouse*:

- Practitioner scholars examine contemporary issues in instructional leadership* (pp. 198–223). Ypsilanti, MI: NCPEA Press.
- Hewitt, K. K. (2015, May 21). Is this thing on? Amplifying the call to stop the use of test data for educator evaluations (at least for now) [invited Web log post]. Retrieved from <http://vamboozled.com/is-this-thing-on-amplifying-the-call-to-stop-the-use-of-test-data-for-educator-evaluations-at-least-for-now/#comment-413731>
- Hill, H., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, 48(3), 794–831. doi: 10.3102/0002831210387916
- Hui, T. K. (2014, February 14). Wake county schools talk about challenges to teacher recruitment efforts. *The News & Observer*. Wake County. Retrieved from <http://www.newsobserver.com/news/local/education/wake-ed-blog/article10293221.html>
- Institute of Education Sciences. (2010). *Technical methods report: Error rates in measuring teacher and school performance based on student test score gains*. Washington, D. C.: NCEE Publications. Retrieved from <http://ies.ed.gov/ncee/pubs/20104004/>
- Jackson, C. K. (2012). *Teacher quality at the high-school level: The importance of accounting for tracks* (Working Paper No. 17722). National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w17722>
- Jiang, J. Y., Sporte, S. E., & Luppescu, S. (2015). Teacher perspectives on evaluation reform: Chicago's REACH students. *Educational Researcher*, 44(2), 105–116. doi: 10.3102/0013189X15575517
- Johnson, S. M. (2015). Will VAMS reinforce the walls of the egg-crate school? *Educational Researcher*, 44(2), 117–126. doi: 10.3102/0013189X15573351
- Kalogrides, D., & Loeb, S. (2013). Different teachers, different peers: The magnitude of student sorting within schools. *Educational Researcher*, 42(6), 304–316. doi: 10.3102/0013189X13495087
- Kane, T. J., Rockoff, J. E., & Staiger, D. (2005). *Identifying effective teachers in New York City*. Cambridge, MA: National Bureau of Economic Research.
- Kane, T. J., Taylor, E. S., Tyler, J. H., & Wooten, A. L. (2011). Identifying effective classroom practices using student achievement data. *Journal of Human Resources*, 46(3), 587–613. doi: 10.1353/jhr.2011.0010
- Kerckhoff, A. C., & Glennie, E. J. (1999). The Matthew effect in American education. In A. M. Pallas (Ed.), *Research in sociology of education and socialization* (Vol. 12, pp. 35–66). Greenwich, CT: JAI Press Inc.
- Koedel, C. (2009). An empirical analysis of teacher spillover effects in secondary school. *Economics of Education Review*, 28(6):682-69. doi: 10.1016/j.econedurev.2009.02.003
- Koedel, C., & Betts, J. R. (2011). Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the Rothstein critique. *Education Finance and Policy*, 6(1), 18–42. doi: 10.1162/EDFP\_a\_00027
- Konstantopoulos, S., & Chung, V. (2011). The persistence of teacher effects in elementary grades. *American Educational Research Journal*, 48(2), 361–386. doi: 10.3102/0002831210382888
- Kupermintz, H. (2003). Teacher effects and teacher effectiveness: A validity investigation of the Tennessee value added assessment system. *Educational Evaluation and Policy Analysis*, 25(3), 287–298. doi: 10.3102/01623737025003287
- Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis*, 24(1), 37–62. doi: 10.3102/01623737024001037

- Linn, R. L. (2008). Methodological issues in achieving school accountability. *Journal of Curriculum Studies*, 40(6), 699–711. doi: 10.1080/00220270802105729
- Litwin, M. S. (20003). *How to assess and interpret survey psychometrics* (2nd. ed.). Thousand Oaks, CA: SAGE Publications.
- Lyons, L. (2014, October 29). *Teachers favor common core standards, not the testing*. Washington, D.C.: Gallup. Retrieved from <http://www.gallup.com/poll/178997/teachers-favor-common-core-standards-not-testing.aspx>
- Mariano, L. T., McCaffrey, D. F., & Lockwood, J. R. (2010). A model for teacher effects from longitudinal data without assuming vertical scaling. *Journal of Educational and Behavioral Statistics*, 35(3), 253–279. doi: 10.3102/1076998609346967
- Martineau, J. (2006). Distorting value added: The use of longitudinal, vertically scaled student achievement data for growth-based, value-added accountability. *Journal of Educational and Behavioral Statistics*, 31(1), 35–62. doi: 10.3102/10769986031001035
- McCaffey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., & Hamilton, L. S. (2004). Let's see more empirical studies on value-added modeling of teacher effects: A reply to Raudenbush, Rubin, Stuart and Zanutto, and Reckase. *Journal of Educational and Behavioral Statistics*, 29(1), 139–143. doi: 10.3102/10769986029001139
- McCaffrey, D. F., & Buzick, H. M. (2014). *Is value-added accurate for teachers of students with disabilities? What we know series: Value-added methods and applications. Knowledge Brief 14*. Stanford, CA: Carnegie Foundation for the Advancement of Teaching. Retrieved from <http://eric.ed.gov/?q=is+VALUE++ADDED+ACCURATE+FOR++TEACHERS+OF+STUDENTS+WITH++DISABILITIES+%3f&id=ED556478>
- Messick, S. (1998). Test validity: A matter of consequence. *Social Indicators Research*, 45(1), 35–44. doi: 10.1023/A:1006964925094
- Morgan, G. B., Hodge, K. J., Trepinski, T. M., & Anderson, L. W. (2014). The stability of teacher performance and effectiveness: Implications for policies concerning teacher evaluation. *Education Policy Analysis Archives*, 22(95), 1–21.
- Murphy, J., Hallinger, P., & Heck, R. H. (2013). Leading via teacher evaluation: The case of the missing clothes? *Educational Researcher*, 42(6), 349–354. doi: 10.3102/0013189X13499625
- 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(23), 1–27.
- Nichols, S. L. and Berliner, D. C. (2005). *The inevitable corruption of indicators and educators through high-stakes testing*. East Lansing, MI: Great Lakes Center for Education Research & Practice.
- North Carolina Department of Public Instruction. (2012). *North Carolina Educator Evaluation System (NCEES) description and using evaluations for decision-making*. Raleigh, NC: Author. Retrieved from <http://www.ncpublicschools.org/docs/fbs/arra/sfsf/reporting/nceesdescription.pdf>
- North Carolina Department of Public Instruction. (2014). *2013-2014 annual report on teachers leaving the profession*. Raleigh, NC: Author. Retrieved from <http://www.ncpublicschools.org/docs/educatoreffectiveness/surveys/leaving/2013-14turnoverreport.pdf>
- Odden, A., & Kelly, J. A. (2008). *Strategic management of human capital in public education*. Madison, WI: Consortium for Policy Research in Education.
- Papay, J. P. (2011). Different tests, different answers: The stability of teacher value-added estimates across outcome measures. *American Educational Research Journal*, 48(1), 163–193. doi: 10.3102/0002831210362589
- Polikoff, M. S., & Porter, A. C. (2014). Instructional alignment as a measure of teaching quality. *Educational Evaluation and Policy Analysis*, 36(4), 399–416. doi: 10.3102/0162373714531851



- Raudenbush, S. W. (2015). Value added: A case study in the mismatch between education research and policy. *Educational Researcher*, 44(2), 138–141. doi: 10.3102/0013189X15575345
- Rhew, A. (2015, June 1). Charlotte panel discusses “urgent” teacher pipeline need. *EducationNC*. Retrieved from <https://www.ednc.org/2015/06/01/charlotte-panel-discusses-urgent-teacher-pipeline-need/>
- Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73(2), 417–458. doi: 10.1111/j.1468-0262.2005.00584.x
- Rockoff, J. E. (2004). The impact of individual teachers on student achievement: evidence from panel data. *American Economic Review*, 94(2), 247–252. doi: 10.1257/0002828041302244
- Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *Quarterly Journal of Economics*, 125(1), 175–214. doi: 10.1162/qjec.2010.125.1.175
- Rothstein, J. (2014). *Revisiting the impacts of teachers*. Berkeley, CA: University of California, Berkeley. Retrieved from [http://eml.berkeley.edu/~jrothst/workingpapers/rothstein\\_cfr.pdf](http://eml.berkeley.edu/~jrothst/workingpapers/rothstein_cfr.pdf)
- Sanders, W. L. (2006, November). *Comparisons among various educational assessment value-added models*. Paper presented at the Power of Two--National Value-Added Conference, Columbus, OH.
- Schmidt, W. H., Houang, R. T., & McKnight, C. C. (2005). Value-added research: Right idea but wrong solution? In Lissitz, R. (Ed.), *Value added models in education: Theory and applications* (pp. 145–164). Maple Grove, MN: JAM Press.
- Sinclair, M., O’Toole, J. (2012). Comparison of response rates and cost-effectiveness for a community-based survey: postal, internet and telephone modes with generic or personalized recruitment approaches. *BMC Medical Research Methodology*, 12, 132. doi:10.1186/1471-2288-12-132
- Springer, M. G., Hamilton, L., McCaffrey, D. F., Ballou, D., Le, V.-N., Pepper, M., ... Stecher, B. M. (2010). *Teacher pay for performance: Experimental evidence from the project on incentives in teaching*. Nashville, TN: National Center on Performance Incentives at Vanderbilt University/RAND Corporation. Retrieved from [http://www.rand.org/content/dam/rand/pubs/reprints/2010/RAND\\_RP1416.pdf](http://www.rand.org/content/dam/rand/pubs/reprints/2010/RAND_RP1416.pdf)
- Stringer, P. M. (2009). Capacity building for school improvement: A case study of a New Zealand primary school. *Educational Research for Policy and Practice*, 8(3), 153–179. doi: 10.1007/s10671-009-9073-6
- Timmermans, A. C., Doolaard, S., & Wolf, I. de. (2011). Conceptual and empirical differences among various value-added models for accountability. *School Effectiveness and School Improvement*, 22(4), 393–413. doi: 10.1080/09243453.2011.590704
- UNC General Administration. (2015). *Enrollment trends*. Raleigh, NC: Author. Retrieved from <http://tqdashboard.northcarolina.edu/preparation/>
- U.S. Department of Education. (nd). *U.S. Department of Education strategic plan for fiscal years 2014-2018*. Washington, D.C.: Author. Retrieved from <http://www2.ed.gov/about/reports/strat/plan2014-18/strategic-plan.pdf>
- Valente, T. W., Palinkas, L. A., Czaja, S., Chu, K., & Brown, C. H. (2015). Social network analysis for program implementation. *PLoS ONE*, 10(6), 1–18. doi: 10.1371/journal.pone.0131712
- Winters, M. A., & Cowen, J. M. (2013). Would a value-added system of retention improve the distribution of teacher quality? A simulation of alternative policies. *Journal of Policy Analysis and Management*, 32(3), 634–654. doi: 10.1002/pam.21705
- Winters, M. A., & Cowen, J. M. (2015). Who would stay, who would be dismissed? An empirical consideration of value-added teacher retention policies. *Educational Researcher*, 44(2), 330–337.

- Woodland, R. H., & Mazur, R. (2015). Beyond hammers versus hugs: Leveraging educator evaluation and professional learning communities into job-embedded professional development. *NASSP Bulletin*, 99(1), 5–25. doi: 10.1177/0192636515571934
- Wright, S. P., White, J. T., Sanders, W. L., & Rivers, J. C. (2010). *SAS EVAAS statistical models*. Raleigh, NC: SAS Institute, Inc.
- Yuan, K. (2015). A value-added study of teacher spillover effects across four core subjects in middle schools. *Education Policy Analysis Archives*, 23(38), 1–24.

## Appendix A

### Factor Analysis of “Prediction” Survey Items

Table 1

*Factor loadings and communalities based on a principle components analysis with varimax orthogonal rotation for 15 items from subscale of predictions regarding effects of the use of EVAAS/Standard 6 in educator evaluation.*

	Effects on Education Quality	Student Effects	Effects on Collegiality	Teacher Effects
Standard will result in a more equitable distribution of good educators across schools.	.756			
Standard 6 will benefit me as an educator.	.761			
Standard 6 will improve the quality of educators in K-12.	.603	.476		
Standard 6 makes education a stronger profession.	.911			
Standard 6 will lead to better teaching.	.940			
Standard 6 will lead to better student learning.	.940			
Standard 6 will lead to higher achievement test scores.	.786			
EVAAS/value-added increases competition amongst educators. (Transformed)			.825	
Educators will leave certain schools because of Standard 6. (Transformed)		.831	.313	
Educators will avoid working with certain students because of Standard 6. (Transformed)		.915		
Standard 6 will hinder me as an educator. (Transformed)				.691
Standard 6 will decrease teacher collaboration. (Transformed)			.849	
Standard 6 will make it harder to recruit people into the teaching profession. (Transformed)		.486		.606
Standard 6 will ultimately harm students. (Transformed)		.686		
Standard 6 makes me feel less valued as an educator. (Transformed)				.786

*Note.* Factor loadings < .3 are suppressed

## Appendix B

### Teacher Context

Table 11

*One-Way Analysis of Variance of Support/Opposition for the Use of Value-Added in Educator Evaluations by Respondent Race*

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>p</i>	$\eta_p^2$
Between groups	5	7.6	1.52	1.78	.12	6.13
Within group	136	116.3	0.86			
Total	141	123.9				

Table 12

*One-Way Analysis of Variance of Support/Opposition for the Use of Value-Added in Educator Evaluations by Respondent Sex*

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>p</i>	$\eta_p^2$
Between groups	3	1.9	0.63	0.70	.55	1.50
Within group	139	123.3	0.89			
Total	142	125.2				

Table 13

*One-Way Analysis of Variance of Support/Opposition for the Use of Value-Added in Educator Evaluations by Age Group*

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>p</i>	$\eta_p^2$
Between groups	7	6.5	0.93	1.06	.39	5.17
Within group	136	118.8	0.87			
Total	143	125.2				

Table 14

*Correlation between Support/Opposition for the Use of Value-Added in Educator Evaluations and Years of Teaching Experience*

	Years of Teaching Experience
Support/Opposition to use of value-added for educator evaluation	-.133

Note. \*  $p < .05$

## Appendix C

### Student Context

Table 15

*Results of t-tests and Descriptive Statistics, Support/Opposition for the Use of Value-Added in Educator Evaluations by Student Demographics*

	Respondents who Teach at Schools that Have this Characteristic			Respondents who Teach at Schools that do Not Have this Characteristic			95% CI for Mean Difference	Cohen's		
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>		<i>t</i>	<i>df</i>	<i>d</i>
>50% minority	1.72	0.94	68	1.94	0.91	82	-0.08, 0.52	1.44	148	0.24
>50% Free/Reduced lunch	1.81	0.91	84	1.88	0.95	66	-0.23, 0.37	0.45	148	0.07
>15% ESL [ELL]	1.76	0.82	62	1.90	1.00	88	-0.16, 0.44	0.91	148	0.15
High student mobility rate	1.74	0.94	42	1.88	0.92	108	-0.19, 0.48	0.84	148	0.14
>1/3 of students are facing significant health, emotional, and/or academic challenges	1.70	0.76	61	1.93	1.02	89	-0.08, 0.53	1.48	148	0.24

*Note.* \*  $p < .05$

## Appendix D

### School-Related Context

Table 16

*Results of t-tests and Descriptive Statistics, Support/Opposition for the Use of Value-Added in Educator Evaluations by Parent Involvement*

	Respondents who Teach at Schools that Have this Characteristic			Respondents who Teach at Schools that do Not Have this Characteristic			95% CI for Mean Difference	<i>t</i>	<i>df</i>	Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>				
Weak parent involvement	1.82	0.94	74	1.86	0.92	76	-0.27, 0.33	0.20	148	0.24

*Note.* \*  $p < .05$

Table 17

*One-Way Analysis of Variance of Support/Opposition for the Use of Value-Added in Educator Evaluations by School Setting (Rural, Suburban, Urban)*

Source	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>p</i>	$\eta_p^2$
Between groups	2	5.6	2.78	3.33	.04	4.34
Within group	147	122.6	0.83			
Total	149	128.2				

## About the Author

Kimberly Kappler Hewitt

University of North Carolina Greensboro

kkhewitt@uncg.edu

Dr. Kimberly Kappler Hewitt is an assistant professor of educational leadership at the University of North Carolina Greensboro. Her research interests focus on the ethical and efficacious use of educational data and sustainable reform. She is the editor of *Postcards from the Schoolhouse: Practitioner Scholars Examine Contemporary Issues in Educational Leadership* (NCPEA Press, 2013) and author of *Differentiation is an Expectation: A School Leader's Guide to Building a Culture of Differentiation* (Routledge, 2011).

orcid.org/0000-0002-5127-9625

## Acknowledgements

Funding to support this research came from a University of North Carolina Greensboro New Faculty Grant (#12363).

---

# education policy analysis archives

Volume 23 Number 76

August 10, 2015

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](mailto:fischman@asu.edu)

Join **EPAA's Facebook community** at <https://www.facebook.com/EPAAAPE> and **Twitter feed** @epaa\_aape.

---