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

education policy analysis archives

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

epaa | aape

Arizona State University

Volume 21 Number 5 January 28th, 2013 ISSN 1068-2341

The Legal Consequences of Mandating High Stakes Decisions Based on Low Quality Information: Teacher Evaluation in the Race-to-the-Top Era

Bruce D. Baker,
Rutgers, The State University of New Jersey

Joseph O. Oluwole,
Montclair State University

Preston C. Green,
The Pennsylvania State University
United States of America

Citation: Baker, B.D., Oluwole, J., Green, P.C. III (2013) The legal consequences of mandating high stakes decisions based on low quality information: Teacher evaluation in the race-to-the-top era. *Education Policy Analysis Archives, 21*(5). This article is part of EPAA/AAPE's Special Issue on *Value-Added: What America's Policymakers Need to Know and Understand*, Guest Edited by Dr. Audrey Amrein-Beardsley and Assistant Editors Dr. Clarin Collins, Dr. Sarah Polasky, and Ed Sloat. Retrieved [date], from http://epaa.asu.edu/ojs/article/view/1298

Abstract: In this article, we explain how overly prescriptive, rigid state statutory and regulatory policy frameworks regarding teacher evaluation, tenure and employment decisions outstrip the statistical reliability and validity of proposed measures of teaching effectiveness. We begin with a discussion of the emergence of highly prescriptive state legislation regarding the use of student testing data within teacher evaluation systems, specifically for purposes of making employment

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

Manuscript received: 7/1/2012
Revisions received: 12/1/2012
Accepted: 12/31/2012

decisions. Next, we explain the most problematic features of those policies, which include a) requirements that test-based measures constitute fixed, non-negotiable weight in final decisions, b) that test-based measures are used to place teachers into categories of effectiveness by applying numerical cutoffs beyond the precision or accuracy of the available data, and c) that professional judgment is removed from personnel decisions by legislating (or regulating) specific actions be taken when teachers fall into certain performance categories. In the subsequent section, we point out that different types of measures are being developed and implemented across states, and we explain that while value-added metrics in particular are, in fact designed to estimate a teacher's effect on student outcomes, descriptive growth percentile measures are not designed for making such inference and thus have no place in making determinations regarding teacher effectiveness. We also explain that, due to the properties of value-added estimates, they have no place in making high-stakes decisions based on rigid policy frameworks like those described herein. We evaluate the legal implications of rigid reliance on measures of teaching effectiveness that a) lack reliability and b) may be entirely invalid.
Keywords: High Stakes; Race to the Top; Value Added Models (VAM)

Las consecuencias jurídicas de imponer decisiones de consecuencias severas basadas en Información de baja calidad: Evaluación docente en la era de *"Carrera hacia la cima"*
Resumen: En este artículo, explicamos cómo marcos reglamentarios y legales altamente prescriptivos y rígidos en materia de políticas sobre evaluación de los docentes, su estabilidad y otras decisiones en materia de empleo superan la fiabilidad estadística y la validez de las medidas propuestas de efectividad de la enseñanza. Empezamos con una discusión de la aparición de lo que consideramos una legislación estatal excesivamente rígida con respecto a la utilización de resultados de exámenes de estudiantes dentro de sistemas de evaluación docente, específicamente para tomar decisiones sobre la estabilidad laboral. Luego, se explican las características más problemáticas de esas políticas, que pueden incluir: a) los requisitos que la prueba de las medidas se constituyen en cuestiones de peso fijas, no negociables en las decisiones finales, b) que los resultados de exámenes de estudiantes se usen para asignar a los docentes en categorías de eficacia más allá de la precisión o exactitud de los datos disponibles, y c) que criterios profesionales sean eliminados de los procesos de toma de decisiones relacionadas con la estabilidad laboral y reemplazados por legislaciones (o regulaciones) que imponen medidas concretas tomadas cuando los profesores son clasificados en determinadas categorías de rendimiento. En la siguiente sección, señalamos que existen diferentes tipos de medidas que se están desarrollando y aplicando en diferentes estados, y explicamos que, si bien medidas de valor agregado en particular están diseñados para estimar el efecto de un docente en los resultados de los estudiantes, las medidas alternativas no están diseñadas para hacer esa inferencia y, por ende, no tienen lugar para influir en las decisiones sobre la eficacia docente. También explicamos que, debido a las propiedades de las estimaciones de valor agregado, estas no deberían ser tomadas en cuenta en decisiones de consecuencias severas basadas en marcos rígidos de política educativa como los discutidos en este trabajo. Por último, evaluamos las consecuencias jurídicas de una dependencia rígida en medidas de eficacia de la enseñanza que a) carecen de fiabilidad y b) pueden ser enteramente inválidas.
Palabras clave: consecuencias severas; "Carrera hacia la cima;" modelos de valor agregado (MVA).

As consequências legais de impor decisões de consequências graves com base em informações de má qualidade: avaliação de professores na era da "carreira para a cima"
Resumo: Neste artigo, explicamos como marcos legais e políticos altamente prescritivos e rígidos usados para tomar decisões sobre emprego estabilidade, avaliação dos docentes excedem a

confiabilidade estatística e validade das medidas propostas sobre a eficácia ensino. Começamos com uma discussão sobre o surgimento do que consideramos uma legislação estadual muito rígida sobre o uso de notas dos alunos em provas no sistema de avaliação de professores, especificamente para tomar decisões sobre a estabilidade do emprego. Em seguida, explicamos as características mais problemáticas dessas políticas, que podem incluir: a) os requisitos que as medidas dos testes constituem questões de peso fixo, não negociável nas decisões finais, b) que os resultados dos testes de alunos sejam utilizados para assignar aos professores em categorias de eficácia além da precisão ou exatidão dos dados disponíveis e, c) que os padrões profissionais sejam retirados dos processos de tomada de decisões relacionadas à segurança no emprego e substituídos pela legislação (ou regulamentos) que imponha medidas específicas quando os professores são classificadas em certas categorias de desempenho. Na próxima seção, constatamos que existem diferentes tipos de medidas que estão sendo desenvolvidas e implementadas em diferentes estados, e explicamos que enquanto as medidas de valor adicionado em particular, são projetados para estimar o efeito de um professor nos resultados dos alunos, medidas alternativas não são projetados para efetuar essa inferência e, portanto, não deveriam influenciar decisões sobre a eficácia do professor. Também explicamos que, devido às propriedades das estimativas de valor adicionado, estes não devem ser levadas em conta nas decisões de consequências graves para a política educacional como os discutidos neste trabalho. Por fim, avaliamos as consequências jurídicas de uma dependência rígida em medidas de eficácia do ensino que: a) não são de confiança e, b) podem ser totalmente inválidas.
Palavras-chave: consequências graves; corrida para a cima; medidas de valor agregado.

Introduction

Spurred by the Race-to-the-Top program championed by the Obama administration and a changing political climate in favor of holding teachers accountable for the performance of their students, many states revamped their tenure laws and passed additional legislation designed to tie student performance to teacher evaluations. States have taken various approaches to these laws. Arizona, for example, uses a range approach for the weight given to student performance data in its teacher evaluations; specifically, the state requires that anywhere between 35% to 50% of teachers' evaluations must be based on student performance data (Arizona Revised Statutes Annotated § 15-203(A)(38) (2012)). Colorado, Florida and Idaho, on the other hand, require that student performance data, at a minimum, constitute 50% of teacher evaluations (Colorado Revised Statute § 22-9-106(1)(e)(II) (2010); Colorado Revised Statute § 22-9-105.5(2)(c)(1) (2010); Florida Statutes Annotated § 1012.34(3)(a)(1) (2011); Idaho Code § 33-514(4) (2012); Idaho Code § 33-515(2) (2012)). Unlike Florida, Idaho and Colorado, however, the District of Columbia Public Schools (DCPS), Ohio and Louisiana do not stipulate a minimum (District of Columbia Public Schools, 2011, p. 6; Ohio Revised Code Annotated § 3319.112(A)(1) (2011); Louisiana Revised Statute Annotated § 17:3902(B)(5) (2010)). Ohio Louisiana and DCPS set aside a fixed percentage (i.e. 50%) of their teacher evaluations for student performance data (District of Columbia Public Schools, 2011, p. 6; Ohio Revised Code Annotated § 3319.112(A)(1) (2011); Louisiana Revised Statute Annotated § 17:3902(B)(5) (2010)). Delaware's approach requires that, in teacher evaluations, student performance data must be "weighted at least as high as any other component" of the evaluation (14 Delaware Code § 1270(c) (2011)). States such as Maine, Maryland, Indiana, Oregon and Illinois simply provide that student performance data must be a "significant factor" in teacher evaluations ((20-A Maine Revised Statute Annotated § 13704(3)(A) (2015) (amended by L.D. 1858); Maryland Code, Education, § 6-202(c)(4)(i) (2010); Indiana Code § 20-28-11.5-4(4)(c)(2) (2012); Oregon Revised Statutes Annotated § 342.856 (2013); Oregon Administrative Rules Compilation

581-022-1723 (2013); 105 Illinois Compiled Statute Annotated 5/24A-5(c) (2011); 105 Illinois Compiled Statute Annotated 5/34-85c(a) (2011)). Utah merely requires that teacher evaluations must factor in evidence of student performance (Utah Administrative Rule 277-531-3(B)(3)(b) (2011); Utah Administrative Rule 277-531-3(C)(1)(b) (2011)). For more on the approaches of various states under the new teacher evaluation movement, see the state tables in the Appendix.

The desire to consequence teachers who fail to meet evaluation standards based on student performance data is a growing political movement that has in fact led to a brewing battle in New York. On June 21, 2012, New York state legislators passed a bill that would limit disclosure of teacher evaluation ratings to the public (Gormley, 2012). New York City Mayor Michael Bloomberg has threatened to circumvent the law by mandating city schools to call parents to disclose the information (Seifman, 2012). Rather than resort to such tactics as threatened by Mayor Bloomberg, in most cases, government officials have sought to consequence teachers for failing to meet evaluation standards by dismissing or terminating teachers. Tenured teachers present the greatest challenge because of laws that restrict their dismissal to specific grounds. For example, Pennsylvania's tenure law provides that once a teacher attains tenure, the teacher cannot be terminated except on any of the following grounds: incompetency; immorality; unsatisfactory performance over a specified time frame (two consecutive unsatisfactory evaluations in a span of at least four months); intemperance; willful neglect of duties; persistent negligence in performance; cruelty; documented mental/physical disability; felony conviction; persistent and willful failure to obey school laws; or participation or advocacy of un-American doctrines (24 Pennsylvania Statutes and Consolidated Statutes § 11-1122 (1996)).

While a number of states such as Pennsylvania have, for several years, had in their tenure statutes provisions for terminating or dismissing a tenured teacher for two consecutive unsatisfactory evaluations, which has become the norm in many states. States have argued that the flexibility to dismiss or terminate teachers whose evaluations do not meet standards based on student performance is necessary to ensure that only quality teachers are in the classrooms. In some states, the law mandates termination or dismissal of teachers who fail to meet evaluation standards; other states leave the decision about termination or dismissal up to the school district. Delaware, for example, provides discretion to districts to decide whether to terminate a teacher with two consecutive ineffective ratings (14 Delaware Code § 1273 (2006); 14 Delaware Code § 1411 (2006); 14 Delaware Code § 1420 (2006); 14 Delaware Code § 1270 (2011)). DCPS, on the other hand, mandates the termination of teachers who are rated minimally effective for two consecutive years (District of Columbia Public Schools, 2011, p. 62). While Florida gives districts discretion to decide whether a teacher with consecutive ratings should be terminated, this authority only applies to employees hired after July 1, 1984 (Florida Statutes Annotated § 1012.33(3) (2011)). Specifically, Florida allows districts to dismiss teachers with two consecutive unsatisfactory performance ratings or three consecutive ratings showing the teacher needs improvement (Florida Statutes Annotated § 1012.33(3)). Indiana's discretion for districts covers teachers with two consecutive ineffective ratings or teachers rated as needing improvement for three years over any five-year span (Indiana Code § 20-28-7.5-1(e)(4) (2011)). Michigan mandates the dismissal of teachers with three consecutive ineffective ratings (Michigan Compiled Laws § 380.1249(2)(h) (2011)). Colorado mandates returning a tenured teacher who has ineffective ratings for two years to probationary status (1 Colorado Administrative Code 301-87:3.0 (2012)). Louisiana does not even require waiting for two years; the state law provides that a tenured teacher rated ineffective must immediately be untenured (Louisiana Revised Statute Annotated § 17:442(C)(1)(2012)). These examples highlight what is at stake for teachers who fail to meet evaluation standards in their states based on student performance data. For other examples, see the table in the Appendix.

Our intent in this article is not to provide a thorough, systematic review of these policies. Rather, in this article, we seek to address what we consider to be prevalent structural problems with the current legislative models states have adopted. This article seeks to bring some urgency to the need to re-examine the current legislative models that put teachers at great risk of unfair evaluation, removal of tenure, and ultimately wrongful dismissal.

Structural Problems with Current Legislative Models

A relatively consistent legislative framework for teacher evaluation has evolved across states in the past few years, largely stimulated by explicit and implicit guidelines for states applying to receive a share of Federal Race to the Top funding (Learning Point Associates, 2010). Many of the risks of unfair treatment, giving rise to legal concerns, do so because of inflexible, arbitrary components of this legislative template. Based on cursory review of recently adopted policies, there appear to be three basic features of the standard model, each of which is problematic in its own regard, and those problems become multiplied when used in combination.

First, common evaluation models proposed in legislation require that *objective measures of student achievement growth* necessarily be considered in a weighting system of simultaneously considered elements. Student achievement growth measures are assigned, for example, a 40 or 50% weight alongside observation and other evaluation measures. Our review of state policies indicates more than 20 states (and the District of Columbia) have adopted a form of this policy component. Colorado requires that "A minimum of 50% of a teacher's evaluation must be based on the "academic growth of the teacher's students."[1] Less specific, Indiana requires "Objective measures of student achievement and growth" must "significantly inform" teacher evaluations.[2]

Placing the measures alongside one another in a weighting scheme assumes all measures in the scheme to be of equal validity and reliability but of varied importance (utility) – varied weight. Validity in this case means that the assigned values or statistical estimates in question measure what they claim to – the effect a teacher has on her students' achievement growth over the course of the year. Reliability in this case means that the measures in question are consistent over time and across tested content. Under common evaluation frameworks, each measure *must* be included, and must be assigned the prescribed weight – with no opportunity to question the validity of any measure. That is, the teacher effect estimate must be included in a teacher's final rating even if the evaluator has reason to believe that the estimate is influenced by some factor outside the teacher's control, or otherwise misrepresents the teacher's true effect.

Such a system also assumes that the various measures included in the system are each scaled such that they can vary to similar degrees. That is, that the observational evaluations will be scaled to produce similar variation to the student growth measures; and that the variance in both measures is equally valid – not compromised by random error or bias. In fact, however, it remains highly likely that some components of the teacher evaluation model will vary far more than others if by no other reasons than that some measures contain more random noise than others or that some of the variation is attributable to factors beyond the teachers' control. Regardless of the assigned weights and regardless of the cause of the variation (real or noise, that is, random variation) the measure that

---

[1] (Colorado Revised Statute § 22-9-106(1)(e)(II) (2010); Colorado Revised Statute § 22-9-105.5(2)(c)(1) (2010); 1 Colorado Administrative Code 301-87:3.0 (2012)). Similarly, for Florida, "The law requires that, at minimum, "50 percent of a performance evaluation must be based upon data and indicators of student learning growth assessed annually by statewide assessments or, for subjects and grade levels not measured by statewide assessments, by school district assessments" (Florida Statutes Annotated § 1012.34(3)(a)(1) (2011)).

[2] (Indiana Code § 20-28-11.5-4(4)(c)(2)(2012)

varies more will carry more weight in the final classification of the teacher as effective or not. In a system that places differential weights, but assumes equal validity across measures, even if the student achievement growth component is only a minority share of the weight, it may easily become the primary tipping point in most high stakes personnel decisions.

Second, the standard evaluation model proposed in legislation requires that teachers be placed into effectiveness categories by assigning arbitrary numerical cutoffs to the aggregated weighted evaluation components. That is, a teacher in the 25th percentile or lower when combining all evaluation components might be assigned a rating of "ineffective," whereas the teacher at the 26th percentile might be labeled effective. Furthermore, the teacher's placement into these groupings may largely if not entirely hinge on their rating in the student achievement growth component of their evaluation. Teachers on either side of the arbitrary cutoff are undoubtedly statistically no different from one another. In many cases as with the recently released teacher effectiveness estimates on New York City teachers, the error ranges for the teacher percentile ranks have been on the order of 35th percentile points (on average, up to 50% with one year of data). Assuming that there is any real difference between the teacher at the 25th percentile and 26th percentile (as their point estimate) is simply not justifiable in such statistical analysis, even where error ranges are much narrower. Placing an arbitrary, rigid, cut-off score into such noisy measures makes distinctions that simply cannot be justified especially when making high stakes employment decisions. Our review of state policies indicates more than twenty states and the District of Columbia have adopted a variation of this requirement – application of cut scores for the creation of performance categories. Indiana uses the following four categories: (i) Highly effective; (ii) Effective; (iii) Improvement Necessary; and (iv) Ineffective.

Third, it is not uncommon in recent legislation to place exact timelines on the conditions for removal of tenure. Recent legislation often dictates that teacher tenure either can or must be revoked after two consecutive years of being rated ineffective (where tenure can only be achieved after three consecutive years of being rate effective). As such, whether a teacher's true effect falls just below or just above the arbitrary cut-offs that define performance categories may have relatively inflexible consequences. Again, more than twenty states have adopted variations on this policy, either mandating that local districts take action on specific timelines, or encouraging or permitting such action on specified time lines. For Colorado, "A nonprobationary teacher who is rated as ineffective for two consecutive years shall lose nonprobationary status."[3] If the teacher fails to improve, he/she could be recommended for dismissal by the evaluator.[4]

Different Measures with Different Purposes

Two broad categories of methods and models have emerged in state policy regarding development and application of measures of student achievement growth to be used in newly adopted teacher evaluation systems. The first general category of methods is known as *value-added* models (VAMs) and the second as *student growth percentiles* (SGPs or MGPs, for "median growth percentile"). Several large urban school districts including New York City and Washington, DC have adopted *value-added* models and numerous states have adopted *student growth percentiles* for use in accountability systems. Among researchers it is well understood that these are substantively different measures by design, one being a possible component of the other. But these measures and their potential uses have been conflated by policymakers wishing to expedite implementation of new

[3] 1 Colorado Administrative Code 301-87:3.0 (2012)
[4] Colorado Revised Statutes Annotated § 22-9-106(4.5)(b) (2010)

teacher evaluation policies and pilot programs (Ehlert, Koedel, Parsons, & Podgursky, 2012; Goldhaber & Walch, 2012).

Arguably, one reason for the increasing popularity of the SGP approach across states is the extent of highly publicized scrutiny and large and growing body of empirical research over problems with using VAMs for determining teacher effectiveness (Baker, Darling-Hammond, Haertal, Ladd, Finn, Ravitch, Rothstein, Shavelson, & Shepard, 2010; Corcoran, 2010; Green, Baker, & Oluwole, 2012). Yet, there has been far less research on using student growth percentiles for determining teacher effectiveness. The reason for this vacuum is not that student growth percentiles are simply immune to problems of value-added models, but that researchers have until recently chosen not to evaluate their validity for this purpose – estimating teacher effectiveness – because they are not designed to infer teacher effectiveness.

Two recent working papers compare SGP and VAM estimates for teacher and school evaluation and both raise concerns about the face validity and statistical properties of SGPs. Goldhaber and Walch (2012) conclude "For the purpose of starting conversations about student achievement, SGPs might be a useful tool, but one might wish to use a different methodology for rewarding teacher performance or making high-stakes teacher selection decisions" (p. 30). Ehlert and colleagues (2012) note "Although SGPs are currently employed for this purpose by several states, we argue that they (a) cannot be used for causal inference (nor were they designed to be used as such) and (b) are the least successful of the three models [Student Growth Percentiles, One-Step & Two-Step VAM] in leveling the playing field across schools"(p. 23).

A value-added estimate uses assessment data in the context of a statistical model (regression analysis), where the objective is to estimate the extent to which a student having a specific teacher or attending a specific school influences that student's difference in score from the beginning of the year to the end of the year – or period of treatment (in school or with teacher). The most thorough of VAMs, more often used in research than practice, attempt to account for several prior year test scores on each student (to account for the extent that having a certain teacher alters a child's trajectory), the classroom level mix of student peers, individual student background characteristics, and possibly school level characteristics. The goal is to identify most accurately the share of the student's or group of students' value-added that should be attributed to the teacher as opposed to other factors outside of the teachers' control. Notably, important corrections such as using multiple years of prior student scores dramatically reduces the number of teachers who may be assigned ratings. For example, when Briggs and Domingue (2011) estimate alternative models to the LA Times (Los Angeles Unified School District) data using additional prior scores, the number of teachers rated drops from about 8,000 to only 3,300, because estimates can only be determined for teachers in grade 5 and above. [5] As such, these important corrections are rarely used in models to be applied for actual teacher evaluation.

By contrast, a student growth percentile is a descriptive measure of the relative change of a student's performance compared to that of all students. That is, the individual scores obtained on these underlying tests are used to construct an index of student growth, where the median student, for example, may serve as a baseline for comparison. Some students have achievement growth on the underlying tests that is greater than the median student, while others have growth from one test to the next that is less. That is, the approach estimates not how much the underlying scores changed, but how much the student moved within the mix of other students taking the same assessments, using a method called quantile regression to estimate the rarity that a child falls in her current position in the distribution, given her past position in the distribution (Briggs & Betebenner, 2009). Student growth percentile measures may be used to characterize each individual student's growth, or

[5] See Briggs & Domingue's (2011) re-analysis of LA Times estimates (pp. 10-12).

may be aggregated to the classroom level or school level, and/or across children who started at similar points in the distribution to attempt to characterize the collective growth of groups of students.

Many, if not most value-added models also involve normative rescaling of student achievement data, measuring in relative terms how much individual students or groups of students have moved within the large mix of students. The key difference is that the value-added models include other factors in an attempt to identify the extent to which having a specific teacher contributed to that growth, whereas student growth percentiles are simply a descriptive measure of the growth itself.

As described by the authors of the Colorado Growth Model:

> A primary purpose in the development of the Colorado Growth Model (Student Growth Percentiles/SGPs) was to distinguish the measure from the use: To separate the description of student progress (the SGP) from the attribution of responsibility for that progress. (Betebenner, Wenning, & Briggs, 2011)

Unlike value-added teacher effect estimates, student growth percentiles are not intended for attribution of responsibility for student progress to either the teacher or the school. But if this limitation is so clearly spelled out, is it plausible that states or local school districts will actually choose to use the measures to make inferences? Below is a brief explanation from a Question & Answer section of the New Jersey Department of Education web site regarding implementation of pilot teacher evaluation programs:

> Standardized test scores are not available for every subject or grade. For those that exist (Math and English Language Arts teachers of grades 4-8), Student Growth Percentages (SGPs), which require pre- and post-assessments, will be used. **The SGPs should account for 35%-45% of evaluations** [emphasis added]**.** The NJDOE (New Jersey Department of Education) will work with pilot districts to determine how student achievement will be measured in non-tested subjects and grades (NJDOE, 2012).

This explanation clearly indicates that student growth percentile data will be used for "evaluation" of teacher effectiveness. In fact, the SGPs alone, as they stand, as descriptive measures "should be used to account for 35% to 45% of evaluations." Other states including Colorado have already adopted (pioneered) the use of SGPs as a statewide accountability measure and have concurrently passed high stakes teacher evaluation legislation. But it remains to be seen how the SGP data will be used in district specific contexts in guiding high stakes decisions.[6]

SGPs can be hybridized with VAMs, by conditioning the descriptive student growth measure on student demographic characteristics. New York State has adopted such a model. However, the state's own technical report found "Despite the model conditioning on prior year test scores, schools and teachers with students who had higher prior year test scores, on average, had higher MGPs. Teachers of classes with higher percentages of economically disadvantaged students had lower MGPs" (American Institutes for Research, 2012, p. 1).

---

[6] In the Spring of 2011, The Colorado State Council for Educator Effectiveness released its report including guidelines for determining teacher effectiveness. This report hedged on causal interpretation of Student Growth Percentiles, identifying one standard of teacher effectiveness as follows: "Standard VI: Teachers take responsibility for student growth" (p. 12). http://www.cde.state.co.us/EducatorEffectiveness/downloads/Report%20&%20appendices/SCEE_Final_Report.pdf. As such, there remains some ambiguity as to how the Colorado Growth Model will actually play into district teacher evaluation frameworks.

Synthesizing the Similarities & Differences

As will be discussed at greater length in the next section, value-added models while intended to estimate teacher effects on student achievement growth, largely fail to do so in any accurate or precise way, whereas student growth percentiles make no such attempt.[7] Specifically, value-added measures tend to be highly unstable from year to year, and have very wide error ranges when applied to individual teachers, making confident distinctions between "good" and "bad" teachers difficult if not impossible (Baker et al., 2010; McCaffrey, Sass, Lockwood, & Mihaly, 2009; Sass, 2008; Schochet & Chiang, 2010). Furthermore, while value-added models attempt to isolate that portion of student achievement growth that is caused by having a specific teacher they often fail to do so and it is difficult if not impossible to discern a) how much the estimates have failed and b) in which direction for which teachers. That is, the individual teacher estimates may be biased by factors not fully addressed in the models and researchers have no clear way of knowing how much. We also know that when different tests are used for the same content, teachers receive widely varying ratings, raising additional questions about the validity of the measures (Corcoran, Jennings & Beveridge, 2010; Gates Foundation, 2010).

While we have substantially less information from existing research on student growth percentiles, it stands to reason that since they are based on the same types of testing data, they will be similarly susceptible to error and noise. But more troubling, since student growth percentiles make no attempt (by design) to consider other factors that contribute to student achievement growth, the measures have significant potential for omitted variables bias. SGPs leave the interpreter of the data to naively infer (by omission) that all growth among students in the classroom of a given teacher must be associated with that teacher. Research on VAMs indicates that even subtle changes to explanatory variables in value-added models change substantively the ratings of individual teachers (Ballou, Mokher, & Cavaluzzo, 2012; Briggs & Domingue, 2010). Omitting key variables can lead to bias and including them can reduce that bias. Excluding <u>all</u> potential explanatory variables, as do SGPs, takes this problem to the extreme by simply ignoring the possibility of omitted variables bias while omitting a plethora of widely used explanatory variables. As a result, it may turn out that SGP measures at the teacher level appear more stable from year to year than value-added estimates, but that stability may be entirely a function of teachers serving similar populations of students from year to year. The measures may contain stable omitted variables bias, and thus may be stable in their invalidity. Put bluntly, SGPs may be more consistent by being more consistently wrong.

In defense of Student Growth Percentiles as accountability measures, Betebenner, Wenning and Briggs (2011) explain that one school of thought is that value-added estimates are also most reasonably interpreted as descriptive measures, and should not be used to infer teacher or school effectiveness: "The development of the Student Growth Percentile methodology was guided by Rubin et al's (2004) admonition that VAM quantities are, at best, descriptive measures" (Betebenner, Wenning, & Briggs, 2011). Rubin, Stuart, and Zanutto (2004) explain:

> Value-added assessment is a complex issue, and we appreciate the efforts of Ballou et al. (2004), McCaffrey et al. (2004) and Tekwe et al. (2004). However, we do not think that their analyses are estimating causal quantities, except under extreme and

[7] Briggs and Betebenner (2009) explain: "However, there is an important philosophical difference between the two modeling approaches in that Betebenner (2008) has focused upon the use of SGPs as a descriptive tool to characterize growth at the student-level, while the LM (layered model) is typically the engine behind the teacher or school effects that get produced for inferential purposes in the EVAAS" (p. 30).

> unrealistic assumptions. We argue that models such as these should not be seen as estimating causal effects of teachers or schools, but rather as providing descriptive measures (Rubin et al., 2004, p. 18).

Arguably, these explanations do less to validate the usefulness of Student Growth Percentiles as accountability measures (inferring attribution and/or responsibility to schools and teachers) and far more to invalidate the usefulness of both Student Growth Percentiles and Value-Added Models for these purposes.

At the Intersection of Legal Claims and Statistical Models

In this section, we address the various legal challenges that might be brought by teachers dismissed under the rigid statutory structures outlined previously in this article. We also address how arguments on behalf of teachers might be framed differently in a context where value-added measures are used versus one where student growth percentiles are used. Where value-added measures are used, we suspect that teachers will have to show that while those measures were intended to attribute student achievement to their effectiveness, the measures failed to do so in a number of ways. That is, where value-added measures are used to assign effectiveness ratings, we suspect that the validity and reliability, as well as understandability of those measures would need to be deliberated at trial. However, where student growth percentiles are used, we would argue that the measures *on their face* are simply not designed for attributing responsibility to the teacher, and thus making such a leap would necessarily constitute a wrongful judgment. That is, one would not necessarily even have to vet the SGP measures for reliability or validity via any statistical analysis, because on their face they are invalid for this purpose.

As Green, Baker, and Oluwole (2012) explain, use of value-added measures in high stakes teacher dismissal cases raise a number of potential legal bases for the teachers to challenge the dismissal. This is especially true within the rigid, arbitrary legislative structures identified at the outset of this article. Specifically, there exists significant possibility that where arbitrary distinctions that cannot be made, are made, that the policies in question violate the due process rights of teachers (see also Harris, 2011; Giordano, 2012; Hill, Charalambous, & Kraft, 2012).

The Due Process Clause of the Fourteenth Amendment provides that no state shall "deprive any person of life, liberty, or property without due process of law" (United States Constitution Amendment XIV, § 1). To bring a Due Process challenge, public school teachers must first demonstrate that the state has deprived them of life liberty or property interest. Teachers might argue that the use of value-added estimates deprives them of a liberty interest by foreclosing their employment opportunities. Such claims may be unsuccessful because findings that teachers have failed to meet professional standards do not prevent them from finding employment elsewhere (Green, Baker, & Oluwole, 2012). On the other hand, teachers might be able to establish a property interest in continued employment based on their state's tenure statute (Green, Baker, & Oluwole, 2012).

Once teachers have established a protectable interest under the Due Process Clause, they may bring either a procedural or substantive due process challenge. Procedural due process "is a right to a fair procedure or set of procedures before one can be deprived of property by the state" (*Seal v. Morgan*, 2000, p. 574). It is more likely that teachers will challenge the technical shortcomings of value-added testing policies on substantive due process grounds. In the context of high school exit examinations, the Fifth Circuit established the following test: "When it encroaches upon concepts of justice lying at the basis of our civil and political institutions, the state is obligated to avoid action which is arbitrary and capricious, does not achieve or even frustrates a legitimate state

interest, or is fundamentally unfair" (*Debra P. v. Turlington*, 1981, p. 404). A Texas federal district court has established an alternative substantive due process analysis for high school exit tests: whether a state's educational determinations "reflect a substantial departure from accepted academic norms as to demonstrate that the person or committee responsible did not actually exercise professional judgment" (*G.I. Forum Image de Tejas v. Texas Education Agency*, 2000, p. 682, quoting *University of Michigan v. Ewing*, 1985, p. 225).

VAMs and SGPs may be vulnerable on both procedural and substantive due process grounds. The technical shortcomings of value-added estimates of teacher effectiveness may be broken down into questions of a) the reliability of those measures, and/or the precision with which they may be interpreted, b) the validity of those measures or the extent to which it may be validly inferred that the teacher had influence over the student outcomes, and c) the understandability of those measures to the teacher and whether the teacher has the ability to control his or her own fate.

Due Process, Rigid Structures & Noisy Measures (Reliability Concerns)

Reliability of measures of teaching effectiveness is critical for making high stakes decisions. It is rather unhelpful for example, if a teacher is rated highly on a given metric one year, and relatively low the next, and then high again the year after that. Such jumps in a performance measure would give most observers pause to think about whether that measure is really providing any useful information about the teacher's true ability. Such is the case in findings from most studies involving value-added measures (Baker et al., 2010; McCaffrey, Sass, Lockwood, Mihaly, 2009; Sass, 2008; Schochet & Chiang, 2010). This lack of reliability has been tested in several different ways:

- The correlation of the value-added measures across a group of teachers from one year to the next.
- The correlation within year across different sections of the same course taught by the same teacher.
- The standard errors around each teacher's predicted value.
- The classification error rates, given the standard errors.

In a value-added model, each teacher has a predicted value of the average achievement growth attributed to them, based on their students. But these predicted values aren't exact. They are estimates, given each teacher's sample of students and given the measures included in the regression model. Each teacher's predicted value has a confidence interval – typically reported as the range within which we can be 95% confident that the teacher's true value-added lies. There is greater likelihood that the teacher's true value-added lies closer to the predicted value than to the extremes of her confidence interval. In value-added models, these error ranges can be very large, meaning that one cannot reasonably distinguish between teachers with relatively similar predicted values.

A plethora of published analyses now raise serious concerns about the stability of teacher's value-added ratings from year to year. Among the earlier studies reporting this concern, the year-to-year correlations for a teacher's value-added rating were only about 0.2 or 0.3—at best a very modest correlation (McCaffrey, Sass, Lockwood, & Mihaly, 2009; Sass, 2008). Sass (2008) also notes that:

> About one quarter to one third of the teachers in the bottom and top quintiles stay in the same quintile from one year to the next while roughly 10 to 15 percent of teachers move all the way from the bottom quintile to the top and an equal proportion fall from the top quintile to the lowest quintile in the next year (Sass, 2008, p. 2).

Furthermore, most of the change or difference in the teacher's value-added rating from one year to the next is unexplainable—by differences in observed student characteristics, peer characteristics, or school characteristics (Sass, 2008). More recent studies have not yielded significant improvement in

year-to-year stability (Gates Foundation, 2010). Preliminary analyses from the MET Project, funded by the Bill and Melinda Gates Foundation, found that "[w]hen the between-section or between-year correlation in teacher value-added is below .5, the implication is that more than half of the observed variation is due to transitory effects rather than stable differences between teachers. That is the case for all of the measures of value-added we calculated" (Gates Foundation, 2010). Rothstein (2010) argues that the MET project findings actually overstated the relative stability in the ratings. Pointing to error ranges of the estimates, Rothstein explains:

> For example, even in the model for value-added on the state math test—the easiest to predict of the measures considered—a teacher whose predicted value-added is at the 25th percentile (that is, lower than 75% of her colleagues) has only about a one-third chance of actually being that far below average and about the same chance of in fact being *above* average. High-stakes decisions made based on predicted value-added will inevitably penalize a large number of teachers who are above average even when judged solely by the narrow metric of value-added for state tests. (p. 4)

While some statistical corrections and multi-year analysis might help, it is hard to guarantee or even be reasonably sure that a teacher would not be dismissed simply as a function of unexplainable low performance for two or three years in a row.

Table 1 provides a practical example drawn from a typical school within the New York City database on teacher value-added estimates released to the media earlier this year. The table includes four teachers from the same school, their predicted values and the upper and lower bounds of their confidence intervals for 2009-10 ratings. The table also includes the rating assigned to the teacher as a function of the strict cutoffs applied to the data. Teacher 1 has the lowest predicted value for math and teacher two for English Language Arts. In those cases, a below average rating is assigned. But it is clear that the confidence intervals are extremely large for these teachers, raising questions, for example, as to whether one can reasonably differentiate between the teacher who has an estimated effectiveness score at the 23rd percentile (Teacher 1, Math) and one at the 39th percentile (Teacher 3, Math). Confidence intervals may narrow for teachers with multi-year ratings, but only 2 of these 4 teachers had multi-year ratings. To begin with, only four of the forty-eight certified staff had value-added estimates to begin with, further questioning the value of these data.[8] In other words, what these data provide us are incredibly imprecise and inconsistent measures of supposed teacher effectiveness for only a tiny handful of teachers in a given school.

Finally, Schochet and Chang (2010), in a report for the U.S. Department of Education's Institute of Education Sciences, evaluated teacher value-added estimates in terms of classification error rates. They found that there is about a 25% chance (if using three years of data) or a 35% chance (if using one year of data) that a teacher who is "average" would be identified as "significantly worse than average" and potentially be fired. Of particular concern is the likelihood that a "good teacher" is falsely identified as a "bad" teacher—in this case a "false positive" identification. According to the study, this occurs one in ten times given three years of data and two in ten times given only one year of data.[9]

Classification errors are especially pertinent where rigid classification schemes are superimposed on these less-than-precise measures. It is difficult to imagine, for example, that a court could perceive as substantively fair, a system which may wrongly classify an average teacher as poor

---

[8] This figure was determined by comparing the number of teachers reported in the teacher effectiveness database (available at: http://www.ny1.com/content/top_stories/156599/now-available--2007-2010-nyc-teacher-performance-data#doereports) with the number of teachers reported in the statewide personnel master file for the same school (New York City school code 01M015.

[9] *Id.*

as much as 35% of the time. In short, the reliability and stability of these measures over time raises serious questions about their practical value at any level of human resource management and educational practice. But even more problematic is the integration of such unreliable and imprecise measures into rigid, high stakes statutory, regulatory and contractual evaluation models.

Table 1
Confidence Intervals for NYC Teacher Ratings in a Selected School

Teacher	Math				ELA			
	Low	Predicted Value	High	Rating	Low	Predicted Value	High	Rating
Tch1(5th)	3	23	68	Below Avg.	12	70	96	Average
Tch2(4th)	20	65	91	Average	0	11	58	Below Avg.
Tch3(4th)	5	39	80	Average	4	37	84	Average
Tch4(5th)	32	71	92	Average	13	68	93	Average

Source: Raw data downloaded on February 27, 2012 from http://www.ny1.com/content/top_stories/156599/now-available--2007-2010-nyc-teacher-performance-data#doereports

Anderson v. Banks (1981), a high school exit examination case, provides some insight as to how courts might analyze substantive due process challenges based on errors in measurement. In *Anderson*, a Georgia school district required candidates for high school graduation to achieve a specific score on the mathematics and reading sections of the California Achievement Test (CAT). Students had four opportunities in the ninth, tenth, eleventh, and twelfth grades to achieve the required scores. Students who were denied a diploma claimed that the CAT policy was not rationally related to the goal of improving education within the district because of the district's failure to account for the standard error of measurement. Specifically, the plaintiffs claimed that if the district had accounted for one standard error of measurement, at least eight out of 42 students who were denied diplomas in 1978 and 1979 would have graduated. The court rejected this claim because students could take the CAT multiple times, thus reducing the errors in measurement.

There is at least one important distinction between the high-stakes exit examination challenged in *Anderson* and teacher evaluation policies that employ value-added testing. In *Anderson*, students had to pass the exit examination in order to earn a diploma. By contrast, in teacher evaluation policies, student achievement scores are one of several components that states used in order to rate teachers. Still, courts might still use the approach adopted in *Anderson* where student achievement data comprise a major portion of the teacher evaluation policy. *Anderson* suggests that states that rely heavily on value-added teacher evaluation policies as grounds for removing tenured teachers may protect themselves from substantive due process challenges based on measurement errors by providing these teachers with multiple opportunities to satisfy the testing requirements. However, states such as Colorado, Florida, Oklahoma, and Tennessee that require student achievement to account for 50% of their teacher evaluation framework mandate the dismissal of teachers after two consecutive years of inadequate performance. Louisiana and Washington, DC appear to permit dismissal after one year of inadequate performance. Thus, the value-added models in these jurisdictions might be vulnerable to a substantive due process challenge for failing to sufficiently reduce errors in measurement.

Due Process & Attribution of Responsibility (Validity Concerns)

The recent release of New York City teacher value-added estimates to several media outlets stimulated much public discussion about standard errors and statistical noise.[10] But lost in that discussion was any emphasis on whether the predicted value-added measures were, to begin with, valid estimates of teacher effects. That is, did they actually represent what they were intended to represent - the teacher's influence on a true measure of student achievement, or learning growth while under that teacher's tutelage. As framed in teacher evaluation legislation, that measure is typically characterized as "student achievement growth," and it is assumed that one can measure the influence of the teacher on "student achievement growth" in a particular content domain.

A brief note on the semantics versus the statistics of evaluation and accountability is in order. At issue are policies involving teacher "evaluation" and more specifically *evaluation of teacher effectiveness*, where in cases of dismissal, the evaluation objective is to identify particularly ineffective teachers. In order to "evaluate" (assess, appraise, estimate) a teacher's effectiveness with respect to student growth, one must be able to "infer" (deduce, conjecture, surmise) that the teacher affected or could have affected that student growth. That is, for example, given one year's bad rating, the teacher had sufficient information to understand how to improve her rating in the following year. Furthermore, one must choose measures that provide some basis for such inference. Inference and *attribution* (ascription, credit, designation) are not separable when evaluating teacher effectiveness. To make an *inference* about teacher effectiveness based on student achievement growth, one must *attribute* responsibility for that growth to the teacher. In some cases, proponents of student growth percentiles alter their wording for general public appeal to argue that SGPs are a measure of student achievement growth, and that obviously student achievement growth is a primary objective of schooling. To that end, they argue that therefore, teachers and schools should obviously be held *accountable* for student achievement growth. Where *accountable* is a synonym for *responsible,* to the extent that SGPs were designed to separate the measurement of student growth from *attribution of responsibility for it,* then SGPs are also invalid on their face for holding teachers *accountable.* For a teacher to be *accountable* for that growth it must be *attributable* to them and one must be using a method that permits such *inference.*

We identify 3 categories of significant compromises to inference and attribution and therefore accountability for student achievement growth:

- The value-added estimate (or SGP) was influenced by something other than the teacher alone
- The value-added (or SGP) estimate given one assessment of the teacher's content domain produces a different rating than the value-added estimate given a different assessment tool
- The value-added estimate (or SGP) is compromised by missing data and/or student mobility, disrupting the *link* between teacher and students.

The first major issue compromising *attribution of responsibility* for or inference regarding teacher effectiveness based on student growth is that some other factor or set of factors actually caused the student achievement growth or lack thereof. A particularly bothersome feature of many value-added models is that they rely on annual testing data. That is, student achievement growth is measured from April or May in one year to April or May in the next, where the school year runs from September to mid or late June. As such, for example, the 4th grade teacher is assigned a rating based on children who attended her class from September to April (testing time), or about 7 months, where 2.5 months were spent doing any variety of other things, and another 2.5 months were spent with their prior grade teacher. Let alone the different access to resources each child has during their

[10] Local news stations convened panels to discuss the usefulness of the teacher ratings, including one on February 27, 2012 on New York's Fox 5 channel, with Sean Corcoran of NYU, Lisa Fleisher of the Wall Street Journal and Heather Brown of Fox 5 TV.

after school and weekend hours during the 7 months over which they have contact with their teacher of record (Lubienski & Crane, 2010).

Students with different access to summer and out-of-school time resources may not be randomly assigned across teachers within a given school or across schools within a district (Rothstein, 2009). And students who had prior year teachers who may have done more or less to advance linearly student achievement in core content areas during the post-testing month of the prior year may also not be randomly distributed. All of these factors go unobserved and unmeasured in the calculation of a teacher's effectiveness, potentially severely compromising the validity of a teacher's effectiveness estimate. Summer learning varies widely across students by economic backgrounds (Alexander, Entwisle, & Olsen, 2001) Furthermore, in the recent Gates MET Studies (2010), the authors found: "The norm sample results imply that students improve their reading comprehension scores just as much (or more) between April and October as between October and April in the following grade. Scores may be rising as kids mature and get more practice outside of school." (p. 8)

Numerous authors have conducted analyses revealing the problems of *omitted variables bias* and the non-random sorting of students across classrooms (Ballou, Mokher, & Cavaluzzo, 2012, Briggs & Domingue, 2011; Rothstein, 2009, 2010, 2011). In short, some value-added models are better than others, in that by including additional explanatory measures, the models seem to correct for at least some biases. Omitted variables bias is where any given teacher's predicted value is influenced partly by factors other than the teacher herself. That is, the estimate is higher or lower than it should be, because some other factor has influenced the estimate. Unfortunately, one can never really know if there are still additional factors that might be used to correct for that bias. Many such factors such as the individual or collective motivation of students in a given class or the influence of disruptive students are simply unobservable or at least unobserved in the available data. Other factors may be measurable and observable but are simply unavailable, or poorly measured in the data. Few if any data systems used for these purposes account for generally disruptive children and few if any data systems used for these purposes precisely parse differences in family income status and education, or even disability classification status (differentiating, for example, between mental retardation and speech impairment under the broad classification of "disability"). While there are some methods that can substantially reduce the influence of unobservables on teacher effect estimates, those methods can typically only be applied to a very small subset of teachers within very large data sets.[11] In a recent conference paper, Ballou and colleagues evaluated the role of omitted variables bias in value-added models and the potential effects on personnel decisions. They concluded:

> In this paper, we consider the impact of omitted variables on teachers' value-added estimates, and whether commonly used single-equation or two-stage estimates are preferable when possibly important covariates are not available for inclusion in the value-added model. The findings indicate that these modeling choices can significantly influence outcomes for individual teachers, particularly those in the tails of the performance distribution who are most likely to be targeted by high-stakes policies (Ballou, Mokher, & Cavaluzzo, 2012, p1).

[11] One approach is known as the student fixed effects specification which requires that each student who passes through each teacher for whom an effect is to be estimated has available multiple years of lagged test scores such that the model can estimate the extent to which any given teacher substantively changes the growth trajectory (within student slope) of students given their prior trajectory. See Briggs & Domingue (2010). Alternatively, but even more restrictive in terms of available sample, is the Chetty, Friedman and Rockoff (2011) bias test which involves evaluating the effectiveness estimates for teachers who move from one setting to another from year to year and across settings where student populations vary in terms of initial performance.

A related problem is the extent to which such biases may not present themselves in obvious patterns across the entire data set, but where specific circumstances or omitted variables may have rather severe effects on predicted values for specific teachers. To reiterate, these are not merely issues of instability or error. These are issues of whether the models are estimating the teacher's effect on students' outcomes, or the effect of something else on students' outcomes. Teachers should not be dismissed for factors beyond their control. Furthermore, statutes and regulations should not require that principals dismiss teachers or revoke their tenure in those cases where the principal understands intuitively that the teacher's rating was compromised by some other cause.

Other factors which severely compromise inference and attribution, and thus validity, include the fact that the measured value-added gains of a teacher's peers – or team members working with the same students – may be correlated, either because of unmeasured attributes of the students or because of spillover effects of working alongside more effective colleagues (one may never know) (Jackson & Bruegmann, 2009; Koedel, 2009,).

Significant evidence of bias existed in the value-added model estimated for the *Los Angeles Times* in 2010 (Felch, Song, & Smith, 2010), including significant patterns of racial disparities in teacher ratings both by the race of the student served and by the race of the teachers (see Green, Baker, & Oluwole, 2012). These model biases raise the possibility that Title VII racially disparate impact claims might also be filed by teachers dismissed on the basis of their value-added estimates, because the model was more likely to classify teachers of certain races as failing not because of their actual effectiveness but because of the students they were more likely to have served. Re-analysis of the LA Times data showed that some of these biases could be reduced by estimating a richer model, including additional prior student scores and additional demographic measures (Briggs & Domingue, 2010).[12]

A handful of studies have also found that teacher ratings vary significantly, even for the same subject area, if different assessments of that subject are used (Corcoran, Jennings, & Beveridge, 2010; Gates Foundation, 2010). If a teacher is broadly responsible for effectively teaching in their subject area, and not the specific content of any one test, different results from different tests raise additional validity concerns. Which test better represents the teacher's responsibilities? If more than one, in what proportions? If results from different tests completely counterbalance, how is one to determine the teacher's true effectiveness in their subject area? Using data on two different assessments used in Houston Independent School District, Corcoran, Jennings, and Beveridge (2010) find:

> [A]mong those who ranked in the top category (5) on the TAKS reading test, more than 17 percent ranked among the lowest two categories on the Stanford test. Similarly, more than 15 percent of the lowest value-added teachers on the TAKS were in the highest two categories on the Stanford. (as cited in Corcoran 2010, p. 17)

The Gates Foundation MET Project also evaluated consistency of teacher ratings produced on different assessments of mathematics achievement. In a review of the Gates findings, Rothstein (2010) explained:

> The data suggest that more than 20% of teachers in the bottom quarter of the state test math distribution (and more than 30% of those in the bottom quarter for ELA) are in the top half of the alternative assessment distribution (p. 5).

And:

[12] The original analysis conducted for the LA times is elaborated in a technical report by Buddin, 2010.

> In other words, teacher evaluations based on observed state test outcomes are only slightly better than coin tosses at identifying teachers whose students perform unusually well or badly on assessments of conceptual understanding. (p. 5)

Finally, student mobility, missing data, and algorithms for accounting for that missing data can severely compromise inferences regarding teacher effectiveness. Corcoran (2010) explains that the extent of missing data can be quite large and can vary by student type:

> Because of high rates of student mobility in this [Houston] population (in addition to test exemption and absenteeism), the percentage of students who have both a current and prior year test score – a prerequisite for value-added – is even lower…Among all grade four to six students in HISD, only 66 percent had both of these scores, a fraction that falls to 62 percent for Black students, 47 percent for ESL students, and 41 percent for recent immigrants. (Corcoran, 2010, pp. 20- 21)

Thus, many teacher effectiveness ratings would be based on significantly incomplete information, and further, the extent to which that information is incomplete would be highly dependent on the types of students served by the teacher.

One statistical resolution to this problem is imputation. In effect, imputation creates pre-test or post-test scores for those students who were not there. One approach is to use the average score for students who were there, or more precisely for otherwise similar students who were there. On its face, imputation is problematic when it comes to attribution of responsibility for student outcomes to the teacher, as some of those outcomes are statistically generated for students who were not even there (Raudenbush, 2004; Rubin, Stuart, & Zanutto, 2004). But not using imputation may lead to estimates of effectiveness that are severely biased, especially when there is substantial missing data. Howard Wainer (2011) in a video presentation in an event held at Educational Testing Services in Princeton, NJ explains somewhat mockingly how teachers might game imputation of missing data by sending all of their best students on a field trip during fall testing days, and then, in the name of fairness, sending the weakest students on a field trip during spring testing days.[13]

Clearly, in such a case of gaming, the predicted value-added assigned to the teacher as a function of the average scores of low performing students at the beginning of the year (while their high performing classmates were on their trip), and high performing ones at the end of the year (while their low performing classmates were on their trip), would not be correctly attributed to the teacher. The teacher might be responsible for her value-added estimate – in a perverse sense, but that does not by any stretch mean that the teacher is "effective."

To summarize, there are a multitude of potential threats to the validity of teacher effectiveness estimates, including non-random assignment, omitted variables bias, missing data problems, and variation in estimates arising from different tests of the same subject area. Each of these threats to validity raises due process concerns for teachers. The strong likelihood that teacher effect estimates are influenced by factors outside the teacher's control raises due process concerns where those estimates affect the teacher's property interests. While the courts have not addressed this question with respect to teachers and their students' achievement, courts have addressed this question with respect to the control individual students have over their own fate under high stakes testing regimes.

Two high school exit examination cases, *Debra P. v. Turlington* (1981, 1984) and *G.I. Forum Imaje de Tejas v. Texas Education Agency* (2000) provide some guidance as to how a court might analyze a substantive due process challenge by teachers based on the failure of a value-added model to account for matters that are outside the control of teachers. In *Debra P.*, minority students alleged that Florida's high school exit test requirement violated the Due Process Clause. A federal district

[13] http://www.njspotlight.com/ets_video2/ .

court agreed because the state failed to give students' sufficient notice before infringing upon their property right to obtain a diploma.

On appeal, the Fifth Circuit held that Florida's high-stakes test had to satisfy accepted standards of instructional validity: that is, whether the test measured what was actually taught in the state's schools. The court declared that the test would violate substantive due process if the test failed to cover material that was not covered in the students' classrooms. The court then remanded the case to determine whether the state had satisfied notions of curricular validity (*Debra P. v. Turlington*, 1981). On remand, the district court held that the test accomplished this goal. The court cited the state's efforts to provide remediation to students who could not master the material and a student survey, which found that 90-95% of students believed that they had been taught the test skills. The court rejected the plaintiffs' assertion that the state needed to focus on students who had failed to pass the high-stakes test in order to establish curricular validity. This was the case because the experts "conceded that there are no accepted educational standards for determining whether a test is [curricularly] valid" (*Debra P. v. Turlington*, 1984, p. 1412).

In the *G.I. Forum* case, minority students alleged that the state of Texas' high-stakes graduation test violated substantive due process. A federal district court rejected the students' challenge. First, the court held that the test satisfied accepted standards of curricular validity because "it measures what it purports to measure and it does so reliably" (*G.I. Forum Imaje de Tejas v. Texas Education Agency*, 2000, p. 682). The court held that the Texas high school exit test was not a substantial departure from accepted academic norms or a failure to use professional judgment. In reaching this conclusion, the court noted: "There was no testimony demonstrating that Texas has rejected current academic standards in designing its education system. Educators and test-designers testified that the design and the use of the test were within accepted norms" (pp. 682-83).

It is important to observe that in the *Debra P.* case, the Fifth Circuit observed that there were no accepted standards for determining whether the high school exit test satisfied curricular validity. Thus, it was easy for the state to establish the validity of the test (Green, Baker, & Oluwole, 2012). Also, in the *G.I. Forum* case, the court found no evidence that Texas' high school exit test fell outside academic norms. By contrast, it is impossible for value-added testing to sufficiently reduce the bias caused by factors outside of teachers' control to make such tests a valid measure of determining teacher effectiveness (see Green, Baker, & Oluwole, 2012 for a summary of multiple sources on this point). As the Economic Policy Institute explains: "[T]here is broad agreement among statisticians, psychometricians, and economists that student test scores alone are not sufficiently reliable and valid indicators of teacher effectiveness to be used in high-stakes personnel decisions, even when the most sophisticated statistical applications such as value-added modeling are employed" (Baker et al., 2010, p. 2).

Conclusions and Implications

As we have explained herein, value-added measures have severe limitations when attempting even to answer the narrow question of the extent to which a given teacher influences tested student outcomes. As such, we argue that it would be foolish to impose on these measures, rigid, overly precise high stakes decision frameworks. One simply cannot parse point estimates to place teachers into one category versus another and one cannot necessarily assume that any one individual teacher's estimate is necessarily valid (non-biased). Furthermore, we have explained how student growth percentile measures being adopted by states for use in teacher evaluation are, on their face, invalid for this particular purpose. Overly prescriptive, rigid teacher evaluation mandates, in our view, are likely to open the floodgates to new litigation over teacher due process rights. This is likely despite

the fact that much of the policy impetus behind these new evaluation systems is the reduction of legal hassles involved in terminating ineffective teachers.

Due process is violated where administrators or other decision-makers place blind faith in the quantitative measures, assuming them to be causal and valid (attributable to the teacher) and applying arbitrary and capricious cutoff-points to those measures (performance categories leading to dismissal). The problem, as we see it, is that some of these new state statutes require these due process violations, even where the informed, thoughtful professional understands full well that she is being forced to make a wrong decision. They require that decision makers take action based on these measures even against their own informed professional judgment.

This is not to suggest that any and all forms of student assessment data should be considered moot in thoughtful decision-making by school leaders and leadership teams. Rather, that incorrect, inappropriate use of this information is simply wrong – ethically and legally (a lower standard) wrong. We accept the proposition that tests of student knowledge and skills can provide useful insights both regarding what students know and potentially regarding what they have learned while attending a particular school or class. We are increasingly skeptical regarding the ability of value-added statistical models to parse any specific teacher's effect on those outcomes. Furthermore, the relative weight in management decision-making placed on any one measure depends on the quality of that measure and likely fluctuates over time and across settings. That is, in some cases, with some teachers and in some years, test data may provide leaders and/or peers with more useful insights. In other cases, it may be quite obvious to informed professionals that the signal provided by the data is simply wrong – not a valid representation of the teacher's effectiveness.

Arguably, a more reasonable and efficient use of these quantifiable metrics in human resource management might be to use them as a knowingly noisy pre-screening tool to identify where problems might exist across hundreds of classrooms in a large district. Value-added estimates might serve as a first step toward planning which classrooms to observe more frequently. Under such a model, when observations are completed, one might decide that the initial signal provided by the value-added estimate was simply wrong. One might also find that it produced useful insights regarding a teacher's (or group of teachers') effectiveness at helping students develop certain tested skills.

School leaders or leadership teams should clearly have the authority to make the case that a teacher is ineffective and that the teacher even if tenured should be dismissed on that basis. It may also be the case that the evidence would actually include data on student outcomes – growth, etc. The key, in our view, is that the leaders making the decision – indicated by their presentation of the evidence – would show that they have reasonably used information to make an informed management decision. Their reasonable interpretation of relevant information would constitute due process, as would their attempts to guide the teacher's improvement on measures over which the teacher actually had control.

References

1 Colorado Administrative Code 301-87:3.0 (2012).
14 Delaware Code § 1270 (2011).
14 Delaware Code § 1273 (2006).
14 Delaware Code § 1411 (2006).
14 Delaware Code § 1420 (2006).
20-A Maine Revised Statute Annotated § 13704(3)(A) (2015) (amended by L.D. 1858).
24 Pennsylvania Statutes and Consolidated Statutes § 11-1122 (1996).

105 Illinois Compiled Statute Annotated 5/24A-5(c) (2011).
105 Illinois Compiled Statute Annotated 5/34-85c(a) (2011).
Alexander, K.L, Entwisle, D.R., & Olsen, L.S. (2001). Schools, Achievement and Inequality: A Seasonal Perspective. *Educational Evaluation and Policy Analysis*, *23*(2), 171-91.
American Institutes for Research. (2012). *2011-12 growth model for educator evaluation technical report: Final.* November, 2012. New York State Education Department.
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). Retrieved [date], from http://epaa.asu.edu/ojs/article/view/1096.
Anderson v. Banks, 520 F. Supp. 472 (S.D.Ga. 1981).
Arizona Revised Statutes Annotated § 15-203(A)(38) (2012).
Ballou, D., Mokher, C.G., & Cavaluzzo, L. (2012, March). Using value-added assessment for personnel decisions: How omitted variables and model specification influence teachers' outcomes. Paper presented at the Annual Meeting of the Association for Education Finance and Policy. Boston, MA. Retrieved June 4, 2012, from http://aefpweb.org/sites/default/files/webform/AEFP-Using%20VAM%20for%20personnel%20decisions_02-29-12.docx.
Baker, E.L., Barton, P.E., Darling-Hammond, L., Haertel, E., Ladd, H.F., Linn, R.L., Ravitch, D., Rothstein, R., Shavelson, R.J., & Shepard, L.A. (2010). *Problems with the use of student test scores to evaluate teachers.* Washington, DC: Economic Policy Institute. Retrieved June 4, 2012, from http://epi.3cdn.net/724cd9a1eb91c40ff0_hwm6iij90.pdf.
Betebenner, D., Wenning, R.J., & Briggs, D.C. (2011). *Student growth percentiles and shoe leather.* Retrieved June 5, 2012, from http://www.ednewscolorado.org/2011/09/13/24400-student-growth-percentiles-and-shoe-leather.
Briggs, D. & Betebenner, D., (2009, April). Is student achievement scale dependent? Paper presented at the invited symposium Measuring and Evaluating Changes in Student Achievement: A Conversation about Technical and Conceptual Issues at the Annual Meeting of the National Council for Measurement in Education, San Diego, CA. Retrieved June 4, 2012, from http://dirwww.colorado.edu/education/faculty/derekbriggs/Docs/Briggs_Weeks_Is%20Growth%20in%20Student%20Achievement%20Scale%20Dependent.pdf.
Briggs, D. & Domingue, B. (2011). Due diligence and the evaluation of teachers: A review of the value-added analysis underlying the effectiveness rankings of Los Angeles Unified School District Teachers by the Los Angeles Times. Boulder, CO: National Education Policy Center. Retrieved June 4, 2012 from http://nepc.colorado.edu/publication/due-diligence.
Buddin, R. (2010, August). How effective are Los Angeles elementary teachers and schools? Retrieved June 4, 2012, from http://www.latimes.com/media/acrobat/2010-08/55538493.pdf.
Chetty, R., Friedman, J., & Rockoff, J. (2011). The long term impacts of teachers: Teacher value added and student outcomes in adulthood. NBER Working Paper # 17699. Retrieved June 4, 2012, from http://www.nber.org/papers/w17699.
Colorado Revised Statute § 22-9-105.5(2)(c)(1) (2010).
Colorado Revised Statute § 22-9-106(1)(e)(II) (2010).
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. Annenberg

Institute for School Reform. Retrieved June 4, 2012, from http://annenberginstitute.org/pdf/valueaddedreport.pdf.

Corcoran, S.P., Jennings, J.L., & Beveridge, A.A. (2010). Teacher effectiveness on high- and low-stakes tests. Paper presented at the Institute for Research on Poverty Summer Workshop, Madison, WI.

Debra P. v. Turlington, 644 F. 2d 397 (5th Cir. 1981).

Debra P. v. Turlington, 730 F.2d 1405 (11th Cir. 1984).

District of Columbia Public Schools (2011). The District of Columbia public schools effectiveness assessment system for school-based personnel: Group 1 general education teachers with individual value-added student achievement data. Washington, DC. Retrieved June 4, 2012, from, http://dcps.dc.gov/DCPS/Files/downloads/TEACHING%20&%20LEARNING/IMPACT/IMPACT%20Guidebooks%202010-2011/Impact%202011%20Group%201-Aug11.pdf

Education Trust (2011). Fact sheet - teacher quality. Washington, DC. Retrieved June 4, 2012, from http://www.edtrust.org/sites/edtrust.org/files/Ed%20Trust%20Facts%20on%20Teacher%20Equity_0.pdf .

Ehlert, M., Koedel, C., &Parsons, E., & Podgursky, M. (2012). Selecting growth measures for school and teacher evaluations. National Center for Analysis of Longitudinal Data in Education Research (CALDAR). Working Paper #80.

Felch, J., Song, J., & Smith, D. (2010, August 14). Who's teaching L.A.'s kids? *Los Angeles Times*. Retrieved February 2, 2011, from http://www.latimes.com/news/local/la-me-teachers-value-20100815,0,2695044.story.

Florida Statutes Annotated § 1012.33(3) (2011).

Florida Statutes Annotated § 1012.34(3)(2011).

Gates Foundation (2010). Learning about teaching: Initial findings from the measures of effective teaching project. MET Project Research Paper. Seattle, Washington: Bill & Melinda Gates Foundation. Retrieved December 16, 2010, from http://www.metproject.org/downloads/Preliminary_Findings-Research_Paper.pdf.

G.I. Forum Image de Tejas v. Texas Educ. Agency, 87 F.Supp.2d 667 (W.D. Tex. 2000).

Giordano, M.A. (2012, March 13). Fallout continues over teacher rankings. *The New York Times – School Book*. Retrieved from http://www.nytimes.com/schoolbook/2012/03/13/fallout-continues-over-teacher-rankings/.

Goldhaber, D. (2010). *When the stakes are high, can we rely on value added?* Washington, DC. Center for American Progress. Retrieved June 4, 2012, from http://www.americanprogress.org/wp-content/uploads/issues/2010/12/pdf/vam.pdf.

Goldhaber, D., & Walch, J. (2012). Does the model matter? Exploring the relationship between different student achievement-based teacher assessments. University of Washington at Bothell, Center for Education Data & Research. CEDR Working Paper 2012-6.

Gormley, M (2012, June 22). NY legislature passes teacher eval disclosure bill. Retrieved June 4, 2012, from http://news.yahoo.com/ny-legislature-passes-teacher-eval-disclosure-bill-154158688--finance.html.

Harris, D.N. (2011). *Value-added measures in education: What every educator needs to know*. Cambridge, MA: Harvard Education Press.

Hill, H.C., Charalambous, C.Y., & Kraft, M.A. (2012). When rater reliability is not enough: Teacher observation systems and a case for the generalizability study. *Educational Researcher, 41*(56). DOI: 10.3102/0013189X12437203.

Idaho Code § 33-514(4) (2012).

Idaho Code § 33-515(2) (2012).
Indiana Code § 20-28-7.5-1(e)(4) (2011).
Indiana Code § 20-28-11.5-4(4)(c)(2) (2012).
Jackson, C. &Bruegmann, E. (2009). Teaching students and teaching each other: The importance of peer learning for teachers. *American Economic Journal: Applied Economics 1*(4), 85–108.
Koedel, C. (2009). An empirical analysis of teacher spillover effects in secondary school. *Economic of Education Review, 28*(6) 682-92.
Koedel, C., & Betts, J. R. (2009). Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the rothstein critique. Working paper.
Learning Point Associates. (2010). Evaluating teacher effectiveness: Emerging trends reflected in the state phase 1 race to the top applications. Accessed December 18, 2012 from http://www.learningpt.org/pdfs/RttT_Teacher_Evaluation.pdf
Louisiana Revised Statute Annotated § 17:3902(B)(5) (2010).
Louisiana Revised Statute Annotated § 17:442(C)(1)(2012).
Lubienski, S. T., & Crane, C. C. (2010) Beyond free lunch: Which family background measures matter? *Education Policy Analysis Archives, 18*(11). Retrieved January 22, 2013 from http://epaa.asu.edu/ojs/article/view/756.
Maryland Code, Education, § 6-202(c)(4)(i) (2010).
McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics, 29*(1), 67 - 101.
McCaffrey, D.F., Sass, T.R., Lockwood, J.R., & Mihaly, K. (2009). The intertemporal variability of teacher effect estimates. *Education Finance and Policy, 4*,(4) 572-606.
McCaffrey, D.F., & Lockwood, J.R. (2011). Missing data in value added modeling of teacher effects. *Annals of Applied Statistics, 5*(2A) 773-97.
Michigan Compiled Laws § 380.1249(2)(h) (2011).
New Jersey Department of Education. Excellent Educators for New Jersey. Retrieved from http://www.state.nj.us/education/EE4NJ/faq/ accessed, May 20, 2012.
Ohio Revised Code Annotated § 3319.112(A)(1) (2011).
Oregon Administrative Rules Compilation 581-022-1723 (2013).
Oregon Revised Statutes Annotated § 342.856 (2013).
Raudenbush, S.W. (2004). What are value-added models estimating and what does this imply for statistical practice? *Journal of Educational and Behavioral Statistics, 29*(1), 121-129. DOI:10.3102/10769986029001121.
Rothstein, J. (2011). Review of "Learning about teaching: Initial findings from the measures of effective teaching project." Boulder, CO: National Education Policy Center. Retrieved June 4, 2011, from http://nepc.colorado.edu/thinktank/review-learning-about-teaching.
Rothstein, J. (2009). Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Education Finance and Policy, 4*(4), 537–71.
Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *Quarterly Journal of Economics*, *125*(1), 175–214.
Rubin, D. B., Stuart, E. A., & Zanutto, E. L. (2004). A potential outcomes view of value-added assessment in education. *Journal of Educational and Behavioral Statistics*, *29*(1), 103–16.
Sass, T.R. (2008). The stability of value-added measures of teacher quality and implications for teacher compensation policy. Retrieved June 4, 2012, from http://www.urban.org/UploadedPDF/1001266_stabilityofvalue.pdf.

Schochet, P.Z. & Chiang, H.S. (2010). Error rates in measuring teacher and school performance based on student test score gains. Institute for Education Sciences, U.S. Department of Education. Retrieved May 14, 2012, from http://ies.ed.gov/ncee/pubs/20104004/pdf/20104004.pdf.

Seifman, D. (2012, June 23). Calling all parents: Mike: We'll get teacher data out anyway! *NY Post.* Retrieved June 24, 2012, from http://www.nypost.com/p/news/local/calling_all_parents_Wo340zWw9Tfyk3bTrUOU2K#ixzz1yjp82gg0.

Utah Administrative Rule 277-531-3(B)(3)(b) (2011).

Utah Administrative Rule 277-531-3(C)(1)(b) (2011).

Working with Teachers to Develop Fair and Reliable Measures of Effective Teaching. (2010). MET Project White Paper. Seattle, Washington: Bill & Melinda Gates Foundation, 1. Retrieved December 16, 2010, from http://www.metproject.org/downloads/met-framing-paper.pdf.

Appendix

State Approaches to the New Teacher Evaluation Movement

The tables below set forth the approaches of various states to the new teacher evaluation movement. Specifically, we set forth the quantitative weight states assign student achievement in their teacher evaluations. We also specify the classifications states use to rate teacher performance under their evaluations. The table also identifies the timelines (if any) provided in state law or policy for dismissing tenured teachers rated ineffective under the state's evaluation system.

Table A1.
Arizona's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Arizona	The model framework for teacher evaluations created by the state board of education must include "quantitative data on student academic progress that accounts for between thirty-three per cent and fifty per cent of the evaluation outcomes" (Arizona Revised Statutes Annotated § 15-203(A)(38) (2012)) (as amended by House Bill 2823. (2012). Retrieved June 1, 2012, from Arizona Legislature Web Site: http://www.azleg.gov/legtext/50leg/2r/laws/0259.pdf)	None[1]	(i) Highly effective; (ii) Effective; (iii) Developing; and (iv) Ineffective (Arizona Revised Statutes Annotated § 15-203(A)(38) (2012)).
		Teachers in the lowest performance classification are offered an intervention only once (Arizona Revised Statute § 15-537(C)(1)-(3) (2012)): Before being dismissed for inadequate performance, the teacher must be given a minimum of sixty instructional days to rectify the inadequate performance; if the teacher fails to show adequate performance within this time frame, the district is required to dismiss the teacher (Arizona Revised Statute § 15-539 (2012)). The notice of inadequate performance must be initiated not later than the teacher's **second** consecutive year in the lowest performance (Arizona Revised Statute § 15-537(C)(4) (2012)).	

[1] In the table, "none" refers to cases where there is either no tenure in the state or where the tenure provision includes no specified timeline for how soon after an ineffective rating a teacher should be dismissed. Note, however, that in Arizona, tenured teachers can be dismissed for inadequate performance (Arizona Revised Statute § 15-539(C) (2012)). The definition of inadequate performance is based on the state's performance classifications for teachers. (Arizona Revised Statute § 15-539(D) (2012)). In Alaska, which currently does not require quantified student achievement as a significant component of evaluations, a tenured teacher who fails to meet district performance standards is provided a plan of improvement. Unless the teacher and the evaluating administrator agree to an extension, the improvement plan must be in effect for at least 90 workdays and at most 180 workdays. During this time, the teacher must be observed at least twice. If the teacher still fails to meet the district performance standards by the end of the term of the improvement plan, the district has the discretion to nonretain the teacher. (Alaska Statute § 14.20.149(e) (2009); Alaska Statute § 14.20.175(b)(1) (2008)). Georgia uses an annual contract for its teachers (Georgia Code Annotated § 20-2-211 (2011); Georgia Code Annotated § 20-2-940 (2013)). In New Hampshire, the law provides that "the grounds for nonrenomination and nonreelection shall be determined at the sole discretion of the school board" (New Hampshire Revised Statute § 189:14-a (2011)).

Table A2.
Arkansas' Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Arkansas	No	No[1]	(i) Distinguished; (ii) Proficient; (iii) Basic; and (iv) Unsatisfactory (Arkansas Code Annotated § 6-17-2805 (2011)).

[1] The Teacher Fair Dismissal Act of 1983 specifically states that Arkansas law does not provide teachers tenure because the law "does not confer lifetime appointment of teachers" (Arkansas Code Annotated 6-17-1503(b) (2005)).

Table A3.
Connecticut's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Connecticut	None	A district is authorized to terminate a tenured teacher at "any time" for incompetency or inefficiency because of the teacher's evaluation based on student academic growth (Connecticut General Statute § 10-151(d)(1)(2011); Connecticut General Statute § 10-151b (2011)).	N/A[1]

[1] N/A = Not Applicable.

Table A4.
Colorado's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Colorado	A minimum of **50%** of a teacher's evaluation must be based on the "academic growth of the teacher's students" (Colorado Revised Statute § 22-9-106(1)(e)(II) (2010); Colorado Revised Statute § 22-9-105.5(2)(c)(1) (2010); 1 Colorado Administrative Code 301-87:3.0 (2012)).	"A nonprobationary teacher who is rated as ineffective for **two** consecutive years shall lose nonprobationary status." (1 Colorado Administrative Code 301-87:3.0 (2012)). If the teacher fails to improve, he/she could be recommended for dismissal by the evaluator (Colorado Revised Statutes Annotated § 22-9-106(4.5)(b) (2010)).	(i) Ineffective; (ii) Partially effective; (iii) Effective; and (iv) Highly effective (1 Colorado Administrative Code 301-87:3.0(3.03) (2012)).

Table A5.
Delaware's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Delaware	The state's evaluation system known as the Delaware Performance Appraisal System II (DPAS II) "must have no more than 5 components and must have a strong focus on student improvement, with 1 component dedicated exclusively to student improvement and weigh	Whenever a teacher is deemed to have a pattern of ineffective teaching based on the state's evaluation system, the district has the discretion of terminating the teacher based on incompetency (14 Delaware Code § 1273 (2006); 14 Delaware Code § 1411 (2006); 14 Delaware Code § 1420 (2006); 14 Delaware Code § 1270 (2011)). "If a teacher's overall Summative Evaluation rating is determined to be 'Needs Improvement' for the third consecutive year, the teacher's rating shall be re-categorized as 'Ineffective'" (14 Delaware Administrative Code 106A(6.2.5) (2011) Teacher Appraisal Process Delaware Performance Appraisal System (DPAS II) Revised. (2011, December 1). Retrieved June 1, 2012, from Delaware Administrative Code Web Site: http://regulations.delaware.gov/AdminCode/title14/100/106A.pdf). The law considers **two** consecutive ratings of 'Ineffective' as a pattern of ineffective teaching (14 Delaware Administrative Code 106A(7.1) (2011) Teacher Appraisal Process Delaware Performance Appraisal System (DPAS II) Revised. (2011, December 1). Retrieved June 1, 2012, from Delaware Administrative Code Web Site: http://regulations.delaware.gov/AdminCode/title14/100/106A.pdf)	The rating categories for each component of a teacher's evaluation are: (i) Satisfactory; (ii) Unsatisfactory (14 Delaware Code § 1270(b) (2011)). For the overall rating of the teacher's performance, the categories are: (i) Highly Effective; (ii) Effective; (iii) Needs Improvement;

Table A5. (Cont.'d)
Delaware's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Delaware			(iv) Ineffective (14 Delaware Administrative Code 106A(6.0) (2011) Teacher Appraisal Process Delaware Performance Appraisal System (DPAS II) Revised. (2011, December 1). Retrieved June 1, 2012, from Delaware Administrative Code Web Site: http://regulations.delaware.gov/AdminCode/title14/100/106A.pdf). A satisfactory evaluation is equivalent to the "overall 'Highly Effective', 'Effective' or 'Needs Improvement' rating on the summative evaluation and shall be used to qualify for a continuing license" (14 Delaware Administrative Code 106A(2.0) (2011) Teacher Appraisal Process Delaware Performance Appraisal System (DPAS II) Revised. (2011, December 1). Retrieved June 1, 2012, from Delaware Administrative Code Web Site: http://regulations.delaware.gov/AdminCode/title14/100/106A.pdf; DPAS II Guide for Teachers . (2011, September 1). Retrieved June 1, 2012, from Delaware Performance Appraisal System Web Site: http://www.doe.k12.de.us/csa/dpasii/ti/DPASIITeacherFullGuide-9-7-11.pdf).

Table A6.
District of Columbia's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
District of Columbia Public Schools (DCPS)	Under IMPACT – the DCPS evaluation system – student achievement data accounts for 50% of teacher evaluations (Group 1 General Education Teachers with Individual Value-Added Student Achievement Data 6. (2011, August). Retrieved June 1, 2012, from The District of Columbia Public Schools Effectiveness Assessment System for School-Based Personnel Web Site: http://dcps.dc.gov/DCPS/Files/downloads/TEACHING%20&%20LEARNING/IMPACT/IMPACT%20Guidebooks%202010-2011/Impact%202011%20Group%201-Aug11.pdf).	Teachers who are rated "'Minimally Effective' for two consecutive years will be subject to separation from the school system" (Group 1 General Education Teachers with Individual Value-Added Student Achievement Data 62. (2011, August). Retrieved June 1, 2012, from The District of Columbia Public Schools Effectiveness Assessment System for School-Based Personnel Web Site: http://dcps.dc.gov/DCPS/Files/downloads/TEACHING%20&%20LEARNING/IMPACT/IMPACT%20Guidebooks%202010-2011/Impact%202011%20Group%201-Aug11.pdf).	(i) Highly Effective; (ii) Effective; (iii) Minimally Effective; or (iv) Ineffective (What Are the Final IMPACT Ratings? (2011). Retrieved June 1, 2012, from District of Columbia Public Schools, An Overview of IMPACT Web Site: http://dcps.dc.gov/DCPS/In+the+Classroom/Ensuring+Teacher+Success/IMPACT+%28Performance+Assessment%29/An+Overview+of+IMPACT).
		For teachers who are rated 'Ineffective', this is an unacceptable performance. Consequently, the two-consecutive-years rule applicable to teachers rated 'Minimally Effective' does not apply; rather teachers rated Ineffective "will be subject to separation from the school system" (Group 1 General Education Teachers with Individual Value-Added Student Achievement Data 62. (2011, August). Retrieved June 1, 2012, from The District of Columbia Public Schools Effectiveness Assessment System for School-Based Personnel Web Site: http://dcps.dc.gov/DCPS/Files/downloads/TEACHING%20&%20LEARNING/IMPACT/IMPACT%20Guidebooks%202010-2011/Impact%202011%20Group%201-Aug11.pdf).	

Table A7.
Florida's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Florida	The law requires that, at minimum, "**50** percent of a performance evaluation must be based upon data and indicators of student learning growth assessed annually by statewide assessments or, for subjects and grade levels not measured by statewide assessments, by school district assessments"[1] (Florida Statutes Annotated § 1012.34(3)(a)(1) (2011)).[2]	Teachers who got continuing contract status before July 1, 1984 will keep that status unless the teacher: (i) willingly gives up the continuing contract status; or (ii) is dismissed on grounds such as incompetency; or (iii) is returned to annual contracts for three years at the discretion of the district for "good and sufficient reasons" (Florida Statutes Annotated § 1012.33(4)(2011)).[3] Teachers employed after July 1, 1984 have a professional service contract which must be renewed annually unless the district chooses to dismiss the teacher who: (i) is charged with unsatisfactory performance; or (ii) has "**two** consecutive annual performance evaluation ratings of unsatisfactory"; or (iii) has "**two** annual performance evaluation ratings of unsatisfactory within a 3-year period"; or (iv) has "**three** consecutive annual evaluation ratings of needs improvement or a combination of needs improvement and unsatisfactory" (Florida Statutes Annotated § 1012.33(3)(2011)).	(i) Highly Effective; (ii) Effective; (iii) Needs Improvement;[4] and (iv) Unsatisfactory (Florida Statutes Annotated § 1012.34(2)(e) (2011)).

[1] School districts granted an exemption pursuant to Florida's Race to the Top Memorandum of Understanding for Phase 2 can use 40% instead of 50% (Florida Statutes Annotated § 1012.341 (2011)).
[2] Additionally, "the student learning growth portion of the evaluation must include growth data for students assigned to the teacher over the course of at least 3 years.
If less than 3 years of data are available, the years for which data are available must be used and the percentage of the evaluation based upon student learning growth may be reduced to not less than 40 percent" Florida Statutes Annotated § 1012.34(3)(a)(1)(a) (2011).
[3] Beginning in July 1, 2011, all new teachers hired in Florida are on annual contracts (Florida Statutes Annotated § 1012.335(2)(2011)). These teachers can be dismissed on various grounds including incompetency (Florida Statutes Annotated § 1012.335(5)(c)(2011)).
[4] For those "instructional personnel in the first 3 years of employment who need improvement" the term used is "developing" instead of "needs improvement" (Florida Statutes Annotated § 1012.34(2)(e)(3) (2011)).

Table A8.
Idaho's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Idaho	Teachers hired after January 31, 2011 operate under two different contract categories: contract A or contract B (Idaho Code § 33-514 (2012)).[1] Unless in a case of reduction in force, if the district decides not to reemploy a category A contract teacher or a category B contract teacher, the decision must be made after an evaluation of the teacher. (Idaho Code § 33-514(2) (2012)). "The objective measure(s) of growth in student achievement shall comprise at least fifty percent (50%) of the total written evaluation" (Idaho Code § 33-514(4) (2012)). This same 50% rule applies to teachers who had acquired tenure status prior to January 31, 2012 (Idaho Code § 33-515(2) (2012)).[2]	None However, before a school district chooses to non-renew teachers with grandfathered renewable contracts, Idaho law entitles such teachers to "a defined period of probation as established by the board, following an observation, evaluation or partial evaluation" (Idaho Code § 33-515(5) (2012)). The length of the probation is not specified.	The suggested categories for districts to use are: (i) Unsatisfactory; (ii) Basic; (iii) Proficient; (iv) Distinguished (Idaho State Department of Education, (2009). Implementation Guidelines. Retrieved May 25, 2012, from http://www.sde.idaho.gov/site/teacherEval/implementationGuidelines.htm).

Table A8. (Cont.'d)
Idaho's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Idaho			The evaluation performance categories used by a district must "at a minimum, address proficient and unsatisfactory practice" (Idaho State Department of Education, (2009). Implementation Guidelines. retrieved May 25, 2012, from http://www.sde.idaho.gov/site/teacherEval/implementationGuidelines.htm).

[1] The category A contract is defined as "a limited one (1) year contract for certificated personnel in the first or greater years of continuous employment with the same school district" (Idaho Code § 33-514(2)(a) (2012)).
The category B contract is defined as "a limited two (2) year contract that may be offered at the sole discretion of the board of trustees for certificated personnel in their fourth or greater year of continuous employment with the same school district" (Idaho Code § 33-514(2)(b) (2012)). Additionally, "[t]he board of trustees may, at its sole discretion, add an additional year to such a contract upon the expiration of the first year, resulting in a new two (2) year contract" (Idaho Code § 33-514(2)(b) (2012)).
[2] Idaho law no longer provides for "vesting of tenure, continued expectations of employment or property rights in an employment relationship" (Idaho Code § 33-515(1) (2012)). Instead, teachers who had tenure rights prior to January 31, 2011 shall operate under grandfathered renewable contracts with "the right to the continued automatic renewal of that employee's employment contract by giving notice, in writing, of acceptance of renewal" (Idaho Code § 33-515(2) (2012)). These automatic renewals could be "for a shorter term, longer term or the same length of term as the length of term stated in the current contract, and at a greater, lesser or equal salary to that stated in the current contract" (Idaho Code § 33-515(2)-(3) (2012)).

Table A9.
Illinois' Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Illinois	Student performance data must be a "significant" factor in teacher evaluations (105 Illinois Compiled Statute Annotated 5/24A-5(c) (2011); (105 Illinois Compiled Statute Annotated 5/34-85c(a) (2011)).	If a teacher is found to have unsatisfactory performance consequent to an evaluation of the teacher, the district could choose to dismiss the teacher for failure to "complete a remediation plan with a rating equal to or better than a 'Proficient' rating" (105 Illinois Compiled Statute Annotated 5/24-16.5(b) (2011)).[1] Additionally, "if a teacher in contractual continued service successfully completes a remediation plan following a rating of 'unsatisfactory' and receives a subsequent rating of 'unsatisfactory' in any of the teacher's annual or biannual overall performance evaluation ratings received during the **36-month period** following the teacher's completion of the remediation plan, then the school district may forego remediation and seek dismissal" of the teacher (Illinois Compiled Statute Annotated 105 ILCS 5/24A-5(n) (2011); Illinois Compiled Statute Annotated 105 ILCS 5/24-12 (2011)).	(i) Excellent; (ii) Proficient; (iii) Needs Improvement; or (iv) Unsatisfactory (105 Illinois Compiled Statute Annotated 5/24A-5(e) (2012); (105 Illinois Compiled Statute Annotated 5/34-85c(a) (2011)).

[1] The law also provides that a "school district may not, through agreement with a teacher or its teacher representatives, waive its right to dismiss a teacher under this Section" (105 Illinois Compiled Statute Annotated 5/24-16.5(b) (2011)).

Table A10.
Indiana's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Indiana	"Objective measures of student achievement and growth" must "**significantly** inform" teacher evaluations (Indiana Code § 20-28-11.5-4(4)(c)(2) (2012);Indiana Department of Education, (2012). Evaluation Law and Guidance. Retrieved May 24, 2012, from http://www.doe.in.gov/improvement/educator-effectiveness/evaluation-law-and-guidance).	Districts can choose to terminate teacher contracts at any time for incompetence which includes (i) "an ineffective designation on **two (2)** consecutive performance evaluations"; or (ii) "an ineffective designation or improvement necessary rating in **three (3) years of any five (5) year period**" (Indiana Code § 20-28-7.5-1(e)(4) (2011)).	(i) Highly effective; (ii) Effective; (iii) Improvement Necessary; and (iv) Ineffective (Indiana Code § 20-28-11.5-4(4)(c)(4) (2012)).

Table A11.
Louisiana's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Louisiana	The evaluation plans used by districts must meet the following: "**fifty** percent of such evaluations shall be based on evidence of growth in student achievement using a value-added assessment model as determined by the board for grade levels and subjects for which value-added data is available. For grade levels and subjects for which value-added data is not available and for personnel for whom value-added data is not available, the board shall establish measures of student growth" (Louisiana Revised Statute Annotated § 17:3902(B)(5) (2010)).	If a tenured teacher is rated "ineffective" under the state's performance evaluation, the teacher "shall **immediately** lose his tenure and all rights related thereto" (Louisiana Revised Statute Annotated § 17:442(C)(1)(2012) (amended by House Bill 974 (2012). Retrieved June 1, 2012, from Lousiana State Legislature Web Site: http://www.legis.state.la.us/billdata/streamdocument.asp?did=793654)). The law also provides that tenured teachers can be terminated for incompetence and willful neglect of duty. A teacher's rating as "ineffective" under the state's performance evaluation "shall constitute sufficient proof of poor performance, incompetence, or willful neglect of duty and no additional documentation shall be required to substantiate such charges" (Louisiana Revised Statute Annotated § 17:443(D) (2012) (House Bill 974 (2012). Retrieved June 1, 2012, from Lousiana State Legislature Web Site: http://www.legis.state.la.us/billdata/streamdocument.asp?did=793654).	(i) Effective; and (ii) Ineffective (Louisiana Revised Statute Annotated § 17:3902(C)(1) (2010)).

Table A12.
Maine's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Maine	"The proportionate weight of the standards and the measures is a local decision, but measurements of student learning and growth must be a **significant** factor in the determination of the rating of an educator" (20-A Maine Revised Statute Annotated § 13704(3)(A) (2015) (amended by Maine Legislature (2012). H.P. 1376 - L.D. 1858: An Act To Ensure Effective Teaching and School Leadership. Retrieved May 21, 2012, from www.mainelegislature.org/legis/bills/getPDF.asp?paper=HP1376&item=4&snum=125); 20-A Maine Revised Statute Annotated § 13705 (2015) (amended by Maine Legislature (2012). H.P. 1376 - L.D. 1858: An Act To Ensure Effective Teaching and School Leadership. Retrieved May 21, 2012, from www.mainelegislature.org/legis/bills/getPDF.asp?paper=HP1376&item=4&snum=125)).	Two consecutive years of summative effectiveness ratings of ineffective "constitutes just cause for nonrenewal of a teacher's contract unless the ratings are the result of bad faith" (20-A Maine Revised Statute Annotated § 13703 (2015) (amended by Maine Legislature (2012). H.P. 1376 - L.D. 1858: An Act To Ensure Effective Teaching and School Leadership. Retrieved May 21, 2012, from www.mainelegislature.org/legis/bills/getPDF.asp?paper=HP1376&item=4&snum=125)).[1]	School districts must use four levels of effectiveness ratings: "At least 2 of the levels must represent effectiveness, and at least one level must represent ineffectiveness" (20-A Maine Revised Statute Annotated § 13704(3)(C) (2015) (amended by Maine Legislature (2012). H.P. 1376 - L.D. 1858: An Act To Ensure Effective Teaching and School Leadership. Retrieved May 21, 2012, from www.mainelegislature.org/legis/bills/getPDF.asp?paper=HP1376&item=4&snum=125); (20-A Maine Revised Statute Annotated § 13702 (2015) (amended by Maine Legislature (2012). H.P. 1376 - L.D. 1858: An Act To Ensure Effective Teaching and School Leadership. Retrieved May 21, 2012, from www.mainelegislature.org/legis/bills/getPDF.asp?paper=HP1376&item=4&snum=125)).

[1] Just cause for dismissal or nonrenewal of teachers who have completed the probationary period is subject to collective bargaining negotiations (20-A Maine Revised Statute Annotated § 13201 (2012) (amended by Maine Legislature (2012). H.P. 1376 - L.D. 1858: An Act To Ensure Effective Teaching and School Leadership. Retrieved May 21, 2012, from www.mainelegislature.org/legis/bills/getPDF.asp?paper=HP1376&item=4&snum=125)).

Table A13.
Maryland's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Maryland	Teacher performance evaluations must "include data on student growth as a **significant** component of the evaluation and as one of multiple measures" (Maryland Code, Education, § 6-202(c)(4)(i) (2010)). However, "[n]o single criterion shall account for more than **35%** of the total performance evaluation criteria" (Maryland Code, Education, § 6-202(c)(5)(ii) (2010)).	None	The minimum categories are: (i) Satisfactory; (ii) Unsatisfactory (Code of Maryland Regulations (COMAR) 13A.07.04.02(A)(3) (2010)).

Table A14.
Massachusetts' Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Massachusetts	The law provides that "[m]ultiple measures of student learning, growth, and achievement" must be used (Code of Massachusetts Regulations (CMR) 603 CMR 35.07(1)(a) (2011)).	None[1]	The four ratings categories used are: (i) Exemplary;[2] (ii) Proficient;[3] (iii) Needs Improvement;[4] (iv) Unsatisfactory[5] (Code of Massachusetts Regulations (CMR) 603 CMR 35.02 (2011); (Code of Massachusetts Regulations (CMR) 603 CMR 35.08(1) (2011)).[6]

[1] The law does provide that teacher evaluations "may be used in decisions to dismiss, demote or remove a teacher" (Massachusetts General Laws Annotated 71 § 38 (1993)).
[2] This refers to where the "educator's performance consistently and significantly exceeds the requirements of a standard or overall" (Code of Massachusetts Regulations (CMR) 603 CMR 35.02 (2011)).
[3] This refers to where the "educator's performance fully and consistently meets the requirements of a standard or overall" (Code of Massachusetts Regulations (CMR) 603 CMR 35.02 (2011)).
[4] This refers to where the "educator's performance on a standard or overall is below the requirements of a standard or overall, but is not considered to be unsatisfactory at this time.
Improvement is necessary and expected" (Code of Massachusetts Regulations (CMR) 603 CMR 35.02 (2011)).
[5] This refers to where the "educator's performance on a standard or overall has not significantly improved following a rating of needs improvement, or the educator's performance is consistently below the requirements of a standard or overall and is considered inadequate, or both" (Code of Massachusetts Regulations (CMR) 603 CMR 35.02 (2011)).
[6] Furthermore, "the evaluator will assign the rating on growth in student performance consistent with Department guidelines: (a) A rating of high indicates significantly higher than one year's growth relative to academic peers in the grade or subject. (b) A rating of moderate indicates one year's growth relative to academic peers in the grade or subject. (c) A rating of low indicates significantly lower than one year's student learning growth relative to academic peers in the grade or subject" (Code of Massachusetts Regulations (CMR) 603 CMR 35.09(3) (2011)).

Table A15.
Michigan's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Michigan	School district evaluations of teachers must comply with the following: "For the annual year-end evaluation for the 2013-2014 school year, at least **25%** of the annual year-end evaluation shall be based on student growth and assessment data" (Michigan Compiled Laws § 380.1249(2)(a)(i) (2011)). "For the annual year-end evaluation for the 2014-2015 school year, at least **40%** of the annual year-end evaluation shall be based on student growth and assessment data" (Michigan Compiled Laws § 380.1249(2)(a)(i) (2011)). "Beginning with the annual year-end evaluation for the 2015-2016 school year, at least **50%** of the annual year-end evaluation shall be based on student growth and assessment data" (Michigan Compiled Laws § 380.1249(2)(a)(i) (2011)).	A district must dismiss a teacher who receives a rating of "ineffective on **3** consecutive annual year-end evaluations"[1] (Michigan Compiled Laws § 380.1249(2)(h) (2011)).[2]	(i) Highly effective; (ii) Effective; (iii) Minimally effective; or (iv) Ineffective (Michigan Compiled Laws § 380.1249(1)(c)(2011)); (Michigan Compiled Laws § 380.1249(2)(e) (2011)).

[1] Additionally, the law provides that "[t]his subdivision does not affect the ability of a school district, intermediate school district, or public school academy to dismiss an ineffective teacher from his or her employment regardless of whether the teacher is rated as ineffective on 3 consecutive annual year-end evaluations" (Michigan Compiled Laws § 380.1249(2)(h) (2011)).

[2] Ironically, even though the choice of three as the number of evaluations is arguably arbitrary, the state law provides that "discharge or demotion of a teacher on continuing tenure may be made only for a reason that is not arbitrary or capricious" (Michigan Compiled Laws § 38.101(1) (2011)). A quick note on Missouri: the state seems poised to introduce student achievement data into its evaluation process in the near future ((JoLynne, 2012). New Teacher Evaluation System on Agenda for Missouri State Board of Education. Retrieved May 21, 2012, from KC Education Enterprise Web Site: http://kceducationenterprise.org/2012/05/17/new-teacher-evaluation-system-on-agenda-for-missouri-state-board-of-education). Mississippi appears to also be on the same path (Hess, J. (2012, January, 18). Mississippi Department of Education Testing Teacher Evaluation System. Retrieved May 21, 2012, from MPB News Web Site: http://mpbonline.org/News/article/mississippi_department_of_education_testing_teacher_evaluation_system). California, on the other hand, seems reluctant to adopt evaluations based on student test scores (Los Angeles Times, (2012, May, 10). State Education Board Wants to Avoid New Teacher Evaluation Plan. retrieved May 21, 2012, from http://latimesblogs.latimes.com/lanow/2012/05/california-education-board-teacher-evaluation.html). Nebraska appears to want to take the approach of merely creating a model evaluation which local schools district can opt to adopt or not adopt (Reutter, H. (2012, March, 24). Education Officials Question Use of Yearly Progress Checks. Retrieved May 21, 2012, from http://www.theindependent.com/news/local/education-officials-question-use-of-yearly-progress-checks/article_003815de-7620-11e1-bb93-0019bb2963f4.html).

North Carolina does not yet use quantified student achievement. However, for low-performing schools, the state "shall dismiss a teacher, principal, assistant principal, director, supervisor, or other licensed personnel when the Secretary receives two consecutive evaluations that include written findings and recommendations regarding that person's inadequate performance" (North Carolina General Statute § 115C-325(p)(1) (2011); North Carolina General Statute § 115C-325(q)(2) (2011); North Carolina State Board of Education, (2009). North Carolina Teacher Evaluation Process. Retrieved May 21, 2012, from http://www.ncpublicschools.org/docs/profdev/training/teacher/teacher-eval.pdf). No timeline is specified for teachers in schools that are not low-performing. The law does, however, allow for the dismissal of career teachers on the grounds of inadequate performance. "Inadequate performance for a teacher shall mean (i) the failure to perform at a proficient level on any standard of the evaluation instrument or (ii) otherwise performing in a manner that is below standard. … For a career teacher, a performance rating below proficient shall constitute inadequate performance unless the principal noted on the instrument that the teacher is making adequate progress toward proficiency given the circumstances" (North Carolina General Statute § 115C-325(e)(3) (2011)). See also North Carolina State Board of Education, (2009). North Carolina Teacher Evaluation Process. retrieved May 21, 2012, from http://www.ncpublicschools.org/docs/profdev/training/teacher/teacher-eval.pdf).

Table A16.
Minnesota's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Minnesota	The local school board and the teacher's union are to negotiate an evaluation process that "must use an agreed upon teacher value-added assessment model for the grade levels and subject areas for which value-added data are available and establish state or local measures of student growth for the grade levels and subject areas for which value-added data are not available as a basis for **35** percent of teacher evaluation results" (Minnesota Statute § 122A.40(8)(a),(b)(8) (2013); Minnesota Statute § 122A.41(5)(a),(b)(8) (2013)).	A school district can choose to terminate a teacher's continuing contract **at the end of the school year** for inefficiency based on the results of the teacher's evaluations (Minnesota Statute § 122A.40(9)(a) (2014); Minnesota Statute § 122A.41(6)(a)(3) (2014)). Furthermore, the law provides that the school district "must discipline" a teacher who fails to make adequate progress in teacher improvement based on the evaluation results. Such discipline "may include a last chance warning, termination, discharge, nonrenewal, transfer to a different position, a leave of absence, or other discipline a school administrator determines is appropriate" (Minnesota Statute § 122A.41(5)(b)(12) (2013)).	None

Table A17.
Nevada's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Nevada	Student achievement data maintained in the state's automated system of accountability information must account for **50%** of the teacher evaluations adopted by each school district (Nevada Revised Statute § 391.3125(2) (2013); Nevada Revised Statute § 391.465(2)(c) (2011); Nevada Revised Statute § 386.650(1)(c)-(e) (2013)).	"A postprobationary employee who receives an unsatisfactory evaluation … or any other equivalent evaluation designating his or her overall performance as below average, for **2** consecutive school years shall be deemed to be a probationary employee … and must serve an additional probationary period" (Nevada Revised Statute § 391.3129 (2013)).	(i) Highly effective; (ii) Effective; (iii) Minimally effective; or (iv) Ineffective (Nevada Revised Statute § 391.465(2)(a) (2011); Nevada Revised Statute § 391.3125(2) (2013)).

Table A18.
New Jersey's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
New Jersey	Fifty percent of the teacher evaluations must be based on student achievement (New Jersey Administrative Code Executive Order No. 42(3)(a) (2010); New Jersey Educator Effectiveness Task Force (2011). Interim Report 15. Retrieved June 1, 2012, from http://www.state.nj.us/education/educators/effectiveness.pdf)).[1]	None	(i) Ineffective; (ii) Partially effective; (iii) Effective; and (iv) Highly Effective (State of New Jersey Department of Education (2011). Department of Education Announces 11 Districts to Participate in a Teacher Evaluation Pilot Program. Retrieved June 1, 2012, from http://www.state.nj.us/education/news/2011/0901ee4nj.htm); New Jersey Educator Effectiveness Task Force (2011). Interim Report 14. Retrieved June 1, 2012, from http://www.state.nj.us/education/educators/effectiveness.pdf).[2]

Table A19.
New York's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
New York	The state's teacher performance evaluation system must be comprised of: (i) a state assessments and other comparable measures subcomponent which shall comprise twenty or twenty-five percent of the evaluation; (ii) a locally selected measures of student achievement subcomponent which shall comprise twenty or fifteen percent of the evaluation; and (iii) an other measures of teacher or principal effectiveness subcomponent which shall comprise the remaining sixty percent of the evaluation, which in sum shall constitute the composite teacher or principal effectiveness score (New York Education Law § 3012-c(1)(a)(1) (2012); New York Education Law § 3012-c(1)(h) (2012)). For subjects and grades with**out** an approved value-added model, "**forty** percent of the composite score of effectiveness shall be based on student achievement measures as follows: (i) **twenty** percent of the evaluation shall be based upon student growth data on state assessments as prescribed by the commissioner or a comparable measure of student growth if such growth data is not available; and (ii) **twenty** percent shall be based on other locally selected measures of student achievement that are determined to be rigorous and comparable across classrooms in accordance with the regulations of the commissioner and as are developed locally in a manner consistent with procedures negotiated pursuant to the requirements of article fourteen of the civil service law (New York Education Law § 3012-c(1)(b)(1) (2012); (New York Education Law § 3012-c(1)(e)(1) (2012)).	A pattern of ineffective teaching or performance shall be defined to mean **two** consecutive annual ineffective ratings received by a classroom teacher pursuant to annual professional performance reviews (New York Education Law § 3012-c(6) (2012); New York Education Law § 3020(1) (2010)).	The overall composite scoring ranges for performance evaluations shall be as follows: (i) Highly Effective if the teacher gets a composite effectiveness score of 91-100; (ii) Effective if the teacher gets a composite effectiveness score of 75-90;

Table A19. (Cont.'d)
New York's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
			(iii) Developing if the teacher gets a composite effectiveness score of 65-74; and (iv) Ineffective if the gets a composite effectiveness score of 0-64 (New York Education Law § 3012-c(1)(a)(2) (2012)).
New York			For subjects and grades with**out** an approved value-added model, "the scoring ranges for the student growth on **state** assessments or other comparable measures subcomponent" of the performance evaluations shall be as follows: (i) A Highly Effective rating in this subcomponent if the teacher's results are well-above the state average for similar students and he/she achieves a subcomponent score of 18-20;

Table A19. (Cont.'d)
New York's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
New York			(ii) An Effective rating in this subcomponent if the teacher's results meet the state average for similar students and he/she achieves a subcomponent score of 9-17; or (iii) A Developing rating in this subcomponent if the teacher's results are below the state average for similar students and he/she achieves a subcomponent score of 3-8; or (iv) An Ineffective rating in this subcomponent, if the teacher's results are well-below the state average for similar students and he/she achieves a subcomponent score of 0-2 (New York Education Law § 3012-c(1)(a)(3) (2012)).

Table A19. (Cont.'d)
New York's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
New York			For subjects and grades with an approved value-added model, "the scoring ranges for the student growth on **state** assessments or other comparable measures subcomponent" of the performance evaluations shall be as be as follows: (i) a highly effective rating in this subcomponent if the teacher's results are well-above the state average for similar students and he/she achieves a subcomponent score of 22-25; (ii) an effective rating in this subcomponent if the teacher's results meet the state average for similar students and he/she achieves a subcomponent score of 10-21; or (iii) a developing rating in this subcomponent if the teacher's results are below the state average for similar students and he/she achieves a subcomponent score of 3-9; or

Table A19. (Cont.'d)
New York's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
New York			(iv) an ineffective rating in this subcomponent, if the teacher's results are well-below the state average for similar students and he/she achieves a subcomponent score of 0-2 (New York Education Law § 3012-c(1)(a)(4) (2012)). For subjects and grades without an approved value-added model, "the scoring ranges for the **locally** selected measures of student achievement subcomponent" of the performance evaluations shall be as follows: (i) a highly effective rating in this subcomponent if the results are well-above district-adopted expectations for student growth or achievement and the teacher gets a subcomponent score of 18-20; or

Table A19. (Cont.'d)
New York's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
New York			(ii) an effective rating in this subcomponent if the results meet district-adopted expectations for growth or achievement and the teacher gets a subcomponent score of 9-17; or (iii) a developing rating in this subcomponent if the results are below district-adopted expectations for growth or achievement and the teacher gets a subcomponent score of 3-8; or (iv) an ineffective rating in this subcomponent if the results are well-below district-adopted expectations for growth or achievement and the teacher gets a subcomponent score of 0-2 (New York Education Law § 3012-c(1)(a)(5) (2012)).

Table A19. (Cont.'d)
New York's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
New York			For subjects and grades with an approved value-added model, "the scoring ranges for the **locally** selected measures of student achievement subcomponent" of the performance evaluations shall be as follows: (i) A Highly effective rating in this subcomponent if the results are well-above district-adopted expectations for student growth or achievement and the teacher gets a subcomponent score of 14-15; or (ii) An Effective rating in this subcomponent if the results meet district-adopted expectations for growth or achievement and the teacher gets a subcomponent score of 8-13; or

Table A19. (Cont.'d)
New York's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
New York			(iii) A Developing rating in this subcomponent if the results are below district-adopted expectations for growth or achievement and the teacher gets a subcomponent score of 3-7; or (iv) An Ineffective rating in this subcomponent if the results are well-below district-adopted expectations for growth or achievement and the teacher gets a subcomponent score of 0-2 (New York Education Law § 3012-c(1)(a)(6) (2012)).

Table A20.
Ohio's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Ohio	Student academic growth must constitute **fifty** percent of the teacher evaluation (Ohio Revised Code Annotated § 3319.112(A)(1) (2011)).	None[1]	(i) Accomplished; (ii) Proficient; (iii) Developing; and (iv) Ineffective (Ohio Revised Code Annotated § 3319.112(B)(1) (2011)).

[1] The law does provide, however, that each school district must "include in its evaluation policy procedures for using the evaluation results for retention and promotion decisions and for removal of poorly performing teachers" (Ohio Revised Code Annotated § 3319.111(E) (2011)).

Table A21.
Oklahoma's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Oklahoma	**Fifty** percent (50%) of the teacher's evaluations must be based on quantitative components divided as follows: (1) **thirty-five** percentage points based on student academic growth using multiple years of standardized test data, as available; and (2) **fifteen** percentage points based on other academic measurements (Oklahoma Statutes Annotated § 6-101.16(B)(4) (2011)).	Oklahoma authorizes dismissal of career teachers (tenured teachers) for instructional ineffectiveness (Oklahoma Statutes Annotated § 6-101.22 (A)(5) (2011)) as follows: (i) A career teacher who has been rated as 'Ineffective' as measured pursuant to the Oklahoma Teacher and Leader Effectiveness Evaluation System (TLE) … for **two (2)** consecutive school years shall be dismissed or not reemployed on the grounds of instructional ineffectiveness by the school district (Oklahoma Statutes Annotated § 6-101.22 (C)(1) (2011)). (ii) A career teacher who has been rated as 'Needs Improvement' or lower pursuant to the TLE for **three (3)** consecutive school years shall be dismissed or not reemployed on the grounds of instructional ineffectiveness by the school district (Oklahoma Statutes Annotated § 6-101.22 (C)(2) (2011)). (iii) A career teacher who has not averaged a rating of at least 'Effective' as measured pursuant to the TLE over a five-year period shall be dismissed or not reemployed on the grounds of instructional ineffectiveness by the school district (Oklahoma Statutes Annotated § 6-101.22 (C)(3) (2011)).	The Oklahoma Teacher and Leader Effectiveness Evaluation System (TLE) uses the following five-tier rating system: (i) Superior; (ii) Highly effective; (iii) Effective; (iv) Needs Improvement; and (v) Ineffective (70 Oklahoma Statutes Annotated § 6-101.16(B)(1) (2011).

Table A22.
Oregon's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Oregon	Student learning must be a **significant** factor in teacher evaluations developed by school districts (Oregon Revised Statutes Annotated § 342.856 (2013); Oregon Administrative Rules Compilation 581-022-1723 (2013); Oregon State Board of Education (2012, May). Educator Effectiveness: Oregon Framework for Teacher and Administrator Evaluation and Support Systems. Retrieved June 1, 2012, from www.ode.state.or.us/stateboard/pdfs/2012-may-17-educator-effectiveness-framework-for-local-teacher-and-admin-evaluation-systems.pdf).[1]	None[2]	The four performance levels to be used are: (i) Level 1 – Unsatisfactory (ii) Level 2 – Basic; (iii) Level 3 – Satisfactory; and (iv) Level 4 – Distinguished (Oregon State Board of Education (2012, May). Educator Effectiveness: Oregon Framework for Teacher and Administrator Evaluation and Support Systems. Retrieved June 1, 2012, from www.ode.state.or.us/stateboard/pdfs/2012-may-17-educator-effectiveness-framework-for-local-teacher-and-admin-evaluation-systems.pdf)

[1] Oregon is in the process of developing its policies (Oregon State Board of Education (2012, May). Educator Effectiveness: Oregon Framework for Teacher and Administrator Evaluation and Support Systems. Retrieved June 1, 2012, from www.ode.state.or.us/stateboard/pdfs/2012-may-17-educator-effectiveness-framework-for-local-teacher-and-admin-evaluation-systems.pdf; Oregon Education Association (2011). Teacher Evaluation. Retrieved June 1, 2012, from http://www.oregoned.org/site/pp.asp?c=9dKKKYMDH&b=6573779; Oregon Department of Education (2012, May). Oregon Framework for Teacher and Administrator Evaluation and Support Systems Draft. Retrieved June 1, 2012, from http://www.google.com/url?sa=t&rct=j&q=oregon%20framework%20for%20teacher%20and%20administrator%20evaluation%20and%20support%20systems&source=web&cd=1&sqi=2&ved=0CGIQFjAA&url=http%3A%2F%2Fwww.ode.state.or.us%2Fstateboard%2Fpdfs%2Fhandout---oregon-framework-for-educators--administrator-evaluations.pdf&ei=IJ7DT6SfLILs6gHI-KHSCg&usg=AFQjCNEvLGt8qj_nBaRIg0OUnUOGvTrQhQ&cad=rja).
[2] While no timeline is specified, the state allows dismissal for inefficiency (Oregon Revised Statute §342.865(1)(a)(1999)), inadequate performance (Oregon Revised Statute § 342.865(1)(g)(1999)) or "[f]ailure to comply with such reasonable requirements as the board may prescribe to show normal improvement and evidence of professional training and growth" (Oregon Revised Statute § 342.865(1)(h)(1999)).

Table A23.
Pennsylvania's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Pennsylvania	The state legislature and the department of education "seem to be gravitating toward counting multiple measures of student achievement and growth as up to 50 percent of a teacher's individual evaluation result, but no final decision has been made" (PSEA Education Services Division (2011). Pennsylvania's New Teacher Evaluation System. Retrieved June 1, 2012, from http://slea.psealocals.org/Portals/444/Advisory%20new%20eval%20system%20FINAL%20Aug%202011.pdf)[1]	A teacher could be terminated for "unsatisfactory teaching performance based on **two (2)** consecutive ratings of the employee's teaching performance that are to include classroom observations, not less than four (4) months apart, in which the employee's teaching performance is rated as unsatisfactory" (24 Pennsylvania Statutes and Consolidated Statutes § 11-1122 (1996)).	To be determined

[1] Pennsylvania is in the process of creating its evaluation policies (Pennsylvania Department of Education, (2011). Teacher Evaluation Project FAQ. retrieved June 1, 2012, from http://www.portal.state.pa.us/portal/server.pt/community/newsroom/7234/teacher_evaluation/1036220; Aument, Ryan (2010). Pennsylvania Department of Education to Begin Statewide Pilot Project to Continue Education Reform Efforts. Retrieved June 1, 2012, fromhttp://repaument.com/NewsItem.aspx?NewsID=12432). A side note about Texas: While the state does not have a statewide requirement of quantified student performance for evaluations, some individual districts in Texas use student performance to evaluate teachers (Texas Education Agency (2011). Teacher Evaluations Including Student Performance. Retrieved June 1, 2012, from ww.tea.state.tx.us/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=2147502760&libID=2147502754; Texas Education Agency (2011). Systems Used to Evaluate Teacher Performance. Retrieved June 1, 2012, from http://www.tea.state.tx.us/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=2147502759&libID=2147502753). Similar information is available for Vermont (Vermont Department of Education (2012, March). Teacher Evaluation Survey. Retrieved June 1, 2012, from http://education.vermont.gov/documents/EDU-ARRA_SFSF_Teacher_%20State_Level_Evaluation_Survey.pdf).

Table A24.
Rhode Island's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Rhode Island	"An educator's overall evaluation of effectiveness is primarily determined by evidence of impact on student growth and academic achievement" (Rhode Island Department of Elementary and Secondary Education (2009). Educator Evaluation System Standards 3. Retrieved June 1, 2012, from http://www.ride.ri.gov/EducatorQuality/EducatorEvaluation/Docs/EdEvalStandards.pdf; Rhode Island Board of Regents (2011). The Rhode Island Model: Guide to Evaluating Building Administrators and Teachers 61-66. Retrieved June 1, 2012, from http://www.ride.ri.gov/EducatorQuality/EducatorEvaluation/Docs/RIModelGuide.pdf; Rhode Island Board of Regents (2011). The Rhode Island Growth Model. Retrieved June 4, 2012, from http://www.ride.ri.gov/assessment/DOCS/RIGM/RIGM_Pamphlet_FINAL-Spring_2011.pdf).	None[1]	The four performance evaluation categories required are: (i) Highly Effective; (ii) Effective,; (iii) Developing; and (iv) Ineffective (Rhode Island Board of Regents (2011). The Rhode Island Model: Guide to Evaluating Building Administrators and Teachers 61. Retrieved June 1, 2012, from http://www.ride.ri.gov/EducatorQuality/EducatorEvaluation/Docs/RIModelGuide.pdf).

[1] "Teachers who are rated as Developing or Ineffective at the end of the year will be placed on an Improvement Plan and will work with an improvement team to assist them with their development over the course of the following year. … The teacher's district will identify personnel actions that may occur if he or she does not adequately improve his or her performance" (Rhode Island Board of Regents (2011). The Rhode Island Model: Guide to Evaluating Building Administrators and Teachers 28. Retrieved June 1, 2012, from http://www.ride.ri.gov/EducatorQuality/EducatorEvaluation/Docs/RIModelGuide.pdf).

Table A25.
South Dakota's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
South Dakota	**Fifty** percent of the teacher evaluation must be "based on quantitative measures of student growth, based on a single year or multiple years of data" (South Dakota Codified Laws § 13-42-34(2) (a) (2014) (amended by South Dakota Legislature (2011). House Bill 1234. Retrieved June 1, 2012, from http://legis.state.sd.us/sessions/2012/Bills/HB1234ENR.pdf)).	A district can choose not to renew a teacher's contract if the teacher is rated unsatisfactory on **two** consecutive evaluations (South Dakota Codified Laws § 13-43-6.3 (2012) (amended by South Dakota Legislature (2011). House Bill 1234. Retrieved June 1, 2012, from http://legis.state.sd.us/sessions/2012/Bills/HB1234ENR.pdf)).	The performance evaluations are based on the following four-tier rating system: (i) Distinguished; (ii) Proficient; (iii) Basic; and (iv) Unsatisfactory (South Dakota Codified Laws § 13-42-34(5) (2014) (amended by South Dakota Legislature (2011). House Bill 1234. Retrieved June 1, 2012, from http://legis.state.sd.us/sessions/2012/Bills/HB1234ENR.pdf)).

Table A26.
Tennessee's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Tennessee	**Fifty** percent (50%) of the teacher performance evaluation in the Tennessee Educator Acceleration Model (TEAM) must be made up of student achievement data divided as follows: (i) **35%** must be "student achievement data based on student growth data as represented by the Tennessee Value-Added Assessment System (TVAAS) … or some other comparable measure of student growth, if no such TVAAS data is available" (Tennessee Code Annotated § 49-1-302(d)(2)(A)(i) (2011)); (ii) the remaining **15%** must use some other student achievement measure chosen from a list of created by the teacher evaluation advisory committee and approved by the state board of education (Tennessee Code Annotated § 49-1-302(d)(2)(A)(ii) (2011)).	"Any teacher who, after acquiring tenure status, receives two (2) consecutive years of evaluations demonstrating an overall performance effectiveness level of 'below expectations' or 'significantly below expectations' … shall be returned to probationary status by the director of schools until the teacher has received two (2) consecutive years of evaluations demonstrating an overall performance effectiveness level of 'above expectations' or 'significantly above expectations'" (Tennessee Code Annotated § 49-5-504(e) (2011)).[1] The state law provides, however, that "[n]o teacher who acquired tenure status prior to July 1, 2011, shall be returned to probationary status" (Tennessee Code Annotated § 49-5-501(11) (2011)). In fact, the law specifically states that the provision about two consecutive years of evaluations mentioned above does not apply to teachers who got tenure before July 1, 2011 (Tennessee Code Annotated § 49-5-504(f) (2011)).	Tennessee Educator Acceleration Model (TEAM) uses the following five categories: (i) Significantly Above Expectations based on a score between 425-500;[2] (ii) Above Expectations based on a score between 350-424.99;[3] (iii) At Expectations based on a score between 275-349.99;[4] (iv) Below Expectations based on a score between 200-274.99;[5] and

Table A26. (Cont.'d)
Tennessee's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Tennessee			(v) Significantly Below Expectations based on a score below 200[6] (Tennessee Department of Education (2011). Tennessee First to the Top Score Calculations 3. Retrieved June 1, 2012, from http://team-tn.org/assets/educator-resources/Calculating_the_Effectiveness_Rating.pdf).

[1] "When a teacher who has returned to probationary status has received two (2) consecutive years of evaluations demonstrating an overall performance effectiveness level of 'above expectations' or 'significantly above expectations,' the teacher is again eligible for tenure and shall be either recommended by the director of schools for tenure or nonrenewed; provided, however, that the teacher cannot be continued in employment if tenure is not granted by the board of education" (Tennessee Code Annotated § 49-5-504(e) (2011)).
[2] "A teacher at this level exemplifies the instructional skills, knowledge, and responsibilities described in the rubric, and implements them without fail. He/she is adept at using data to set and reach ambitious teaching and learning goals. He/she makes a significant impact on student achievement and should be considered a model of exemplary teaching" (Tennessee Department of Education (2011).
Tennessee First to the Top Score Calculations 3. Retrieved June 1, 2012, from http://team-tn.org/assets/educator-resources/Calculating_the_Effectiveness_Rating.pdf).
[3] "A teacher at this level comprehends the instructional skills, knowledge, and responsibilities described in the rubric and implements them consistently. He/she is skilled at using data to set and reach appropriate teaching and learning goals and makes a strong impact on student achievement" (Tennessee Department of Education (2011). Tennessee First to the Top Score Calculations 3.
Retrieved June 1, 2012, from http://team-tn.org/assets/educator-resources/Calculating_the_Effectiveness_Rating.pdf).
[4] "A teacher at this level understands and implements most of the instructional skills, knowledge, and responsibilities described in the rubric. He/she uses data to set and reach teaching and learning goals and makes the expected impact on student achievement" (Tennessee Department of Education (2011). Tennessee First to the Top Score Calculations 3. Retrieved June 1, 2012, from http://team-tn.org/assets/educator-resources/Calculating_the_Effectiveness_Rating.pdf).
[5] "A teacher at this level demonstrates some knowledge of the instructional skills, knowledge, and responsibilities described in the rubric, but implements them inconsistently. He/she may struggle to use data to set and reach appropriate teaching and learning goals. His/her impact on student achievement is less than expected" (Tennessee Department of Education (2011). Tennessee First to the Top Score Calculations 3. Retrieved June 1, 2012, from http://team-tn.org/assets/educator-resources/Calculating_the_Effectiveness_Rating.pdf).
[6] "A teacher at this level has limited knowledge of the instructional skills, knowledge, and responsibilities described in the rubric, and struggles to implement them. He/she makes little attempt to use data to set and reach appropriate teaching and learning goals, and has little to no impact on student achievement" (Tennessee Department of Education (2011). Tennessee First to the Top Score Calculations 3. Retrieved June 1, 2012, from http://team-tn.org/assets/educator-resources/Calculating_the_Effectiveness_Rating.pdf).

Table A27.
Utah's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Utah	Currently, the law merely requires that evaluation systems adopted by school districts must include "evidence of student growth" (Utah Administrative Rule 277-531-3(B)(3)(b) (2011); Utah Administrative Rule 277-531-3(C)(1)(b) (2011); Utah Administrative Rule 277-531-3(F)(4)(b) (2015)). The state board of education will determine the weight of each component of the evaluation (Utah Administrative Rule 277-531-3(F)(5) (2011)).	"If the district intends to terminate a career employee's contract during its term for reasons of unsatisfactory performance or discontinue a career employee's contract beyond the current school year for reasons of unsatisfactory performance, the unsatisfactory performance must be documented in at least two evaluations conducted at any time within the preceding three years in accordance with district policies or practices" (Utah Code Annotated § 53A-8-104(2) (2011)).[1]	To be determined (Utah Administrative Rule 277-531-1(2011); Utah State Office of Education (2012). Teaching and Learning Licensing. Retrieved June 3, 2012, from http://www.schools.utah.gov/cert/Educator-Effectiveness-Project.aspx)

[1] Each district's evaluation policy must specify the employment consequences of teachers' failure to meet performance requirements (Utah Administrative Rule 277-531-3(F)(7) (2011)). Utah is still developing its evaluation policy (Utah State Office of Education (2012). Teaching and Learning Licensing. Retrieved June 3, 2012, from http://www.schools.utah.gov/cert/Educator-Effectiveness-Project.aspx).

Table A28.
Virgina's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Virginia	The Guidelines for Uniform Performance Standards and Evaluation Criteria model set forth by the state board of education "recommends that 40 percent of teachers' evaluations be based on student academic progress, as determined by multiple measures of learning and achievement, including, when available and applicable, student-growth data" (Virginia Department of Education (2012). Teaching in Virginia: Performance and Evaluation. Retrieved June 3, 2012, from http://www.doe.virginia.gov/teaching/performance_evaluation/; Virginia Department of Education (2012). The Guidelines for Uniform Performance Standards and Evaluation Criteria for Teachers 5, 67-68. Retrieved June 3, 2012, from http://www.doe.virginia.gov/teaching/performance_evaluation/guidelines_ups_eval_criteria_teachers.pdf).	None[1]	(i) Exemplary; (ii) Proficient; (iii) Developing/Needs Improvement; (iv) Unacceptable (Virginia Department of Education (2012). The Guidelines for Uniform Performance Standards and Evaluation Criteria for Teachers 58. Retrieved June 3, 2012, from http://www.doe.virginia.gov/teaching/performance_evaluation/guidelines_ups_eval_criteria_teachers.pdf).

[1] However, incompetence, one of the grounds for dismissal of continuing contract teachers "may be construed to include, but shall not be limited to, consistent failure to meet the endorsement requirements for the position or performance that is documented through evaluation to be consistently less than satisfactory" (Virginia Code Annotated § 22.1-307(B) (2008)). While Virginia does not explicitly identify a timeline specific to teachers with continuing contracts, it specifies that for teachers in the state who are rated 'Unacceptable', the school district could opt to recommend the teacher for dismissal. If the teacher is not dismissed, he/she will participate in a Performance Improvement Plan. If the teacher gets a second 'Unacceptable' rating, the district must recommend the teacher for dismissal (Virginia Department of Education (2012). The Guidelines for Uniform Performance Standards and Evaluation Criteria for Teachers 77. Retrieved June 3, 2012, from http://www.doe.virginia.gov/teaching/performance_evaluation/guidelines_ups_eval_criteria_teachers.pdf). For teachers with continuing contracts who get a rating of 'Unacceptable', the guidelines provide that "a Performance Improvement Plan will be developed and implemented. Following implementation of the Performance Improvement Plan, additional performance data, including observations as applicable, will be collected" (Virginia Department of Education (2012). The Guidelines for Uniform Performance Standards and Evaluation Criteria for Teachers 77. Retrieved June 3, 2012, from http://www.doe.virginia.gov/teaching/performance_evaluation/guidelines_ups_eval_criteria_teachers.pdf).

Table A29.
Washington's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Washington	Not specified yet[1]	"When a continuing contract employee with five or more years of experience receives a comprehensive summative evaluation performance rating below level 2 for two consecutive years, the school district shall, within ten days of the completion of the second summative comprehensive evaluation or May 15th, whichever occurs first, implement the employee notification of discharge" (Revised Code of Washington Annotated 28A.405.100(4)(c) (2012) amended by Senate Bill 5895; Revised Code of Washington Annotated 28A.405.300 (2010)).	Teacher summative performance evaluations ratings use the following four categories: (i) Level 1 - unsatisfactory; (ii) Level 2 - basic; (iii) Level 3 - proficient; and (iv) Level 4 – distinguished (Revised Code of Washington Annotated 28A.405.100(2)(a) (2012) amended by S.B. 5895; Washington State Legislature (2012). Office of Superintendent of Public Instruction, Teacher/Principal Evaluation Pilot. Retrieved June 3, 2012, from http://www.k12.wa.us/EdLeg/TPEP/default.aspx).

[1] Washington state is in the process of creating its evaluation policy (Dorn, Randy (2012). Teacher and Principal Evaluation Pilot: Report to the Legislature. Retrieved June 3, 2012, from State Superintendent of Public Instruction Web Site: http://tpep.files.wordpress.com/2011/07/tpep_leg_report-july_2011_full.pdf; Washington State Legislature (2012). Office of Superintendent of Public Instruction, Teacher/Principal Evaluation Pilot. Retrieved June 3, 2012, from http://www.k12.wa.us/EdLeg/TPEP/default.aspx; Washington's Teacher/Principal Evaluation Pilot (2012). Retrieved June 3, 2012, from State Superintendent of Public Instruction Web Site: http://tpep-wa.org/).

Table A30.
West Virginia's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
West Virginia	"Fifteen percent of the evaluation shall be based on evidence of the learning of the students assigned to the educator … and five percent of the evaluation shall be based on student learning growth measured by the school-wide score on the state summative assessment" (West Virginia Code §18A-3C-2(c)(2) (2013)).	A teacher who receives a rating of 'Unsatisfactory' must be given a performance improvement plan and provided a reasonable time though not more than 12 months to comply with the plan.[1] If the teacher's evaluation following the period of improvement plan rates the teacher as 'Unsatisfactory' the evaluator could choose to recommend dismissal of the teacher (West Virginia Code §18A-3C-2(c)(2) (2013)).	(i) Satisfactory; and (ii) Unsatisfactory (West Virginia Code §18A-3C-2(h) (2013)).

[1] West Virginia allows dismissal of teachers with continuing contracts based on unsatisfactory performance (West Virginia Code § 18A-2-8(a) (2007)). Unsatisfactory performance is determined by the teacher's evaluation (West Virginia Code § 18A-2-8(b) (2007)). Moreover, a new law in West Virginia provides that the results of teacher evaluations will constitute "documentation for a dismissal on the grounds of unsatisfactory performance" (West Virginia Code §18A-3C-2(e)(7) (2013)).

Table A31.
Wisconsin's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Wisconsin	"Fifty percent of the total evaluation score assigned to a teacher or principal shall be based upon measures of student performance, including performance on state assessments, district-wide assessments, student learning objectives, school-wide reading at the elementary and middle-school levels, and graduation rates at the high school level" (Wisconsin Statutes Annotated 115.415(2)(a) (2014);Wisconsin Department of Public Instruction (2011). Wisconsin Framework for Educator Effectiveness: Preliminary Report and Recommendations 8)).	While no specific timeline is provided, the Wisconsin Department of Public Instruction indicates that "[a]n educator will not be allowed to remain at the developing level and continue to practice indefinitely. If an educator is rated as developing over a time period the educator will undergo an intervention phase to improve on the areas rated as developing. If, at the end of the intervention phase, the educator is still developing, the district shall move to a removal phase" (Wisconsin Department of Public Instruction (2011). Wisconsin Framework for Educator Effectiveness: Preliminary Report and Recommendations 8)).	(i) Developing; (ii) Effective; and (iii) Exemplary (Wisconsin Statutes Annotated 115.415(2)(c) (2014; Wisconsin Department of Public Instruction (2011). Wisconsin Framework for Educator Effectiveness: Preliminary Report and Recommendations 8)).

Table A31.
Wyoming's Approaches to the New Teacher Evaluation Movement

State	Teacher Evaluation Significantly Based On Quantified Student Achievement?	Timelines For Dismissing A Tenured Teacher Rated Ineffective	Teacher Performance Categories
Wyoming	Not specified[1] (Wyoming Rules and Regulations Education General Chapter 29 §§ 4 - 6 (2010); Borchardt, Jackie (2011, October 26). Report: Wyoming Educator Evaluations Could Be Stronger. Retrieved June 3, 2012, from Star-Tribune Web Site: http://trib.com/news/state-and-regional/report-wyoming-educator-evaluations-could-be-stronger/article_12776e10-604e-532e-a0cd-a9dc54fdebed.html)	Starting with the 2013-2014 school year, local school boards can choose to dismiss a teacher for "inadequate performance as determined through annual performance evaluation tied to student academic growth" (Wyoming Statutes Annotated § 21-7-110(a)(vii) (2012); Wyoming Statutes Annotated § 21-3-110(a)(xvii)-(xix) (2012)).[2]	(i) Highly Effective; (ii) Effective; and (iii) Ineffective (Wyoming Statutes Annotated § 21-2-304(b)(xv) (2012)).

[1] School districts seem to have some flexibility in the weighting allocated to student growth, though student growth must be included (Wyoming Department of Education (2011). Certified Personnel Evaluation System-Chapter 29. Retrieved June 3, 2012, from http://edu.wyoming.gov/Programs/certifiedpersonnelevaluationsytem.aspx)
[2] The state law does provide that "[s]ubject to satisfactory performance evaluation … a continuing contract teacher shall be employed by each school district on a continuing basis from year to year without annual contract renewal at a salary determined by the board of trustees of each district, said salary subject to increases from time to time as provided for in the salary provisions adopted by the board" (Wyoming Statutes Annotated § 21-7-104(a) (2012)).

About the Authors

Bruce Baker
Rutgers, The State University of New Jersey
bruce.baker@gse.rutgers.edu
Dr. Baker is Professor in the Department of Educational Theory, Policy and Administration at Rutgers, The State University of New Jersey in New Brunswick.

Joseph Oluwole
Montclair State University
oluwolej@mail.montclair.edu
Dr. Oluwole is an Associate Professor in the Department of Counseling and Educational Leadership at Montclair State University in New Jersey.

Preston C. Green III
The Pennsylvania State University
pcg10@psu.edu
Dr. Green is Professor of Education and Law, and Harry L. Batschelet II Chair of Educational Administration at The Pennsylvania State University.

About the Guest Editor and Assistant Guest Editors

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

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

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

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

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

education policy analysis archives

Volume 21 Number 5 January 28th, 2013 ISSN 1068-2341

of Education at Arizona State University Articles are indexed in CIRC (Clasificación Integrada de Revistas Científicas, Spain), DIALNET (Spain), Directory of Open Access Journals, EBSCO Education Research Complete, ERIC, Education Full Text (H.W. Wilson), QUALIS A2 (Brazil), SCImago Journal Rank; SCOPUS, SOCOLAR (China).

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

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