

Putting Growth and Value-Added Models on the Map: A National Overview

by Clarin Collins & Audrey Amrein-Beardsley

Background: *Within the last few years, the focus on educational accountability has shifted from holding students responsible for their own performance to holding those shown to impact student performance responsible—students’ teachers. Encouraged and financially incentivized by federal programs, states are becoming ever more reliant on statistical models used to measure students’ growth or value added and are attributing such growth (or decline) to students’ teachers of record. As states continue to join the growth and value-added model movement, it is difficult to find inclusive resources documenting the types of models used and plans for each state.*

Objective: *To capture state initiatives in this area, researchers collected data from all 50 states and the District of Columbia to provide others with an inclusive national growth and value-added model overview. Data yielded include information about the types of growth or value-added models used in each state, the legislature behind each state’s reform efforts, the standardized tests used for growth or value-added calculations, and the strengths and weaknesses of each state’s models as described by state personnel.*

Method: *This article synthesizes qualitative and quantitative themes as identified from data collected via multiple phone interviews and emails with state department of education personnel in charge of their own state’s initiatives in this area, as well as state websites. These data provide the most inclusive and up-to-date resource on national growth and value-added data usage, noting however that this is changing rapidly across the nation, given adjustments in policies, pieces of legislation, and the like.*

Conclusions: *Findings from this study provide a one-stop resource on what each state has in place or in development regarding growth or value-added model use as a key component of its state-based teacher evaluation systems. Despite widespread use, however, not one state has yet articulated a plan for formative data use by teachers. Federal and state leaders seem to assume that implementing growth and value-added models leads to simultaneous data use by teachers. In addition, state representatives expressed concern that the current emphasis on growth and value-added models could be applied to only math and English/language arts teachers with state standardized assessments (approximately 30% of all teachers). While some believe the implementation of the Common Core State Standards and its associated tests will help to alleviate such issues with fairness, more research is needed surrounding (the lack of) fairness and formative use associated with growth and value-added models.*

THE POLICY TOPOGRAPHY

Debates surrounding educational accountability have amplified over the past decade, given the federally incentivized stratagem to push states away from focusing on student-level accountability to now focus on teacher-level accountability and the “value” teachers “add” to student learning over time. In fact, many states have begun to reward and penalize measurably effective and ineffective teachers respectively, whereas effectiveness is increasingly, yet narrowly defined by the gains teachers’ students make on large-scale standardized tests over time (Barnett & Amrein-Beardsley, 2011).

This has occurred largely because of the educational policies and practices backed by U.S. Secretary of Education Arne Duncan and his Race to the Top (RtT) competition. This has been spurred by inimical actions of the media, first in Los Angeles (Felch, Song, & Smith, 2010) and most recently in New York City (Gonen, 2012; see also Rubenstein, 2012), during which major news outlets made public the names of teachers and their high or low “value-added” scores. This has also been stimulated by econometricians and statisticians claiming, for the first time in educational measurement history, that they can help states and districts reliably and precisely identify good and bad teachers using students’ composite levels of growth on large-scale standardized test scores over time. States and districts are left to trust that growth and value-added models can properly identify

effective teachers, and can do so well enough to help states make highly consequential decisions about teacher effectiveness and teachers' professional lives (e.g., merit pay, termination, and tenure).

To date, only six states have *not* applied for the No Child Left Behind (NCLB) waivers (Philips, 2012) Duncan put into place to excuse states from not meeting NCLB's prior goal that 100% of the students in their schools would be academically proficient by the year 2014. The other 44 states and the District of Columbia (D.C.) that have applied for waivers in exchange for even stronger accountability mechanisms, are (or will soon be) using student growth scores as an integral component of their new teacher evaluation systems. States still have local control under such federal requirements; however, they are grappling with how exactly to "meaningfully differentiate [teacher] performance" with an evaluation "including as a significant factor, data on student growth for all students" (U.S. Department of Education [USDOE], 2012c). Additionally, the 18 states, D.C., and 16 districts that have won competitive RttT grants are required to measure teacher effectiveness using student performance data as a "significant factor" (USDOE, 2009, 2012a, 2012b). With limited resources (i.e., time, personnel, money, and human resource capacities), however, the majority of states are opting to use or pilot preexisting growth or value-added models to track the academic growth of students in their states from one year to the next.

While traditional teacher evaluations consist of multiple indicators of teacher effectiveness (e.g., including supervisor observations in which teacher content knowledge, delivery of content, classroom management, organization, etc. are valued), the observation method commonly used to assess such indicators during one or at most a few short visits to a classroom can be subjective. This was demonstrated, for example, in the New Teacher Project's "Widget Effect" report in which researchers found that 99% of teachers whose evaluation reports were examined received ratings of "satisfactory" (Weisberg, Sexton, Mulhern, & Keeling, 2009; see also Corcoran, 2010; Goldhaber & Hansen, 2010; Hanushek, 2011). Accordingly, growth and value-added models that are based on student achievement data are seen and increasingly positioned as a more objective measure of teacher effectiveness. As well, they are being used now more often in conjunction with observations to "better" evaluate teacher effectiveness. Some states, for example, are using growth and value-added estimates to account for as much as 50% of a teacher's total evaluation.

Growth models measure students' progress on standardized test scores from one point to the next in relation to academically similar students; they also help to measure students' progress toward proficiency standards (see, for example, Colorado Department of Education, 2012). Growth models are typically used for more descriptive purposes (Betebenner, 2011; Betebenner & Linn, 2010; Briggs & Betebenner, 2009; Linn, 2008) unless growth estimates correlate well with other measures of teacher effectiveness (Briggs & Betebenner, 2009). Researchers have found varying levels of correlations in their respective studies and are continuously discussing what size correlations might be appropriate or sufficient (Hill, Kapitula, & Umland, 2011; Kane & Staiger, 2012; Kersting, Chen, & Stigler, 2013; Sass & Harris, 2012). However, correlations are not inordinately high, and they are typically no greater than $r = 0.50$. Observational scores typically do not explain more than 25% of the variance in growth and value-added scores, or vice versa.

Value-added models (VAMs) *better* estimate teachers' impacts on student growth over time, and as such are being used by states for more consequential purposes. This is due to VAMs' advanced methodologies and often (but not always) the statistical controls that are used to block, or control for, the student background, risk, and other extraneous variables (e.g., race, ethnicity, gender, poverty, attendance, English proficiency, and involvement in special education, gifted, or other programs) that otherwise make it *impossible* to determine teachers' impact on student growth over time (Baker, 2011, 2012; Harris, 2011; McCaffrey, 2012; Reckase, 2004). Whether this in fact works is not without controversy, however (see, for example, Baker et al., 2010; Darling-Hammond, Amrein-Beardsley, Haertel, & Rothstein, 2012; Darling-Hammond & Haertel, 2012; Sparks, 2011), and is likely one reason that growth model use is more widespread.

SURVEYING THE TERRAIN

Researchers in this study set out to examine if and how all 50 states and Washington D.C. are currently using growth or value-added models to measure teacher effectiveness as a component of teacher evaluations. Researchers did this to help others, particularly educational researchers, practitioners, and policymakers, better contextualize and understand what is happening in this area across the country, as well as explore the national landscape in its current yet rapidly evolving form. It was also the aim of researchers to create one document that encapsulated all of this information for easy consumer use and understanding.

Several organizations have released reports on the various growth and value-added models used by states (e.g., Assessment and Accountability Comprehensive Center [AACC], 2011; Council of Chief State School Officers [CCSSO], 2011; Education Commission of the States [ECS], 2010; National Conference of State Legislatures [NCSL], 2010; National Council on Teacher

Quality [NCTQ], 2011; and Southern Regional Education Board [SREB], 2011); however, these reports are limited due to the use of aggregate data without state-level identification, the inclusion of only some states and not others, and some reports include model information while others focus only on legislation. In addition, none of these reports include information regarding the strengths and weaknesses of the models, for example as interpreted by state personnel, nor the consequences attached to the growth and value-added data generated. Information about how model-generated data are used in practice is also void, which is particularly important as researchers often criticize whether these data indeed help teachers improve their instruction, and, at a grander level, help to increase student learning and achievement (Eckert & Dabrowski, 2010; Harris, 2010, 2011; Lozier, 2012).

As such, this study stands as the first to include all of this information for each state and Washington D.C. in one report. In addition, this is the first study to present more perceptual data, using the impressions and expressed acumen of the representatives working most closely within each state in this area. This is also the first study to look at the consequences attached to model output to discern what is actually occurring across the nation in this area as well. It should be noted, however, that this is an ever-changing territory, and even though researchers verified all final data and findings, it is possible that new developments or state policy shifts may not be represented. Additionally, it should be noted that individual school districts may have teacher evaluations that deviate from the state-level data reported herein, as the state-level data indicates just the minimum requirements to which districts must adhere. Examples of districts often highlighted in the academic literature and popular press include Dallas Independent School District, Houston Independent School District, Milwaukee Public Schools, New York City Public Schools, and Los Angeles Unified School District (see, for example, Amrein-Beardsley & Collins, 2012; Cody, McFarland, Moore, & Preston, 2010; Corcoran, 2010; Goe, 2008; Harris, 2011; Hill et al., 2011; Meyer & Dokumaci, 2010).

To do this, researchers collected data for seven months via multiple phone interviews and emails with state department of education personnel in charge of their own state's initiatives in this area and via multiple websites, most often referenced by each source. Researchers collected information about (a) the type of growth or value-added model used; (b) related legislation, in particular requiring student achievement data to be used to measure teacher effectiveness; (c) the student achievement data used to measure growth or value added; (d) the grade levels and subject areas measured; (e) whether the model accounts for student background and demographic variables; (f) the longitudinal data systems in place to facilitate model calculations; and (g) the percentage (if any) of the teacher evaluation system growth or value-added output that count. In terms of data use, researchers collected information about how growth or value-added data are (h) used in general ways; (i) used in formative ways, and (j) the consequences attached to output. Researchers also collected information about the (k) limitations and (l) strengths of the models, again given the perspectives of the state personnel in charge (see questionnaire protocol in Appendix).

METHODS

Researchers contacted representatives from various positions (Title II A, Accountability & Assessment divisions, Leadership & Evaluation offices, and Teacher Quality or Leadership departments) from each state via email or phone to collect data. Nineteen states and D.C. provided information via phone and email, six states completed the questionnaire by phone, and 21 states participated in the questionnaire solely via email. Despite multiple contact efforts, four states opted to not participate in the study and researchers referred to their department of education websites as sole sources for teacher effectiveness information.¹ After researchers captured all data by state, researchers calculated the forthcoming frequencies and descriptive statistics. All of these data can be found in Table 1. Researchers also collapsed and quantified all free response items (of the 47/51 total respondents who participated either by phone or email [97%]) to represent overarching themes (Miles & Huberman, 1994), specifically in regard to the expressed strengths and weaknesses of the growth and value-added models currently being piloted, implemented, or already in use. These and the descriptive findings are illustrated next.

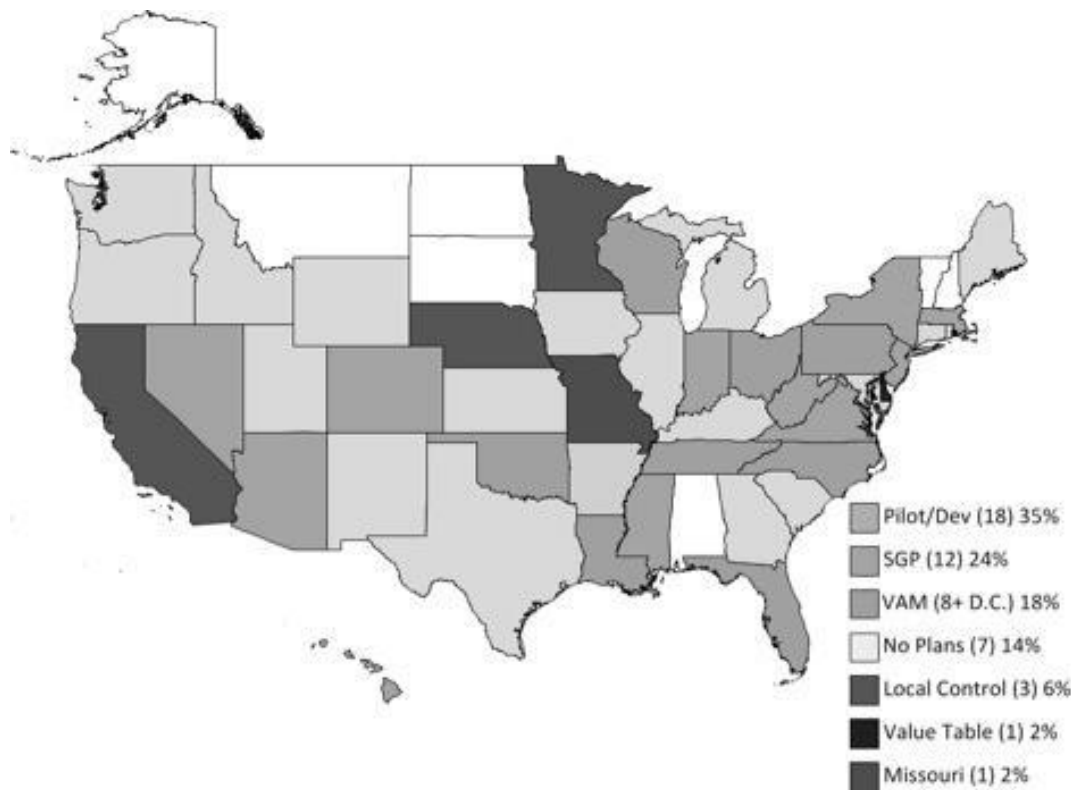
THE NATIONAL LANDSCAPE

The growth and value-added models used across the country include "off the shelf" extant models, the most popular of which include the SAS Education Value-Added Assessment System (EVAAS);² the Student Growth Percentiles (SGP) model (also commonly recognized as the Colorado Growth Model [CGM]);³ the Value-Added Research Center (VARC) model;⁴ homegrown models (e.g., value-table models; see, for example, Delaware Department of Education, 2008); and hybrid versions with state systems incorporating different components from various models, including other VAMs being developed, for example, by the American Institutes for Research (AIR) and Mathematica Policy Research (see, for example, Goldhaber & Theobald, 2012).

Currently, 40 states and D.C. (80%) are using, piloting, or developing, some type of growth or value-added model. The SGP model is the most commonly used model, as it is used or piloted by 12 states (24%);⁵ eight states and D.C. (18%)⁶ are using or

piloting a VAM; Missouri is piloting both the SGP and a VAM; and Delaware is using its value table model. Additionally, 18 states (35%)⁷ indicated that they are piloting or currently developing a model to be used statewide, but they did not specify a particular model. In three states (6%),⁸ growth or value-added model use is locally controlled at the district-level, and seven states (14%)⁹ indicated that they do not currently have plans to develop a statewide growth or value-added model for evaluating teacher effectiveness (see Figure 1), although some are using such measures to evaluate school effectiveness.

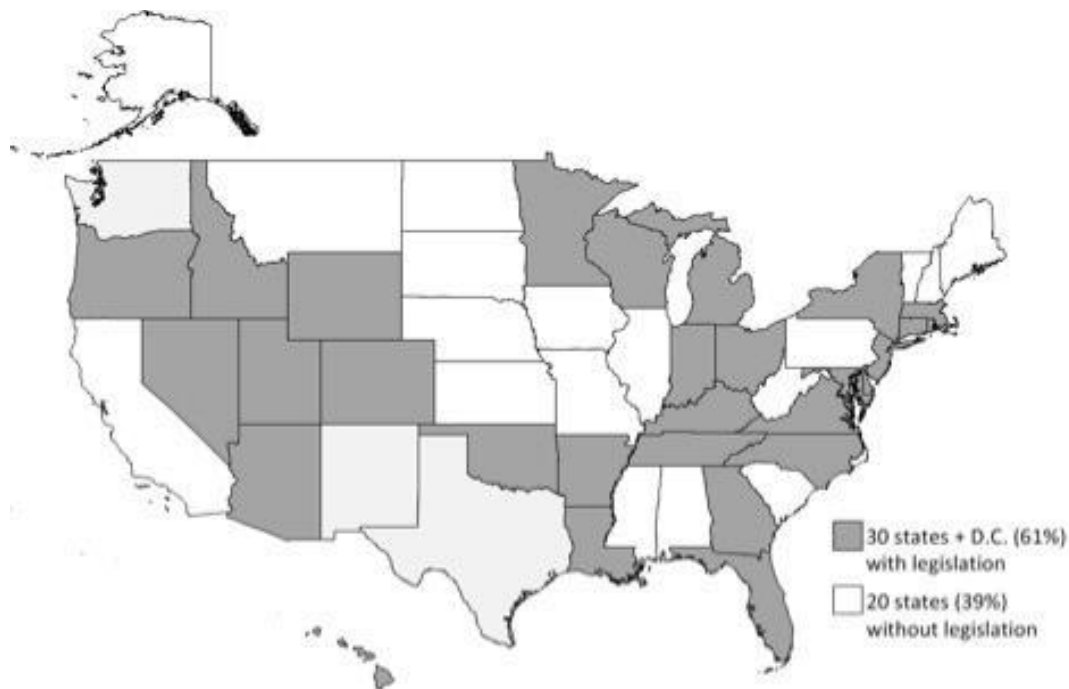
Figure 1. Current and planned growth and VAM use



In order to calculate growth within these models, states need data systems capable of linking individual students with their teachers of record. Since 2005, the Data Quality Campaign (DQC) has supported states' efforts in this area. According to the DQC in 2011, only seven states (14%) had yet to connect teachers of record with their students (DQC, 2012). According to responses collected here, it seems that now only four states (8%)¹⁰ do not yet have such a data system in place, although 12 states (24%)¹¹ are still in the development or piloting stages.

In addition, given the federal requirements for teacher accountability (i.e., NCLB waivers and RttT grant requirements), individual states are also developing state legislation and policies in accord. Results here indicate that 30 states and D.C. (61%) now have legislation or regulations that require student achievement data be used to "significantly" inform the criteria for the evaluation of teacher effectiveness and subsequent decision-making efforts (see Figure 2).

Figure 2. State legislation requiring that teacher evaluation systems use growth or VAM output



In terms of consequences, in four states (8%),¹² teacher consequences attached to growth or value-added data are locally controlled. In 15 states (29%),¹³ respondents indicated that teacher consequences attached to student performance data have yet to be determined. And, in 15 states and D.C. (31%),¹⁴ respondents indicated that teacher consequences are (or will ultimately be) attached and heavily influenced by growth or value-added output as a main component of the overall teacher evaluation. Ten states and D.C. (22%)¹⁵ tie (or are planning to tie) teacher tenure decisions to such output. Related, nine states and D.C. (20%)¹⁶ use (or are planning to use) these data to make teacher termination decisions, although the number of states using growth or value-added estimates to make termination decisions will likely increase as a result of increased state legislation favoring the removal of tenure (see, for example, Banchemo & Kesmodel, 2011; Underwood & Mead, 2012). And, nine states and D.C. (20%)¹⁷ use (or are planning to use) such output to differentiate levels of teacher compensation, award merit pay, or make pay-for-performance decisions. Otherwise, 13 states and D.C. (27%)¹⁸ use (or are planning to use) growth or value-added data to inform professional development efforts.

Thirteen states (25%)¹⁹ indicated that their teacher evaluations include (or will include) multiple measures of *student* growth data beyond value-added or growth scores. South Carolina, for example, uses student work samples, which can be collected from all teachers' classrooms, and not just the typical 30% of the teachers who teach the core curricular areas typically assessed using large-scale standardized tests to measure growth or value-added (Harris, 2011). Maryland is currently using local or school-level data to contribute to student growth calculations. New Jersey also uses school-level data as a measure of student growth, and districts have the option to include additional student achievement data such as portfolios or supplemental assessment data. Kentucky supplements the state's annual test data with interim assessment and student portfolio data. The inclusion of multiple measures of student data is more in line with the field standards developed by the prominent national associations on educational measurement and testing (AERA, APA, & NCME, 2000). It should be mentioned, however, that these standards were never specifically named or mentioned by the state representatives. These standards most importantly note that high-stakes decisions "should not be made on the basis of test scores alone. Other relevant information should be taken into account to enhance the overall validity of such decisions" (AERA, 2000).

Related, researchers also collected information about the tests (by content area and grade level) used for calculating growth or value-added data. Out of the 22 states and D.C. currently using growth and value-added models, 100% of them use their state's large-scale standardized tests in mathematics and English/language arts to generate accountability output. Nine states (18%)²⁰ indicated that they plan to evaluate teacher effectiveness at the high school level, using end-of-course exams. South Carolina is the only state that is thus far evaluating early childhood teachers, using the Northwest Evaluation Association Measures of Academic Progress (NWEA MAP) for grades K–3, although Wisconsin has plans to evaluate K–3 teachers as well.

Finally, state representatives reported whether the growth or value-added model they use accounts or "controls for" student

demographic information (e.g., race, ethnicity, gender, poverty, attendance, English proficiency, and involvement in special education, gifted, or other programs). This is a highly-contested topic, especially between model architects (see, for example, Sanders, Wright, Rivers, & Leandro, 2009; Sanders & Horn, 1998) and critics (see, for example, Amrein-Beardsley, 2008; Amrein-Beardsley & Collins, 2012; Ballou, 2012; Braun, 2005; Cody et al., 2010; Darling-Hammond et al., 2012; Rothstein, 2009). Regardless, 21 states and D.C. (43%)²¹ responded that demographic information is not or will not be accounted for in the growth or value-added model used in their states, six states (12%)²² indicated that demographic information is or will be accounted for, and nine states (18%)²³ indicated that this has yet to be determined. Please see all other state information in Table 1.

Table 1. National Growth and Value-Added Model Summary Statistics

Table 1a. States: Alabama—Missouri

	AL	AK	AZ	AR	CA	CO	CT	DE	FL	GA	HI	ID	IL	IN	IA	KS	KY	LA	ME	MD	MA
Legislation requiring evaluation of teacher effectiveness to be linked to student growth/achievement data	no	no	SEDC	HB 778	no	SB 0049	SB 000	HB 1000	HB 1000	no	no	no	no	no	no	no	no	no	no	no	no
Growth or Value-Added Model Used	no	no	SEP	Value-Added Model	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
Assessment type and grade levels included	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
Account for student demographic information	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
Data system in place to link individual student data with teachers	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
Teacher consequences attached to data output	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
Percentage of teacher evaluation student growth included	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no

Notes. ¹ Indicates data were collected via state websites.

Table 1b. States: Montana—Washington D.C.

	MT	NE	NV	NH	NJ	NM	NY	NC	ND	OH	OK	OR	PA	R	SC	SD	TN	TX	UT	VT	VA
Legislation requiring evaluation of teacher effectiveness to be linked to student growth/achievement data	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
Growth or Value-Added Model Used	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
Assessment type and grade levels included	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
Account for student demographic information	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
Data system in place to link individual student data with teachers	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
Teacher consequences attached to data output	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no
Percentage of teacher evaluation student growth allotted	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no	no

Notes. ¹ Indicates data were collected via state websites.

STRENGTHS AND WEAKNESSES

A unique characteristic of this study is that state department of education personnel, along with the descriptive information they provided, were asked to share their perceptions of the strengths and weaknesses of the models used in their states. Because these models and their respective legislative mandates seemingly exist in a fluid climate where adaptations and changes are constant, this information should also provide valuable information to other states as states continue to develop, adapt, and redesign their growth and value-added approaches. This information should also prove useful to researchers who continue to follow and research the changes in this area.

Regardless of the type of model currently used or in development, most state representatives (71%) expressed, first and foremost, their concerns about assessing student progress for teachers of nontested grades and subject areas. The issue of fairness (see also Darling-Hammond et al., 2012; Glazerman, Goldhaber, Loeb, Raudenbush, Staiger, & Whitehurst, 2011; Goe, 2008) is most troublesome. Recall that 100% of the states currently calculating (or with plans to calculate) these data are using (or are planning to use) their large-scale, state-level, standardized test score data, predominantly collected in grades 4–8 in the core subject areas of mathematics and English/language arts. This means that a minority of teachers, approximately 30% (Harris, 2011), are currently eligible for these types of high-stakes evaluations. As a representative from one state noted, some entire buildings lack these types of scores (e.g., early childhood/primary and high school campuses), making evaluative comparisons nearly impossible, not to mention similar enough to even approach being fair. As a result, this representative said the state had yet to determine how to use growth or value-added models to assess teacher quality across and within districts.

In terms of reliability, some respondents (14%) expressed concerns given the existing research in the literature about the lack of reliability found across many, if not all, growth and value-added models. For example, a teacher classified as effective using these models has a 25–50% chance of being classified as ineffective the following year, and vice versa (Baeder, 2010; Baker et al., 2010; Haertel, 2011; Koedel & Betts, 2007; Newton, Darling-Hammond, Haertel, & Thomas, 2010; Papay, 2010). While not directly asked, respondents expressed freely or of their own accord that they were grappling with such inconsistencies among or across teachers from one year to the next, but did not mention that they were running such analyses. They simply noted they were aware of these reliability issues.

In terms of validity, while four state representatives (8%) boasted about the perceived strengths of their state standardized testing systems, others (6%) expressed concerns about whether the tests used to measure growth or value added were appropriately designed to capture teacher effectiveness over time. Three state representatives (6%) mentioned that their perceived “weak tests” weakened validity, while states with self-reported “stronger tests” felt their tests strengthened model validity. Though not directly asked, representatives did not discuss whether the multiple measures of teacher quality used in their state teacher evaluations contradicted or marginally correlated with growth or value-added output. Issues with alignment may have been underreported because respondents did not mention such concerns freely or of their own accord, although current research on the topic suggests that growth and value-added scores are mildly related to supervisor evaluation scores and even less related to teacher portfolio scores (see, for example, Jacob & Lefgren, 2007; Milanowski, Kimball, & White, 2004; Wilson, Hallman, Pecheone, & Moss, 2007), although both supervisor evaluation and teacher portfolio scores have their own methodological issues as well (e.g., subjectivity).

Other validity-related challenges mentioned (8%) include ensuring proper linkages between students and their teachers of record for accurate data analyses, reporting, and recording. Several state representatives (12%) also pointed out that teaching consists of multifaceted, collective, and cumulative efforts that occur all year long, whereas growth and value-added measures intend to capture teacher effectiveness using only one test every year, or two tests when growth is measured from point x to y under the tutelage of different teachers. To counter this issue, respondents in particular who use (or plan to use) multiple measures of student achievement expanded further, stressing the numerous ways student achievement would be captured and measured in their states beyond using progress on standardized state tests alone.

Growth and value-added models provide measures of teacher effectiveness, but the benefits that are to come from these models, such as improving teacher effectiveness, require formative use of the data. Of the 22 states and D.C. (45%) that have implemented (or are piloting) growth and value-added models, not one state representative was able to articulate a statewide plan for formative data use. In general, and according to respondents, the sophistication and complexity of the statistics used across models made it difficult to clearly explain the process to virtually anyone outside of the creators of the models themselves, including principals and teachers. Regardless, state representatives, at least conceptually, support the formative use of model output as a means to increase teacher quality. For example, one state representative noted that “value-added data used as a formative tool to direct instruction is a powerful tool.” Another called the state model “an excellent tool to help develop strong

teachers.” However, when probed for specific ways teachers were using the data to inform instruction, they were unable to express what this looked like. No states have yet developed statewide plans for the formative use of the growth or value-added output derived.

Otherwise, in terms of model-specific issues, representatives from states using the SGP model (24%) considered the fact that SGP does not require a vertical scale (the process of statistically linking the scores from two or more tests) to be a strength of the model (see also Barlevy & Neal, 2012; Betebenner, 2009; Briggs, 2012; Goldschmidt, Choi, & Beaudoin, 2012). Four representatives (8%) from states using SGP stated that having a growth score generated for every single student allows teachers the opportunity to reflect on their teaching practices at the student level, more so than what they perceived other models did (see also Barlevy & Neal, 2012; Bonk, 2011). In one state, teachers received the SGP data for English/language arts the same spring as their students were assessed. This was noted as an asset in that it allows teachers to, at least conceptually, apply what they learn from data output directly with the students who construct the results.

Personnel representing the four states using the EVAAS (8%) listed as a strength that the model was able to show growth for all achievement levels of students, from the lowest performing to the highest performing students. Like the SGP, this provides information on groups’ of students, individuals’ progress, or their potential “to catch up.” One state also saw predictive value in the EVAAS. The respondent, for example, noted as a strength how the model permits the use of student-level data to predict future achievement, largely to inform the placement of students in courses with appropriate levels of rigor. Additionally, the four representatives using the EVAAS appreciated that the model is able to account for students with missing test score data “in a sophisticated way.” However, a common concern was that the model is proprietary, and although SAS (the sponsoring statistical company) has been willing to share many of the workings of the model with states, there are, in the words of one respondent, “lingering concerns about the transparency of the model.” As another state coordinator expanded, “the calculations of EVAAS results are not easily explained or visible” and that this in itself prohibits the accessibility and usability of the EVAAS system.

Elsewhere, personnel representing states still in the development and piloting phase (35%) stressed their desires to build functional, sustainable, and instrumental models instead of hastily adopting something other states had adopted or endorsed. This is not to say that these states do not want to learn from others, particularly from their successes and mistakes, but they expressed wanting to maintain some level of local control and develop homegrown models using stakeholder input and others’ experiences thus far.

CONCLUSIONS

In sum, researchers found that almost all states are moving forward with growth and value-added models as a key component of their state-based teacher evaluation systems. As well, researchers found that states view teacher evaluation reform as a continuous process, and this includes the states that are already employing growth and value-added models. The percentage of student achievement data accounting for teacher evaluations, as well as the consequences attached to these data, appear to be in a state of flux, still, among the strong majority. Yet, as teacher tenure continues to disintegrate (see, for example, Banchemo & Kesmodel, 2011; Underwood & Mead, 2012), it is likely that more states will increasingly use growth and value-added measures to at least attempt to make more consequential decisions. Whether this will ever be done with enough consistency and accuracy is still a hotly debated issue however (Amrein-Beardsley & Collins, 2012; Baker et al., 2010; Baker, Oluwole, & Green, 2013; Capitol Hill Briefing, 2011; Darling-Hammond et al., 2012; Darling-Hammond & Haertel, 2012; Newton et al., 2010; Rothstein, 2009, 2010; Schochet & Chiang, 2010).

On the flip side, a few state representatives (6%) outright rejected the notion of quantifying teacher effectiveness and vehemently opposed building these models for increased chances at RttT funding. One respondent indicated that her state is interested in taking its time to build a “sustainable system” that will last “without the support of Washington [D.C.]” Another said, “We are fighting against a system that is narrowing everything about education and unfortunately, our current evaluation focus seems to go right along with that narrowness.” However, given that 80% of states are using, piloting, or developing a growth or value-added model, it is clear that while some personnel in these states might have great intentions, they are still up against a strong national movement.

Otherwise, several other findings from this study stand out, findings that should prove useful to state representatives, researchers, and policymakers alike. First and most disturbing is that not one state representative was able to articulate a plan for formative use of growth or value-added data. In other words, states are assuming that implementation of such models automatically implies data utilization by teachers. Much more attention and research is needed to investigate and learn if and how teachers are using growth and value-added data, and how states and districts are putting in place the resources necessary for teachers to use these

data in formative ways.

To date, there is no research evidence that demonstrates that providing teachers with increased access to growth or value-added estimates will increase teachers' abilities to understand or use this information in instructionally meaningful ways (see also Braun, 2008; Briggs & Domingue, 2011; Graue, Delaney, Karch, & Romero, 2011; Kraemer, 2011). Though some (including growth and value-added contractors) have attempted to provide instructions for teachers and administrators on how to use value-added estimates in meaningful ways, for example, by stressing the importance of building a culture of trust with teachers and emphasizing transparency and open communication (see, for example, Harris, 2011; Kennedy, Peters, & Thomas, 2012), no research evidence suggests that doing any of this works to improve instruction, much less increase subsequent levels of student achievement or growth. In many states, the luxury of time and the resources to educate administrators and teachers about the models and data use were not realities before implementation began.

Second, it is evident that even though state representatives are aware that growth and value-added models can only be used for approximately 30% of all teachers, the nation continues to implement them regardless of issues with fairness (see also Baker et al., 2010; Buddin, 2011; Glazerman et al., 2011; Goldhaber & Hansen, 2010; Kersting et al., 2013). Putting aside the numerous concerns surrounding the reliability and validity of growth and value-added models, the fact that approximately two-thirds of all teachers cannot be evaluated by these models should demand continued research and investigation on how to more accurately evaluate *all* teachers. Some believe the implementation of the Common Core State Standards and its associated tests meant to evaluate teachers of noncore subject areas will alleviate such issues with fairness. Although, with this comes a "devil's bargain" by which making things fairer might ultimately create excessively more standardized testing than there already is (e.g., more than 200 tests throughout grades 2 through 12; see, for example, Richardson, 2012). This is not to mention the other fairness-related issues with class sizes (Baker et al., 2010; Brophy, 1973; Buddin, 2011; Goldhaber & Hansen, 2010; Kupermintz, 2003; Sanders, Saxton, & Horn, 1997) and the relative sizes of comparison schools (Linn & Haug, 2002; Newton et al., 2010; Reardon & Raudenbush, 2009) that will likely prevail regardless of the Common Core State Standards.

Nonetheless, at a time when almost all states are employing, piloting, or developing growth and value-added models to help them measure teacher effectiveness, and in 15 states and D.C. to inform highly consequential decisions, findings from this study should provide a good, one-stop resource to see what each state already has in place or in development. Findings should also provide information about how model data are (or will be) used to evaluate teachers, the consequences associated with outcome data, and the perceived and purported strengths and weaknesses associated with the various models. As well, and for the time being, this study provides the most inclusive national growth model overview for practitioner, researcher, and policymaker consumption and use. Results from this study should inform others about the important factors they might consider when building teacher evaluation systems based, at least in part, on growth or value-added data.

Notes

1. Illinois, New Jersey, New York, Oklahoma
2. For more information about the SAS[®] EVAAS[®], please see <http://www.sas.com/govedu/edu/k12/evaas/index.html>
3. For more information about the SGP, please see Betebenner, 2009 and Betebenner & Linn, 2010. The RAND Corporation also offers a comprehensive overview (<http://www.rand.org.proxy.library.vanderbilt.edu/education/projects/measuring-teacher-effectiveness/student-growth-percentiles.html>). For more information about the SGP as per SGP users, please see the information constructed by states including Colorado, the state in which the SGP originated (<http://www.schoolview.org/ColoradoGrowthModel2.asp>); Rhode Island (<http://www.ride.ri.gov/assessment/RIGM.aspx>); and West Virginia (<http://wvde.state.wv.us/growth/>).
4. For more information about VARC, please see <http://varc.wceruw.org/>
5. Arizona, Colorado, Hawaii, Indiana, Massachusetts, Mississippi, Nevada, New Jersey, New York, Rhode Island, Virginia, West Virginia
6. Florida, Louisiana, North Carolina, Ohio, Oklahoma, Pennsylvania, Tennessee, Wisconsin
7. Arkansas, Connecticut, Georgia, Iowa, Idaho, Illinois, Kansas, Kentucky, Maine, Maryland, Michigan, New Mexico, Oregon, South Carolina, Texas, Utah, Washington, Wyoming
8. California, Minnesota, Nebraska
9. Alabama, Alaska, Montana, New Hampshire, North Dakota, South Dakota, Vermont
10. Alaska, Iowa, Nebraska, Vermont
11. Colorado, Illinois, Maine, Maryland, Minnesota, Montana, New Jersey, North Dakota, Oregon, Pennsylvania, South Dakota, Wisconsin
12. Arizona, Colorado, Massachusetts, New Jersey

13. Arkansas, Georgia, Idaho, Illinois, Mississippi, Missouri, Nebraska, New Mexico, Ohio, Oklahoma, Oregon, Pennsylvania, Washington, Wisconsin, Wyoming
14. Florida, Hawaii, Indiana, Kentucky, Louisiana, Maine, Maryland, Michigan, Minnesota, New York, North Carolina, South Carolina, Rhode Island, Tennessee, Virginia
15. Florida, Hawaii, Indiana, Kentucky, Louisiana, Michigan, Minnesota, New York, Rhode Island, Tennessee
16. Hawaii, Kentucky, Louisiana, Maryland, Michigan, Minnesota, New York, Rhode Island, Tennessee
17. Florida, Indiana, Maine, Maryland, New York, North Carolina, South Carolina, Tennessee, Virginia
18. Florida, Hawaii, Louisiana, Maine, Maryland, Michigan, New Jersey, New York, North Carolina, Ohio, Rhode Island, South Carolina, Tennessee
19. Georgia, Hawaii, Kentucky, Maryland, Nevada, New Hampshire, New Jersey, North Carolina, South Carolina, Virginia, Washington, West Virginia, Wisconsin
20. Florida, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, Utah, Virginia, Wisconsin
21. Arizona, California, Colorado, Florida, Hawaii, Indiana, Massachusetts, Minnesota, Nebraska, Nevada, North Carolina, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, Tennessee, Virginia, West Virginia, Wyoming
22. Delaware, Iowa, Louisiana, Missouri, Utah, Wisconsin
23. Arkansas, Georgia, Maryland, Michigan, Mississippi, New Mexico, New York, Texas, Washington

References

- American Educational Research Association (AERA). (2000, July). *AERA position statement on high-stakes testing in Pre-K–12 education*. Retrieved from <http://www.aera.net.proxy.library.vanderbilt.edu/AboutAERA/AERARulesPolicies/AERAPolicyStatements/PositionStatementonHighStakesTesting/tabid/11083/Default.aspx>
- American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2000). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Amrein-Beardsley, A. (2008). Methodological concerns about the education value-added assessment system. *Educational Researcher*, 37(2), 65–75. doi: 10.3102/0013189X08316420
- Amrein-Beardsley, A., & Collins, C. (2012). The SAS Education Value-Added Assessment System (SAS[®] EVAAS[®]) in the Houston Independent School District (HISD): Intended and unintended consequences. *Education Policy Analysis Archives*, 20(12). Retrieved from <http://epaa.asu.edu.proxy.library.vanderbilt.edu/ojs/article/view/1096>
- Assessment and Accountability Comprehensive Center (AACC). (2011). *Measuring teacher effectiveness: An overview of state policies and practices related to pre-K–12 teacher effectiveness or teacher evaluation*. Washington, DC: Gallagher, Rabinowitz, & Yeagley.
- Baeder, J. (2010, December 21). Gates' measures of effective teaching study: More value-added madness. *Education Week*. Retrieved from http://blogs.edweek.org.proxy.library.vanderbilt.edu/edweek/on_performance/2010/12/gates_measures_of_effective_teaching_study_more_value-added_madness.html
- Baker, B. D. (2011). Take your SGP and VAMit, Damn it! [Web log post]. Retrieved from <http://schoolfinance101.wordpress.com/2011/09/02/take-your-sgp-and-vamit-damn-it/>
- Baker, B. D. (2012, March 31). Firing teachers based on bad (VAM) versus wrong (SGP) measures of effectiveness: Legal note [Web log post]. Retrieved from <http://schoolfinance101.wordpress.com/2012/03/31/firing-teachers-based-on-bad-vam-versus-wrong-sgp-measures-of-effectiveness-legal-note>
- Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., . . . Shepard, L. A. (2010). *Problems with the use of student test scores to evaluate teachers*. Washington, DC: Economic Policy Institute. Retrieved from <http://www.epi.org/publications/entry/bp278>
- 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). Retrieved from <http://epaa.asu.edu.proxy.library.vanderbilt.edu/ojs/article/view/1298/1043>

Ballou, D. (2012). Review of “The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood.” [Review of the report *The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood*, by R. Chetty, J. Friedman, & J. Rockoff]. Boulder, CO: National Education Policy Center. Retrieved from <http://nepc.colorado.edu.proxy.library.vanderbilt.edu/thinktank/review-long-term-impacts>

Banchero, S., & Kesmodel, D. (2011, September 13). Teachers are put to the test: More states tie tenure, bonuses to new formulas for measuring test scores. *The Wall Street Journal*. Retrieved from <http://online.wsj.com.proxy.library.vanderbilt.edu/article/SB10001424053111903895904576544523666669018.html>

Barlevy, G., & Neal, D. (2012). Pay for percentile. *American Economic Review*, 102(5), 1805–1831. doi: 10.1257/aer.102.5.1805

Barnett, J. H., & Amrein-Beardsley, A. (2011). Actions over credentials: Moving from highly qualified to measurably effective [Commentary]. *Teachers College Record*. Retrieved from <http://www.tcrecord.org.proxy.library.vanderbilt.edu/Content.asp?ContentID=16517>

Betebenner, D. W. (2009). Norm- and criterion-referenced student growth. *Education Measurement: Issues and Practice*, 28(4), 42–51. doi:10.1111/j.1745-3992.2009.00161.x

Betebenner, D. W. (2011, April). *Student growth percentiles*. National Council on Measurement in Education (NCME) training session presented at the annual conference of the American Educational Research Association (AERA), New Orleans, LA.

Betebenner, D. W., & Linn, R. L. (2010). *Growth in student achievement: Issues of measurement, longitudinal data analysis, and accountability*. Exploratory Seminar: Measurement Challenges Within the Race to the Top Agenda: Center for K–12 Assessment and Performance Management. Retrieved from www.k12center.org/rsc/pdf/BetebennerandLinnPolicyBrief.pdf

Bonk, W. (2011). *District and school use of Colorado Growth Model results*. Report prepared for the Unit of Accountability and Improvement. Denver, CO: Colorado Department of Education. Retrieved from www.cde.state.co.us/.../GrowthStandardsAccountability.pdf

Braun, H. I. (2005). *Using student progress to evaluate teachers: A primer on value-added models*. Princeton, NJ: Educational Testing Service.

Braun, H. I. (2008). *Vicissitudes of the validators*. Presentation made at the 2008 Reidy Interactive Lecture Series, Portsmouth, NH. Retrieved from www.cde.state.co.us/cdedocs/OPP/HenryBraunLectureReidy2008.ppt

Briggs, D. (2012, April). Understanding the Colorado Growth Model. *New Jersey Education Association (NJEA) Review*. Retrieved from <http://www.njea.org/news-and-publications/njea-review/april-2012/understanding-the-colorado-growth-model>

Briggs, D. C., & Betebenner, D. (2009). *Is growth in student achievement scale dependent?* Paper presented at the annual meeting of the National Council for Measurement in Education (NCME), San Diego, CA.

Briggs, D., & Domingue, B. (2011, February). 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 from nepc.colorado.edu/publication/due-diligence

Brophy, J. (1973). Stability of teacher effectiveness. *American Educational Research Journal*, 10(3), 245–252. doi:10.2307/1161888

Buddin, R. J. (2011). *How effective are Los Angeles elementary teachers and schools?* MPRA Paper, University Library of Munich, Germany. Retrieved from <http://econpapers.repec.org/paper/pramprapa/27366.htm>

Capitol Hill Briefing. (2011, September 14). *Getting teacher evaluation right: A challenge for policy makers*. A briefing by E. Haertel, J. Rothstein, A. Amrein-Beardsley, and L. Darling-Hammond. Washington DC: Dirksen Senate Office Building (research in brief). Retrieved from <http://www.aera.net.proxy.library.vanderbilt.edu/Default.aspx?id=12856>

- Cody, C. A., McFarland, J., Moore, J. E., & Preston, J. (2010, August). *The evolution of growth models*. Public Schools of North Carolina. Raleigh, NC. Retrieved from <http://www.dpi.state.nc.us/docs/intern-research/reports/growth.pdf>
- Colorado Department of Education. (2012). *The Colorado Growth Model: Explore growth and achievement of Colorado districts and schools*. Denver, CO. Retrieved from <http://www.schoolview.org/ColoradoGrowthModel2.asp>
- Corcoran, S. P. (2010). *Can teachers be evaluated by their students' test scores? Should they be? The use of value-added measures of teacher effectiveness in policy and practice*. Providence, RI: Annenberg Institute for School Reform. Retrieved from <http://www.annenberginstitute.org/products/Corcoran.php>
- Council of Chief State School Officers (CCSSO). (2011). *Interstate teacher assessment and support consortium (InTASC) model core teaching standards: A resource for state dialog*. Washington, DC: Council of Chief State School Officers. Retrieved from www.ccsso.org/.../intasc_model_core_teaching_standards_2011.pdf
- Darling-Hammond, L., Amrein-Beardsley, A., Haertel, E., & Rothstein, J. (2012). Evaluating teacher evaluation. *Phi Delta Kappan*, 93(6), 8–15.
- Darling-Hammond, L., & Haertel, E. (2012, November 5). A better way to grade teachers. *Los Angeles Times* [op-ed]. Retrieved from <http://www.latimes.com/news/opinion/commentary/la-oe-darling-teacher-evaluations-20121105,0,650639.story>
- Data Quality Campaign. (2012). *Data for action 2011: Empower with data*. Washington, DC. Retrieved from dataqualitycampaign.org/files/DFA2011%20Annual%20Report.pdf
- Delaware Department of Education. (2008, August 1). *Delaware's growth model for AYP determinations*. Dover, DE. Retrieved from <http://www.doe.k12.de.us/aab/files/2008%20School%20Accountability%20-%20Growth%20Model%20Presentation%208-1-08.pdf>
- Eckert, J., & Dabrowski, J. (2010). Should value-added measures be used for performance pay? *Phi Delta Kappan*, 91(8), 88–92.
- Education Commission of the States (ECS). (2010). *Teacher evaluation: New approaches for a new decade*. Denver, CO. Retrieved from www.ecs.org/html/Document.asp?chouseid=8621
- Felch, J., Song, J., & Smith, D. (2010, August 14). Who's teaching L.A.'s Kids? *Los Angeles Times*. Retrieved from: <http://www.latimes.com/news/local/la-me-teachers-value-20100815,0,258862,full.story>
- Glazerman, S., Goldhaber, D., Loeb, S., Raudenbush, S., Staiger, D. O., & Whitehurst, G. J. (2011, April 26). *Passing muster: Evaluating teacher evaluation systems*. Washington, DC: The Brookings Institution. Retrieved from www.brookings.edu/reports/2011/0426_evaluating_teachers.aspx
- Goe, L. (2008). *Using value-added models to identify and support highly effective teachers*. Washington DC: National Comprehensive Center for Teacher Quality. Retrieved from <http://www2.tqsource.org/strategies/het/UsingValueAddedModels.pdf> on 2 June 2010
- Goldhaber, D., & Hansen, M. (2010). *Is it just a bad class? Assessing the stability of measured teacher performance* (CEDR Working Paper 2010-3). Seattle, WA. Retrieved from <http://www.cedr.us/publications.html>
- Goldhaber, D., & Theobald, R. (2012, October 15). Do different value-added models tell us the same things? *Carnegie Knowledge Network*. Retrieved from <http://www.carnegieknowledgenetwork.org/briefs/value-added/different-growth-models/>
- Goldschmidt, P., Choi, K., & Beaudoin, J. B. (2012, February). *Growth model comparison study: Practical implications of alternative models for evaluating school performance*. Washington, DC: Council of Chief State School Officers. Retrieved from http://www.ccsso.org/Resources/Publications/Growth_Model_Comparison_Study_Practical_Implications_of_Alternative_Models_for_Evaluating_School_Performance_.html
- Gonen, Y. (2012, February 24). NYC makes internal ratings of 18,000 public school teachers available. *The New York Post*. Retrieved from <http://www.nypost.com/p/news/local>

/nyc_makes_internal_ratings_of_public_4nzYXTN1L4LQXU17G2YktO#ixzz1okkfxFDT

- Graue, B., Delaney, K., Karch, A., & Romero, C. (2011, April). *Data use as a reform strategy*. Paper presented at the annual convention of the American Educational Research Association (AERA).
- Haertel, E. (2011). *Using student test scores to distinguish good teachers from bad*. Paper presented at the annual conference of the American Educational Research Association (AERA), New Orleans, LA.
- Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education Review*, 30, 466–479.
- Harris, D. N. (2010). Clear away the smoke and mirrors of value-added. *Phi Delta Kappan*, 91(8), 66–69.
- Harris, D. N. (2011). *Value-added measures in education: What every educator needs to know*. Cambridge, MA: Harvard Education Press.
- Hill, H. C., Kapitula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, 48, 794–831.
- Jacob, B. A., & Lefgren, L. (2007, June). *Can principals identify effective teachers? Evidence on subjective performance evaluation in education*. Brigham Young University. Retrieved from economics.byu.edu/Documents/Lars%20Lefgren/.../principals.pdf
- Kane, T., & Staiger, D. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Seattle, WA: Bill and Melinda Gates Foundation.
- Kennedy, K., Peters, M., & Thomas, M. (2012). *How to use value-added analysis to improve student learning: A field guide for school and district leaders*. Thousand Oaks, CA: Sage.
- Kersting, N. B., Chen, M., & Stigler, J. W. (2013). Value-added teacher estimates as part of teacher evaluations: Exploring the effects of data and model specifications on the stability of teacher value-added scores. *Education Policy Analysis Archives*, 21(7). Retrieved from <http://epaa.asu.edu.proxy.library.vanderbilt.edu/ojs/article/view/1167/1049>
- Koedel, C., & Betts, J. R. (2007, April). *Re-examining the role of teacher quality in the educational production function* (Working Paper No. 2007-03). Nashville, TN: National Center on Performance Initiatives.
- Kraemer, S. (2011, April). *A human factors engineering framework for effective data use in education reform and accountability*. Paper presented at the annual convention of the American Educational Research Association (AERA).
- Kupermintz, H. (2003). Teacher effects and teacher effectiveness: A validity investigation of the Tennessee Value-Added Assessment System. *Educational Evaluation and Policy Analysis*, 25, 287–298. doi:10.3102/01623737025003287
- Linn, R. L. (2008). Methodological issues in achieving school accountability. *Journal of Curriculum Studies*, 40, 699–711. doi:10.1080/00220270802105729
- Linn, R. L., & Haug, C. (2002). Stability of school-building accountability scores and gains. *Educational Evaluation and Policy Analysis*, 24, 29–36. doi:10.3102/01623737024001029
- Lozier, C. (2012, July 18). What the PGA can teach us about value-added modeling [Web log post]. Retrieved from <http://gettingsmart.com/blog/2012/07/what-pga-can-teach-us-about-value-added-modeling/>
- McCaffrey, D. F. (2012, October 15). Do value-added methods level the playing field for teachers? *Carnegie Knowledge Network*. Retrieved from <http://www.carnegieknowledgenetwork.org/briefs/value-added/level-playing-field>
- Meyer, R. H., & Dokumaci, E. (2010, July). *Value-added models and the next generation of assessments*. Center for K–12 Assessment & Performance Management. Retrieved from <http://www.k12center.org/rsc/pdf/MeyerDokumaciPresenterSession4.pdf>

- Milanowski, A., Kimball, S. M., & White, B. (2004). *The relationship between standards-based teacher evaluation scores and student achievement: Replication and extensions at three sites*. Madison, WI: University of Wisconsin-Madison, Center for Education Research, Consortium for Policy Research in Education. Retrieved from www.cpre-wisconsin.org/papers/3site_long_TE_SA_AERA04TE.pdf
- Miles, M., & Huberman, A. M. (1994). *Qualitative Data Analysis*. Thousand Oaks, CA: Sage Publications.
- National Conference of State Legislatures (NCSL). (2010). Educators (teachers/principals) 2010 enacted evaluation legislation [Web page]. Washington, DC. Retrieved from <http://www.ncsl.org/default.aspx?tabid=21155>
- National Council on Teacher Quality (NCTQ). (2011). *State of the states: Trends and early lessons on teacher evaluation and effectiveness policies*. Washington, DC. Retrieved from www.nctq.org/p/publications/docs/nctq_stateOfTheStates.pdf
- Newton, X. A., Darling-Hammond, L., Haertel, E., & Thomas, E. (2010). Value-added modeling of teacher effectiveness: An exploration of stability across models and contexts. *Education Policy Analysis Archives*, 18(23). Retrieved from <http://epaa.asu.edu.proxy.library.vanderbilt.edu/ojs/article/view/810>
- Papay, J. P. (2010). Different tests, different answers: The stability of teacher value-added estimates across outcome measures. *American Educational Research Journal*, 48(1), 163–193. doi:10.3102/0002831210362589
- Philips, R. H. (2012, Sep. 10). No child left behind waiver sought by Alabama, six other states. *Press Register*. Retrieved from http://blog.al.com/live/2012/09/no_child_left_behind_waiver_so.html
- Reardon, S. F., & Raudenbush, S. W. (2009). Assumptions of value-added models for estimating school effects. *Education Finance and Policy*, 4(4), 492–519. doi:10.1162/edfp.2009.4.4.492
- Reckase, M. D. (2004). The real world is more complicated than we would like. *Journal of Educational and Behavioral Statistics*, 29(1), 117–120. doi:10.3102/10769986029001117
- Richardson, W. (2012, September 27). Do parents really want more than 200 separate state-mandated assessments for their children? *The Huffington Post*. Retrieved from http://www.huffingtonpost.com/will-richardson/do-parents-really-want-ov_b_1913704.html
- Rothstein, J. (2009). Student sorting and bias in value-added estimation: Selection on observables and unobservables. *Education Finance and Policy*, (4)4, 537–571. doi:<http://dx.doi.org.proxy.library.vanderbilt.edu/10.1162/edfp.2009.4.4.537>
- Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *Quarterly Journal of Economics*, 175–214. doi:10.1162/qjec.2010.125.1.175
- Rubenstein, G. (2012). Analyzing released NYC value-added data [parts 1–4]. *Teach for Us*. Retrieved from <http://garyrubenstein.teachforus.org/2012/02/26/analyzing-released-nyc-value-added-data-part-1/>
- Sanders, W. L., & Horn, S. (1998). Research findings from the Tennessee Value-Added Assessment System (TVAAS) database: Implications for educational evaluation and research. *Journal of Personnel Evaluation in Education*, 12(3), 247–256.
- Sanders, W. L., Saxton, A. M., & Horn, S. P. (1997). The Tennessee Value-Added Accountability System: A quantitative, outcomes-based approach to educational assessment. In J. Millman (Ed.), *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* (pp. 137–162). Thousand Oaks, CA: Corwin Press.
- Sanders, W. L., Wright, S. P., Rivers, J. C., & Leandro, J. G. (2009, November). *A response to criticisms of SAS EVAAS*. Retrieved from http://www.sas.com/resources/asset/Response_to_Criticisms_of_SAS_EVAAS_11-13-09.pdf
- Sass, T., & Harris, D. (2012). *Skills, productivity and the evaluation of teacher performance* (W. J. Usery Workplace Research Group Paper No. 2012-3-1). Retrieved from <http://ssrn.com.proxy.library.vanderbilt.edu/abstract=2020717>

Schochet, P. Z., & Chiang, H. S. (2010). *Error rates in measuring teacher and school performance based on student test score gains*. Washington, DC: National Center for Educational Evaluation and Regional Assistance, Institute of Education Sciences, United States Department of Education.

Southern Regional Education Board (SREB). (2011). *Focus on teacher reform legislation in SREB states: Evaluation policies*. Atlanta, GA. Retrieved from http://www.sreb.org/cgi-bin/MySQLdb?VIEW=/public/docs/view_one.txt&docid=1599

Sparks, S. D. (2011, November 15). Value-added formulas strain collaboration. *Education Week*. Retrieved from <http://www.edweek.org.proxy.library.vanderbilt.edu/ew/articles/2011/11/16/12collab-changes.h31.html?tkn=OVMFb8PQXxQi4wN6vpelNlr7%2BNhOFCbi71mI&intc=es>

Underwood, J., & Mead, J. F. (2012, February 29). A smart ALEC threatens public education. *Phi Delta Kappan*. Retrieved from http://www.edweek.org.proxy.library.vanderbilt.edu/ew/articles/2012/03/01/kappan_underwood.html

U.S. Department of Education. (2009). *Race to the Top program: Executive summary*. Washington, DC. Retrieved from <http://www2.ed.gov.proxy.library.vanderbilt.edu/programs/racetothetop/executive-summary.pdf>

U.S. Department of Education (2012a). *Awards*. Washington DC. Retrieved from <http://www2.ed.gov.proxy.library.vanderbilt.edu/programs/racetothetop/awards.html>

U.S. Department of Education (2012b). *Education department announces 16 winners of Race to the Top—District competition* [Web page]. Washington DC. Retrieved from <http://www.ed.gov.proxy.library.vanderbilt.edu/news/press-releases/education-department-announces-16-winners-race-top-district-competition>

U.S. Department of Education. (2012c). *ESEA flexibility* [Web page]. Washington DC. Retrieved from <https://www-ed.gov.proxy.library.vanderbilt.edu/esea/flexibility>

Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. Brooklyn, NY: The New Teacher Project.

Wilson, M., Hallman, P. J., Pecheone, R., & Moss, P. (2007, October). *Using student achievement test scores as evidence of external validity for indicators of teacher quality: Connecticut's Beginning Educator Support and Training [BEST] Program*. Stanford, CA. Retrieved from <http://edpolicy.stanford.edu.proxy.library.vanderbilt.edu/pages/pubs/pubs.html>

APPENDIX

Question Protocol (for phone interviews and/or emails)

1. Contact Info (name, email, phone, title):
2. Does your state have any legislation requiring the evaluation of teacher effectiveness?
3. What growth/value-added model is your state currently using?
4. What assessment are you using to measure growth/value added? Which grade levels does this include?
5. Does your model account for student demographic information?
6. Does your state currently have a data system in place to track growth/value added? What kind of data system?
7. How are the results or outcomes from the growth/value added measure currently being used? What consequences are being attached to the output? (Ex: bonuses, termination, etc.)
8. What percentage (if any) of the teacher evaluation does growth/value added account for?
9. Are teachers using growth/value added in formative ways? How so?
10. What limitations or weaknesses does the model have?
11. What strengths does the model have?

Cite This Article as: *Teachers College Record* Volume 116 Number 1, 2014, p. -

<http://www.tcrecord.org.proxy.library.vanderbilt.edu/library> ID Number: 17291, Date Accessed: 2/17/2014 5:00:46 PM