

February 2019

## Tennessee's National Impact on Teacher Evaluation Law & Policy: An Assessment of Value-Added Model Litigation

Mark A. Paige

*University of Massachusetts - Dartmouth*

Audrey Amrein-Beardsley

*Mary Lou Fulton Teachers College*

Kevin Close

*Arizona State University*Follow this and additional works at: <https://trace.tennessee.edu/tjlp>Part of the [Education Law Commons](#), and the [Law and Politics Commons](#)

### Recommended Citation

Paige, Mark A.; Amrein-Beardsley, Audrey; and Close, Kevin (2019) "Tennessee's National Impact on Teacher Evaluation Law & Policy: An Assessment of Value-Added Model Litigation," *Tennessee Journal of Law and Policy*. Vol. 13 : Iss. 2 , Article 4.

Available at: <https://trace.tennessee.edu/tjlp/vol13/iss2/4>

This Article is brought to you for free and open access by Volunteer, Open Access, Library Journals (VOL Journals), published in partnership with The University of Tennessee (UT) University Libraries. This article has been accepted for inclusion in Tennessee Journal of Law and Policy by an authorized editor. For more information, please visit <https://trace.tennessee.edu/tjlp>.

# TENNESSEE JOURNAL OF LAW AND POLICY

---

VOLUME 13

WINTER 2019

ISSUE 2

---

## ARTICLE

### TENNESSEE'S NATIONAL IMPACT ON TEACHER EVALUATION LAW & POLICY

AN ASSESSMENT OF VALUE-ADDED MODEL  
LITIGATION

*Mark A. Paige\**  
*Audrey Amrein-Beardsley\*\**  
*Kevin Close†*

#### **Abstract**

*Over the last decade or so, federal and state education policymakers embraced the use of value-added models (VAMs) to evaluate teachers' performance and make high-stakes employment decisions (e.g., tenure, merit pay,*

---

\* Mark A. Paige, J.D., Ph.D., is an associate professor of public policy at the University of Massachusetts-Dartmouth. Before becoming a professor, he was an education law attorney representing school districts.

\*\* Audrey Amrein-Beardsley, Ph.D., is currently a Professor in Mary Lou Fulton Teachers College. Her research interests include educational policy, educational measurement and research methods.

† Kevin Close, M.Ed., is a doctoral candidate and graduate assistant researcher at the Arizona State University.

*termination of employment). VAMs are complicated statistical models that attempt to estimate a teacher's contribution to student test scores, particularly those in mathematics and reading. Educational researchers, as well as many teachers and unions, however, have objected to the use of VAMs noting that these models fail to adequately account for variables outside of teachers' control that contribute to a student's education performance. Subsequently, many teachers challenged the use of VAMs through the courts. This article assesses those challenges.*

I. Introduction	525
II. VAMS: Promise and Controversy	530
A. A Brief History of VAMs in Educational Policy	530
1. The Rise of VAMs in National Education Policy: Race to the Top	532
B. Statistical and Practical Controversies	535
1. Reliability	535
2. Validity	537
3. Bias	538
4. Transparency	540
5. Fairness	542
6. Consequential Use	543
7. Intended Consequences	544
8. Unintended Consequences	545
III. The Cases	547
A. Federal Substantive Due Process Rights & Equal Protection Arguments: VAMs May Be Unwise But Still Constitutional	547
1. <i>Cook v. Bennett</i>	547
2. <i>Trout v. Knox County Board of Education</i>	551
3. <i>Wagner v. Haslam</i>	553
4. <i>Matter of Lederman v. King</i>	554
B. Legislative State Agency Authority Questioned	555
1. <i>Leff v. Clark County School District</i>	555
2. <i>Stapleton v. Skandera</i>	557
3. <i>Louisiana Federation of Teachers v. State</i>	559

4. <i>Robinson v. Stewart</i>	560
5. Filed but not Adjudicated	561
C. Process & “Fundamental Fairness” Cases	563
1. <i>Houston Federation of Teachers</i>	563
2. <i>Washington Teachers’ Union v. D.C. Public Schools</i>	566
IV. Current Policy Landscape in Wake of ESSA	569
A. ESSA Reauthorization	570
B. State Plans	572
V. Conclusions	573

## I. Introduction

In March of 2017, William “Bill” Sanders passed away in Tennessee.<sup>1</sup> To most policymakers outside of education (and many within it) he was a relatively unknown statistician. His work in education policy started far away from schoolhouses. Indeed, after he received his degree in statistics at the University of Tennessee, he began assessing the impact of radiation on farm animals.<sup>2</sup>

But his career trajectory changed markedly. In 1982, after reading a newspaper article about how Tennessee Governor Lamar Alexander sought a model of teacher compensation that would pay teachers for performance, Mr. Sanders concluded he had the answer.<sup>3</sup> He wrote to Alexander explaining that he developed a statistical model that could determine who the “best” teachers were—a so-called “value-added” model (e.g., the Tennessee Value-Added Assessment System (TVAAS))

---

<sup>1</sup> Kevin Carey, *The Little-Known Statistician Who Taught Us to Measure Teachers*, N.Y. TIMES (May 19, 2017), <https://www.nytimes.com/2017/05/19/upshot/the-little-known-statistician-who-transformed-education.html> [<https://perma.cc/2VBF-CZWY>].

<sup>2</sup> *Id.*

<sup>3</sup> *Id.*

which is more generally known as the Education Value-Added Assessment (EVAAS)).<sup>4</sup> This model estimates a teacher's contribution to student achievement on standardized tests,<sup>5</sup> and it formed the basis for his private company that developed algorithms for the models.<sup>6</sup> Tennessee ultimately incorporated value added models into policies and laws, linking high-stakes employment decisions and evaluation to student test scores.<sup>7</sup>

Mr. Sanders's models—sparked by this random collision of events—has had profound impact on national educational policy. In 2009, President Obama's Race to the Top (RttT) program conditioned state receipt of federal education dollars on states' use of VAMs to evaluate and make employment decisions for teachers. States seeking much-needed federal money during the

---

<sup>4</sup> *Id.* VAMs have a policy history that precede Mr. Sanders's adoption of the term in education. They had been used in economics since the 1960s. *See, e.g.,* Douglas Harris, *Would Accountability Based on Teacher Value Added Be Smart Policy? An Examination of the Statistical Properties and Policy Alternatives*, 4 J. EDUC. FIN. & POL'Y 319, 321 (2009). Yet Sanders is widely credited as the one who popularized the use of VAMs for educational accountability. *E.g.,* Carey *supra* note 1.

<sup>5</sup> *E.g.,* EDWARD WILEY, A PRACTITIONER'S GUIDE TO VALUE ADDED ASSESSMENT 5 (2006) <https://nepc.colorado.edu/publication/a-practitioners-guide-value-added-assessment-educational-policy-studies-laboratory-resea> [<https://perma.cc/EH6R-S7QN>].

<sup>6</sup> SAS® EVAAS® FOR K-12, [https://www.sas.com/en\\_si/software/evaas.html](https://www.sas.com/en_si/software/evaas.html) [<https://perma.cc/65TE-VEFG>] (crediting the development of this particular model sold by a private company to Mr. Sanders).

<sup>7</sup> TENN. CODE ANN. §§ 49-1-302(a)(2)(C), 49-5-503(4) (2016); TENN. STATE BD. OF EDUC., TEACHER AND ADMINISTRATOR POLICY § 5.201 (2017) (statutory and regulatory framework delegating authority to state department of education to develop policy for evaluation and further linking that evaluation to tenure determinations).

“Great Recession” eagerly complied.<sup>8</sup> As a consequence, VAMs became codified in state teacher evaluation and employment laws across the country.<sup>9</sup>

Despite their widespread adoption, the use of these statistical models in improving public schools is a source of considerable debate in law and policy. Some scholars applaud their use, arguing that they provide a clear measure of a teacher’s worth and address a persistent policy dilemma: How to improve the quality of our public school teachers.<sup>10</sup> Detractors insist that a teacher’s value is much more than the measure of test scores and, more importantly, that VAMs are statistically flawed.<sup>11</sup> Critics note that VAMs fail to account for the complexity of teaching and cannot accurately control for the impact of other variables (e.g., students’ individual

---

<sup>8</sup> See generally Rhoda Freelon et al., *Overburdened and Underfunded: California Public Schools Amidst the Great Recession*, 2 MULTIDISCIPLINARY J. EDUC. RES., 152 (2012) (documenting the impact of the Great Recession on public schools in California, but also noting the broader impact of the recession on schools and institutions beyond California).

<sup>9</sup> KATHRYN M. DOHERTY & SANDI JACOBS, STATE OF THE STATES 2013: CONNECT THE DOTS: USING EVALUATIONS OF TEACHER EFFECTIVENESS TO INFORM POLICY AND PRACTICE 10 (2013) (noting that in 2013 at least 31 states had adopted the use of standardized test in their teacher evaluation protocols); see also MARK A. PAIGE, BUILDING A BETTER TEACHER: UNDERSTANDING VALUE-ADDED MODELS IN THE LAW OF TEACHER EVALUATION 15, 16 (2016) (describing the links between teacher evaluation systems and teacher employment statutes, such as tenure, and warning against such use for high-stakes decisions).

<sup>10</sup> See, e.g., Eric A. Hanushek, *Conceptual and Empirical Issues in the Estimation of Educational Production Functions*, 14 J. HUM. RESOURCES 351, 353 (arguing for the adoption of production function models to evaluate teachers).

<sup>11</sup> E.g., Linda Darling-Hammond, *Can Value-Added Add Value to Teacher Evaluation?*, 44 EDUC. RESEARCHER 132, 133 (placing the use of value added models in the larger policy debate about how to improve teacher quality).

motivation) that impact student achievement.<sup>12</sup> Because of these issues, commentators cautioned against the use of VAMs in high-stakes employment decisions (e.g. termination), noting such use would invite legal action.<sup>13</sup>

Notwithstanding these warnings, many states embraced VAMs. Florida, for example, amended their teacher evaluation statutes to ensure that VAMs played a controlling role in teacher employment status, including tenure decisions.<sup>14</sup> Teachers and unions almost immediately challenged the use of VAMs through legal means. Lawsuits ranged from violations of the Federal Constitution<sup>15</sup> to assertions that requirements to use VAMs violated the non-delegability doctrine.<sup>16</sup> Many of these received widespread attention in the popular press.<sup>17</sup>

---

<sup>12</sup> *Id.*; see also SEAN P. CORCORAN, 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 22 (2010).

<sup>13</sup> PAIGE, *supra* note 9, at 22 n.28; see also Preston C. Green III et al., *The Legal and Policy Implication of Value-Added Teacher Assessment Policies*, 2012 BYU EDUC. & L.J. 1, 15–16 (2012).

<sup>14</sup> *E.g.*, FLA. STAT. ANN. § 1012.22(1)(c)(5) (West 2013) (connecting teacher salary to an evaluation system that requires use of VAMs).

<sup>15</sup> *E.g.*, *Cook v. Bennett*, 792 F.3d 1294, 1298 (11th Cir. 2015) (alleging use of VAMs violated substantive and procedural due process clauses, as well as the Equal Protection Clause of the 14th Amendment).

<sup>16</sup> *E.g.*, *State ex rel. Stapleton v. Skandera*, 346 P.3d 1191, 1194 (N.M. App. 2015).

<sup>17</sup> *E.g.*, Peter Greene, *Over a Year Ago a Federal Court Struck Down VAM: Why Are We Still Using it to Evaluate Teachers?*, FORBES (June 25, 2018, 08:23 PM), <https://www.forbes.com/sites/petergreene/2018/06/25/over-a-year-ago-a-federal-court-struck-down-vam-why-are-we-still-using-it-to-evaluate-teachers/> [<https://perma.cc/AA4M-NRQ5>]; Patricia MacGregor-Mendoza, *Court Finds Teacher Evaluation System Flawed*, LAS CRUCES SUN NEWS (May 26, 2017, 07:17 PM), <https://www.lcsun->

It has been almost ten years since Race to the Top brought Mr. Sander's idea of VAMs from Tennessee to a national scale, and it seems an appropriate moment to assess their legal and policy ramifications. Indeed, as we note, the use of VAMs has triggered a wave of litigation and policy change that continues today. Many states continue to use VAMs, while others have reduced their use under new federal laws.<sup>18</sup> Thus, assessing the legal and policy landscape forms the basis of this article.

Generally speaking, three lines of legal challenges have emerged. First, some are grounded in the substantive Due Process Clause and Equal Protection Clause of the 14th Amendment, arguing that the laws do not pass rational basis scrutiny.<sup>19</sup> Second, a line of cases challenges the authority or jurisdiction of a particular agency (e.g., state Department of Education) to enact evaluation regulations or laws that use VAMs. Third, some cases advance what we refer to as "process" arguments. These contend that the use of VAMs violates some agreed-upon or standing procedural terms found in the Procedural Due Process Clause or collective bargaining agreements (CBAs). As we note, plaintiffs have captured the most success (although not always) on this third line of argument.

That litigants have experienced more success arguing VAMs offend certain procedural protections comports with common understanding of procedural due

---

news.com/story/opinion/2017/05/26/court-finds-teacher-evaluation-system-flawed/102219102/ [<https://perma.cc/ESS8-SXWX>]; Valerie Strauss, *Judge Calls Evaluation of N.Y. Teacher "Arbitrary" and "Capricious" in Case Against New U.S. Secretary of Education*, WASH. POST (May 10, 2016), <https://www.washingtonpost.com/news/answer-sheet/wp/2016/05/10/judge-calls-evaluation-of-n-y-teacher-arbitrary-and-capricious-in-case-against-new-u-s-secretary-of-education/> [<https://perma.cc/Y645-2T82>].

<sup>18</sup> See *infra* Part III.

<sup>19</sup> See, e.g., *Cook*, 792 F.3d at 1298, 1300.

process. At its core, procedural due process ensures “fundamental fairness” when the government moves to take away a protected interest, such as employment. While courts generally have not overruled a legislature’s policy choice to use VAMs as violative of the substantive due process, they (including a federal appeals court case) have questioned the wisdom of the legislature’s decision.<sup>20</sup> Where they have overturned the use of VAMs, they have done so on procedural grounds.<sup>21</sup> This allows courts to stay within “their lane” and avoid jurisdictional overreach into the policy area.

The article is organized as follows. Part I overviews VAMs, their link to teacher evaluation and employment, and the controversy surrounding their use, especially as a factor in high-stakes employment decisions. Part II provides the most current assessment of cases where the statistical controversy has led to legal action. Part III discusses the recent policy and legal developments with respect the use of VAMs in evaluation that have occurred because changes in federal education law. In conclusion, we note that VAMs have receded, somewhat, in terms of their role in evaluation and employment matters.

## II. VAMs: Promise and Controversy

### A. A Brief History of VAMs in Educational Policy

In the simplest of terms, VAMs (e.g., Tennessee’s TVAAS) are statistical models used to measure the predicted and the actual “value” a teacher “adds” to (or detracts from) student achievement from the point at which students enter a teacher’s classroom to the point students leave. This is typically done using student

---

<sup>20</sup> *See id.* at 1301.

<sup>21</sup> *See id.* at 1301–02.

achievement growth as measured by large-scale standardized test scores (i.e., the tests mandated by the No Child Left Behind (NCLB) Act of 2001). The models attempt to statistically control for outside variables, including students' prior test performance, and student-level background variables (e.g., whether students are eligible for free-and-reduced lunches).<sup>22</sup>

The most widely used VAM is the EVAAS, developed and used in Tennessee.<sup>23</sup> EVAAS comes in different versions for different states (e.g., the EVAAS in Ohio, North Carolina, and South Carolina, the PVAAS in Pennsylvania, the TVAAS in Tennessee, and the TxVAAS in Texas) and different ones based on large and small school districts (e.g., located within Arkansas, Georgia, Indiana, Texas, and Virginia). For each consumer, EVAAS modelers choose one of two sophisticated statistical models.<sup>24</sup>

Using these models, student growth scores are aggregated at the teacher or classroom level to yield teacher-level value-added *estimates*. Depending on where

---

<sup>22</sup> See e.g., Sean Corcoran & Dan Goldhaber, *Value Added and Its Uses: Where You Stand Depends on Where You Sit*, 8 EDUC. FIN. & POL'Y 418, 421 (2013). Other variables include things such as, English language learners (ELLs), gifted, receiving special education services, and classroom and school-level variables (e.g., class sizes, school resources, school leadership).

<sup>23</sup> The EVAAS is advertised as “the most comprehensive reporting package of value-added metrics available in the educational market” in that the EVAAS offers states, districts, and schools “precise, reliable and unbiased results that go far beyond what other simplistic [value-added] models found in the market today can provide.” SAS® EVAAS ® FOR K-12, [https://www.sas.com/en\\_us/software/evaas.html](https://www.sas.com/en_us/software/evaas.html) [<https://perma.cc/76AY-G47W>].

<sup>24</sup> For a comprehensive statistical summation of the various models and options available, see WHITE PAPER: SAS® EVAAS® FOR K12 STATISTICAL MODELS, [https://www.sas.com/content/dam/SAS/en\\_us/doc/whitepaper1/sas-evaas-k12-statistical-models-107411.pdf](https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/sas-evaas-k12-statistical-models-107411.pdf) [<https://perma.cc/F5EW-WCB6>].

teachers' EVAAS estimates fall, as compared to similar teachers to whom they are compared (e.g., within districts) at the same time, teachers' value-added determinations are made.<sup>25</sup> Thereafter, EVAAS modelers make relativistic comparisons and rank teachers hierarchically along a continuum.<sup>26</sup> Teachers whose students grow significantly more than the average and/or surpass projected levels of growth are identified as "adding value"; teachers whose students grow significantly less and/or fall short of projected levels are identified as "detracting value."<sup>27</sup> Teachers whose students grow at rates that are not statistically different from average (i.e., falling within one standard deviation of the mean) are classified as Not Detectibly Different (NDD).<sup>28</sup>

### 1. The Rise of VAMs in National Education Policy: Race to the Top

In 2007, TVAAS/EVAAS entered the national education policy discussion when developer Dr. William L. Sanders shared his research with Congress. Specifically, he testified before the U.S. House of Representatives Committee on Education and the Workforce on how TVAAS could improve teacher

---

<sup>25</sup> For a general overview of the use of VAMs and the concepts noted herein, see WILEY, *supra* note 5.

<sup>26</sup> *Id.*

<sup>27</sup> *Id.*; Audrey Amrein-Beardsley & Clarin Collins, *The SAS Education Value-Added Assessment System (SAS® EVAAS®) in the Houston Independent School District (HISD): Intended and Unintended Consequences*, 20 EDUC. POL'Y ANALYSIS ARCHIVES, no. 12, Apr. 2012, at 1, 7 n.2.

<sup>28</sup> WILEY, *supra* note 5; Amrein-Beardsley & Collins, *supra* note 27, at 7 n.2; see, e.g., WILLIAM L. SANDERS, COMPARISONS AMONG VARIOUS EDUCATIONAL ASSESSMENT VALUE-ADDED MODELS 18 (2006).

accountability and promote educational reform.<sup>29</sup> His testimony spurred the U.S. Department of Education's piloting of VAMs.<sup>30</sup>

The use of VAMs nationally grew under the Race to the Top program. By way of background, RttT was a competitive federal grant program that amounted to an injection of \$4.35 billion to selected states to support educational reform efforts.<sup>31</sup> Receipt of the grant was conditioned on states developing teacher evaluation laws and policy that used VAMs.<sup>32</sup> States that attached relatively more serious consequences (e.g., employment status) to teachers' VAM-based output were viewed more favorably than those that did not.<sup>33</sup> High-stakes consequences included, but were not limited to: teachers' permanent files being flagged, thus preventing teachers from changing jobs within states; the revocation of teacher licenses; teacher tenure; salary increases, decreases, and merit pay; and teacher probation and termination.<sup>34</sup>

Beyond RttT, the federal government used other mechanisms to embed VAMs in state evaluation and employment matters as a matter of law and policy. In 2011, the federal government required that states adopt the accountability practices discussed above

---

<sup>29</sup> CHRISTOPHER B. SWANSON & JANELLE BARLAGE, INFLUENCE: A STUDY OF THE FACTORS SHAPING EDUCATION POLICY 41 (2016), [https://secure.edweek.org/media/influence\\_study.pdf](https://secure.edweek.org/media/influence_study.pdf) [<https://perma.cc/346S-HJSX>].

<sup>30</sup> *Id.*

<sup>31</sup> U.S. DEP'T OF EDUC., RACE TO THE TOP FACT SHEET (2009), <https://www2.ed.gov/programs/racetothetop/factsheet.pdf> [<https://perma.cc/35GG-Y3HM>].

<sup>32</sup> *Id.*

<sup>33</sup> Arne Duncan, Sec'y, Dep't of Educ., Remarks at The Race to the Top Program Announcement: The Race to the Top Begins (July 4, 2009), <https://www.ed.gov/news/speeches/race-top-begins> [<https://perma.cc/3RD5-RP7A>].

<sup>34</sup> *See generally* PAIGE, *supra* note 9 (noting that VAMs became required factors for employment decisions).

(notwithstanding if a state applied or received RttT funds) to secure waivers from the penalties that they would incur for non-compliance with the No Child Left Behind Act of 2001.<sup>35</sup> NCLB, passed with bipartisan support in 2001, required 100 percent of students to attain proficiency in math and reading state standardized tests.<sup>36</sup> The utopian goal has been widely criticized as impractical.<sup>37</sup> Nevertheless, the federal government required states to submit waivers to escape the punitive measures of non-compliance (e.g., intervention of state authorities in the operation of local schools). More specifically, these waivers buttressed the core policy drivers of RttT by continuing to incorporate student test scores as a means to hold teachers accountable for their “value added,” or lack thereof.<sup>38</sup>

The cumulative impact of RttT and federal waivers on the use of VAMs in teacher evaluations was substantial. By 2014, 40 states and Washington, D.C.,

---

<sup>35</sup> KEVIN CLOSE ET AL., STATE-LEVEL ASSESSMENTS AND TEACHER EVALUATION SYSTEMS AFTER THE PASSAGE OF THE EVERY STUDENT SUCCEEDS ACT: SOME STEPS IN THE RIGHT DIRECTION 5 (Nat'l Educ. Policy Ctr. ed., 2018), [https://nepc.colorado.edu/sites/default/files/publications/PB%20Close-Beardsley-Collins\\_1.pdf](https://nepc.colorado.edu/sites/default/files/publications/PB%20Close-Beardsley-Collins_1.pdf) [<https://perma.cc/RG4N-B8N2>].

<sup>36</sup> No Child Left Behind Act of 2001, Pub. L. No. 107-110, § 1001, 115 Stat. 1425 (requiring all students obtain proficiency in specified test areas) (repealed 2015).

<sup>37</sup> See, e.g., Bruce Meredith & Mark A. Paige, Opinion, *Rethinking Federal Role in Education Makes Sense. Trump's Plan Does Not*, ATLANTA J.-CONST.: GET SCHOOLED (Oct. 3, 2018, 11:15 AM) <https://www.myajc.com/blog/get-schooled/opinion-rethinking-fed-education-role-makes-sense-trump-plan-does-not/T19cWIKAznnDpcoxmvr1nJ/> [<https://perma.cc/S3J4-B4FW>] (characterizing the NCLB goal of proficiency as unrealistic, especially in light of the lack of support from the federal government to education and other important public policy areas that impact education success, like housing and health care).

<sup>38</sup> CLOSE ET AL., *supra* note 35, at 8.

(80%) were using or still developing some type of VAM for increased teacher accountability purposes.<sup>39</sup> While state department of education leaders recognized and encouraged the use of VAMs, they did not develop support mechanisms and resources to help teachers understand and subsequently use their VAM-based data to improve their effectiveness.<sup>40</sup> Put differently, information from VAMs was not actionable. This disconnect has been the source of serious contention and concern about the VAM-based teacher and educational reform enterprise.

## B. Statistical and Practical Controversies

Significant statistical and practical concerns surround VAMs, and these are best understood with reference to the professional guidelines that govern education and psychological professions, the *Standards for Educational and Psychological Testing*<sup>41</sup> (hereinafter “Standards”). These issues include, but are not limited to: (1) reliability, (2) validity, (3) bias, (4) transparency, and (5) fairness, with emphasis also on (6) whether VAMs are being used to make consequential decisions using concrete (e.g., not arbitrary) evidence, and (7) unintended consequences. These are discussed below.

### 1. Reliability

Reliability is the degree to which test- or measurement-based scores “are consistent over repeated applications of a measurement procedure (e.g., a VAM) and hence and inferred to be dependable and consistent”

---

<sup>39</sup> *Id.*

<sup>40</sup> *Id.* at 14.

<sup>41</sup> AM. EDUC. RESEARCH ASS'N, AM. PSYCHOLOGICAL ASS'N & NAT'L COUNCIL ON MEASUREMENT IN EDUC., STANDARDS FOR EDUCATIONAL AND PSYCHOLOGICAL TESTING (2014) [hereinafter STANDARDS].

for the individuals (e.g., teachers) to whom the scores pertain.<sup>42</sup> VAMs are reliable when within-group (same school or district) VAM estimates of teacher effectiveness are more or less consistent over time, from one year to the next, regardless of the type of students and subject areas teachers teach. Consistency over time is typically captured using particular statistical tools such as standard errors, reliability coefficients per se, and generalizability coefficients, among others.<sup>43</sup> These situate and make explicit VAM estimates and their (sometimes sizeable) errors and, importantly, help others understand the errors that come along with VAM estimates.

Research has documented serious concerns with respect to VAM reliability (or intemporal stability). Indeed, teachers classified as “effective” one year might have a 25–59% chance of being classified as “ineffective” the next year, or vice versa, with other permutations possible.<sup>44</sup> If a teacher who is classified as a “strong” teacher this year is classified as a “weak” teacher next year, and vice versa, this casts doubt on the reliability of VAMs for the purpose of identifying and making high-stakes decisions regarding teachers. Accordingly, across VAM, reliability is a hindrance, especially when unreliable measures are to be used for consequential purposes like decisions to terminate or deny tenure.

---

<sup>42</sup> *Id.* at 222–23.

<sup>43</sup> *Id.* at 33.

<sup>44</sup> For a comprehensive overview of these concepts, see José Felipe Martínez et al., *Approaches for Combining Multiple Measures of Teacher Performance: Reliability, Validity, and Implications for Evaluation Policy*, 38 EDUC. EVALUATION & POL’Y ANALYSIS 738-56 (2016); see also Peter Z. Schochet & Hanley S. Chiang, *What are Error Rates for Classifying Teacher and School Performance Using Value-Added Models?*, 38 J. EDUC. & BEHAV. STAT. 142-71 (2013).

## 2. Validity

Validity is “the degree to which evidence and theory support the interpretations of test scores for [the] proposed uses of tests.”<sup>45</sup> It is measured by “the degree to which all the accumulated evidence supports the intended interpretation of [the test-based] scores for [their] proposed use[s].”<sup>46</sup> Put another way, validity asks: Does the model assess what it is supposed to assess?<sup>47</sup> Accordingly, one must be able to support validity arguments with quantitative or qualitative evidence that the data derived allows for accurate inferences.

There are various means to assess validity, but of particular focus for researchers is validity as it concerns “concurrent-related evidences.”<sup>48</sup> This helps to assess, for example, whether teachers who post large and small

---

<sup>45</sup> STANDARDS, *supra* note 41, at 11.

<sup>46</sup> *Id.* at 14.

<sup>47</sup> There are sub areas of validity that have been the subject of considerable research as it relates to VAMs.

These are: (1) content-related evidence of validity; (2) concurrent-related evidence of validity; (3) predictive-related evidence of validity; and (4) consequence-related evidence of validity. See Michael T. Kane, *Validating the Interpretations and Uses of Test Scores*, 50 J. EDUC. MEASUREMENT 1, 2, 8 (2013); see generally Samuel Messick, *Validity*, 3 J. EDUC. MEASUREMENT 1, 8–103 (1989). However, while all these evidences of validity help to support construct-related evidence of validity, in VAM research most researchers rely on gathering concurrent-related evidence of validity.

<sup>48</sup> *E.g.*, Edward Sloat, Audrey Amrein-Beardsley & Jessica Holloway, *Different Teacher-Level Effectiveness Estimates, Different Results: Inter-Model Concordance Across Six Generalized Value-Added Models (VAMs)*, 30 EDUC. ASSESSMENT EVALUATION & ACCOUNTABILITY 367, 372 (2018); see also Pam Grossman et al., *The Test Matters: The Relationship Between Classroom Observation Scores and Teacher Value Added on Multiple Types of Assessment*, 43 EDUC. RESEARCHER 293, 293-303 (2014).

value-added gains or losses over time are the same teachers deemed effective or ineffective, respectively, over the same period using other independent quantitative and qualitative measures of teacher effectiveness. Other measures might include supervisors' observational scores. If all measures line up and theoretically validate one another, then confidence in them as independent measures increases.<sup>49</sup> If all indicators point in different directions, something may be wrong with either or both indicators (the VAM tool or observational scores, or both).<sup>50</sup>

Researchers have questioned whether measures of teacher value-added are substantively related to at least one other criterion of teacher effectiveness (e.g., teacher observational or student survey indicators).<sup>51</sup> Moreover, they question whether the concurrent-related evidence of validity that does exist is strong or substantive enough to warrant valid inference-making.

### 3. Bias

Bias pertains to the validity of the inferences that stakeholders draw from test-based scores.<sup>52</sup> Specific to

---

<sup>49</sup> Kane, *supra* note 47, at 6–8, 37, 40, 64.

<sup>50</sup> *Id.*

<sup>51</sup> *E.g.*, Morgan S. Polikoff & Andrew C. Porter, *Instructional Alignment as a Measure of Teaching Quality*, 36 EDUC. EVALUATION & POL'Y ANALYSIS 399, 399–401 (2014); Tanner LeBaron Wallace, Benjamin Kelcey & Erik Ruzek, *What Can Student Perception Surveys Tell Us About Teaching? Empirically Testing the Underlying Structure of the Tripod Student Perception Survey*, 53 AM. EDUC. RES. J. 1834, 1835, 1837–38 (2016).

<sup>52</sup> The *Standards* define bias as follows: as the “construct underrepresentation of construct-irrelevant components of test scores that differentially affect the performance of different groups of test takers and consequently the . . . validity of interpretations and uses of their test scores.” STANDARDS, *supra* note 41, at 216. Biased estimates, also known as

VAMs, unpredictable characteristics (variables outside of the control of a teacher or school) of students can bias estimates about teachers' contributions. Student characteristics include: students' individual motivation, capability to learn, and levels of academic achievement.<sup>53</sup> Because schools do not randomly assign teachers, these variables are not controlled in a way to mitigate bias.<sup>54</sup> Biased results are quite possible, especially when relatively homogeneous sets of students (e.g., English Language Learners (ELLs), gifted and special education students, or free-or-reduced lunch eligible students) are non-randomly concentrated into schools, purposefully placed into classrooms, or both.

Statistical models—even the most sophisticated—cannot control for such bias.<sup>55</sup> One influential study illustrated VAM-based bias when it found that a

---

systematic error as concerning “[t]he systematic over- or under-prediction of criterion performance” are observed when said criterion performance varies for “people belonging to groups differentiated by characteristics not relevant to the criterion performance” of measurement. STANDARDS, *supra* note 41, at 216, 222.

<sup>53</sup> See generally Noelle A. Paufler & Audrey Amrein-Beardsley, *The Random Assignment of Students into Elementary Classrooms: Implications for Value-Added Analyses and Interpretations*, 51 AM. EDUC. RES. J. 328, 328–62 (2014).

<sup>54</sup> See, e.g., Charles T. Clotfelter, Helen F. Ladd, & Jacob L. Vigdor, *Teacher-Student Matching and the Assessment of Teacher Effectiveness*, J. HUM. RESOURCES 778, 779–82 (2006) (noting the various ways teachers are assigned to schools). Class assignments in schools are historically a function of a host of factors, including: pressure from parents for particular class placement and pressure from teachers for placement of particular students, especially those who may tend to be considered “high-achieving.” *Id.* at 781. Additionally, placement among schools within a district is similarly subject to other variables, such as housing patterns. *Id.*

<sup>55</sup> See, e.g., Paufler & Amrein-Beardsley, *supra* note 53, at 335.

student's 5th grade teacher was a better predictor of a student's 4th grade growth than was the student's 4th grade teacher.<sup>56</sup> The absurdity of that finding raises serious questions about the ability of VAMs to control for bias. Notwithstanding, the primary debate raging across articles concerns whether statistically controlling for potential bias by using complex statistical approaches to account for non-random student assignment makes bias negligible, or rather "strongly ignorable."<sup>57</sup>

#### 4. Transparency

Transparency is defined as the extent to which something is accessible and understandable.<sup>58</sup> In terms of VAMs, this relates to the extent to which VAM-based estimates may not make sense to those receiving the information. In education, teachers and principals may not understand the models being used to evaluate their performance. Because of this, they are generally unlikely to use the VAM-generated information for formative purposes (i.e., as a tool to gather information and change practice as soon as possible).<sup>59</sup> Practitioners often

---

<sup>56</sup> Jesse Rothstein, *Student Sorting and Bias in Value-added Estimation: Selection and Observables and Unobservables*, 4 EDUC. FIN. & POL'Y 537, 546–47 (2009); Jesse Rothstein, *Teacher Quality in Educational Production*, Q.J. ECON. 175, 210 (2010).

<sup>57</sup> Sean Reardon & Stephen Raudenbush, *Assumptions of Value-Added Models for Estimating School Effects*, 4 EDUC. FIN. & POL'Y 492, 496–97 (2009).

<sup>58</sup> STANDARDS, *supra* note 44.

<sup>59</sup> Jonathan M. Eckert & Joan Dabrowski, *Should Value-Added Measures Be Used for Performance Pay?*, KAPPAN, May 2010, at 88, 89–90; Rachel Gabriel & Jessica Nina Lester, *Sentinels Guarding the Grail: Value-Added Measurement and the Quest for Education Reform*, 21 EDUC. POL'Y ANALYSIS ARCHIVES 1, 1–30 (2013); Ellen Goldring et al., *Make Room Value Added: Principals' Human Capital Decisions and the Emergence of*

describe value-add data reports as confusing, not comprehensive in terms of the key concepts and objectives taught, ambiguous regarding teachers' efforts at both the student and composite levels, and often received months after students leave teachers' classrooms.

For example, teachers in Houston, Texas, expressed that they are learning little about what they did effectively or how they might use their value-added data to improve their instruction.<sup>60</sup> Teachers in North Carolina reported that they were "weakly to moderately" familiar with their value-added data.<sup>61</sup> Tennessee teachers maintained that there was very limited support or explanation helping teachers use their value-added data to improve upon their practice.<sup>62</sup>

Quite apart from the statistical concerns noted above, the "black-box" nature of VAMs raises additional questions in the field. Indeed, the purported strength of VAMs is that they will improve instruction by providing a wealth of positive diagnostic information. The models are supposed to give practitioners useful, actionable information. Yet, if practitioners have problems understanding the models, the value (if you will) of VAMs is greatly diminished. Unfortunately, statisticians that have developed the models make "no apologies for the

---

*Teacher Observation Data*, 44 EDUC. RESEARCHER 96, 96–97 (2015).

<sup>60</sup> Clarin Collins, *Houston, We Have a Problem: Teachers Find No Value in the SAS Education Value-Added Assessment System*, 22 EDUC. POL'Y ANALYSIS ARCHIVES 1, 4, 15, 22 (2014).

<sup>61</sup> Kim Kappler Hewitt, *Educator Evaluation Policy That Incorporates EVAAS Value-Added Measures: Undermined Intentions and Exacerbated Inequities*, 23 EDUC. POL'Y ANALYSIS ARCHIVES 1, 11 (2015).

<sup>62</sup> See Eckert & Dabrowski, *supra* note 59, at 90.

fact that [their] methods [are] too complex for most of the teachers whose jobs depended on them to understand.”<sup>63</sup>

### 5. Fairness

General questions of fairness have been raised concerning the use of VAMs, especially in the context of high-stakes employment decisions. Fairness is the impartiality of “test score interpretations for intended use(s) for individuals from *all* relevant subgroups.”<sup>64</sup> But issues of fairness arise when a test or test use impacts some more than others in unfair or prejudiced, yet often consequential ways.<sup>65</sup>

Fairness issues are amplified as VAMs are applied in the field. Indeed, VAMs are generally only directly applicable to teachers who instruct in areas that are subjected to standardized tests (typically, math and reading).<sup>66</sup> States and districts can only produce VAM-based estimates for approximately 30–40% of all teachers.<sup>67</sup> The other 60–70%, which sometimes includes entire campuses of teachers (e.g., early elementary and high school teachers) or teachers who do not teach the core subject areas assessed using large-scale standardized tests (e.g., mathematics and English/language arts), cannot be evaluated or held accountable using teacher-level value-added data.<sup>68</sup> Importantly, when districts use this information to make

---

<sup>63</sup> Carey, *supra* note 1, at 13; *see also* Gabriel & Lester, *supra* note 59, at 20.

<sup>64</sup> STANDARDS, *supra* note 41, at 219 (emphasis added).

<sup>65</sup> This concern is consistent with the general argument of this paper. To wit, courts have sustained objections to the use of VAMs where they violate procedural due process, the basic “fundamental fairness.” *See* Cook v. Bennett, 792 F.3d 1294, 1301 (11th Cir. 2015).

<sup>66</sup> *E.g.*, Green et al., *supra* note 13 (noting that the models only apply to 30–40% of teachers).

<sup>67</sup> *Id.*; *see also* Gabriel & Lester, *supra* note 59, at 7.

<sup>68</sup> Green et al., *supra* note 13, at 15, 27–28.

consequential, high-stakes employment decisions the unfairness can have considerable consequences. Some teachers in certain grades or subject areas experience the negative or positive consequences of these VAM-based data more than their colleagues.<sup>69</sup>

## 6. Consequential Use

Assessing the appropriate use of tests must consider the social and ethical concerns<sup>70</sup> in addition to more sterile concerns about statistical methodology.<sup>71</sup> The *Standards* recommend ongoing evaluation of both the intended and unintended consequences of any test as an essential part of any test-based system, including those based upon VAMs.<sup>72</sup>

Typically, ongoing evaluation of social and ethical consequences rests on the shoulders of the governmental bodies that mandate such test-based policies.<sup>73</sup> In this case, local and state education departments would be the agencies in charge of assessing the social costs and ethical issues associated with the use of VAMs in high-stakes contexts. This is because they “provide resources for a continuing program of research and for dissemination of research findings concerning both the

---

<sup>69</sup> This has formed the basis of substantive due process claims against school districts. *E.g.*, *Cook*, 792 F.3d 1294 (agreeing that the system of Florida that adopted VAM ratings that apply to all teachers, including those in non-tested subject areas, was unwise and unfair but upholding it under rational basis test).

<sup>70</sup> *E.g.*, Messick, *supra* note 47, at 8 noting that “[t]he only form of validity evidence [typically] bypassed or neglected in these traditional formulations is that which bears on the social consequences of test interpretation and use.”

<sup>71</sup> *See also* Kane, *supra* note 47.

<sup>72</sup> STANDARDS, *supra* note 41.

<sup>73</sup> *Id.*

positive and the negative effects of the testing program.”<sup>74</sup>

However, this rarely occurs. The burden typically rests on the research community who must provide evidence about the positive and negative effects and explain these effects to external constituencies, including policymakers. This group must collectively determine whether VAM use, given the consequences and issues identified above, warrant the financial, time, and human resource investments.<sup>75</sup> Local and state departments of education typically have not (perhaps for political reasons) acknowledged or sought to examine the consequences of their policy actions.

### 7. Intended Consequences

As noted, the primary intended consequence of VAM use is to improve teaching and help teachers (and schools/districts) become better at educating students by measuring and then holding teachers accountable for their effects on students. The stronger the consequences, the stronger the motivation leading to stronger intended effects. Secondary intended consequences include

---

<sup>74</sup> *Position Statement on High-Stakes Testing in Pre-K–12 Education*, AM. EDUC. RES. ASS’N (2000), <http://www.aera.net/About-AERA/AERA-Rules-Policies/Association-Policies/Position-Statement-on-High-Stakes-Testing> [<https://perma.cc/969R-8RMR>]; see also STANDARDS, *supra* note 41.

<sup>75</sup> Arguably, some “reformers” assume that their ideas are inviolable and opposition is simply a reflection of a recalcitrant system, at best, or teachers’ unions at worst. See e.g., Michelle Rhee, *Opting Out of Standardized Tests? Wrong Answer*, WASH. POST (Apr. 4, 2014) [https://www.washingtonpost.com/opinions/michelle-rhee-opting-out-of-standardized-tests-wrong-answer/2014/04/04/37a6e6a8-b8f9-11e3-96ae-f2c36d2b1245\\_story.html](https://www.washingtonpost.com/opinions/michelle-rhee-opting-out-of-standardized-tests-wrong-answer/2014/04/04/37a6e6a8-b8f9-11e3-96ae-f2c36d2b1245_story.html) [<https://perma.cc/JD5L-6APK>] (suggesting that an organization she founded always keeps students’ interests first and also implying that teachers’ unions do not, especially in regards to the use of standardized tests).

replacing the nation's antiquated teacher evaluation systems which have been criticized by all corners of the education research.<sup>76</sup>

Yet, in practice, research evidence supporting whether VAM use has led to these intended consequences is suspect. Indeed, numerous studies have noted that there is a *lack* of evidence linking VAMs to improved teacher quality. First, VAM estimates have not produced useable information for teachers about how teachers, schools, and states might improve upon their instruction, or how all involved might collectively improve student learning and achievement over time.<sup>77</sup> Likewise, recent evidence suggests the use of VAMs has not led to improvements in teacher evaluation systems.<sup>78</sup> In sum, strong evidence suggest that VAMs have not promoted the intended benefits of providing actionable information for teachers to improve instruction or teacher evaluation systems.

## 8. Unintended Consequences

Simultaneously, ethical and research standards require that the use of testing data must also recognize VAMs' unintended consequences.<sup>79</sup> Policymakers must present evidence on whether VAMs cause unintended effects and if those effects outweigh their intended impact. This means that the educative goals at issue (e.g., increased student learning and achievement) should be

---

<sup>76</sup> See, e.g., DANIEL WEISBERG ET AL., *THE WIDGET EFFECT* (2009) (criticizing the evaluation models that treat teachers as “widgets” and fail to recognize their differences and value).

<sup>77</sup> Henry Braun, *The Value in Value-Added Depends on the Ecology*, 44 EDUC. RES. 2 (2015); Corcoran, *supra* note 12.

<sup>78</sup> Matthew A. Kraft & Allison Gilmour, *Revisiting the Widget Effect: Teacher Evaluation Reforms and the Distribution of Teacher Effectiveness*, 46 EDUC. RES. 234–49 (2017).

<sup>79</sup> See AM. EDUC. RES. ASS'N, *supra* note 74; STANDARDS, *supra* note 41.

examined alongside the positive and negative implications for both the science and ethics of using VAMs in practice.<sup>80</sup>

Researchers have produced an exhaustive list of these unintended consequences.<sup>81</sup> First, the use of VAMs leads to teacher isolation whereby teachers “literally or figuratively ‘close their classroom door’ and revert to working alone.”<sup>82</sup> Sadly, teacher isolation is at cross-purposes with collaboration among colleagues, something that is an essential part to improving schools.<sup>83</sup> Second, the use of high-stakes testing causes teachers to leave the profession and avoid high-needs schools that most need the best teachers.<sup>84</sup> Because of the very nature of VAM-based teacher evaluation which rewards testing achievement, teachers avoid teaching high-needs students. This is rational: if they perceive themselves to be at greater risk of teaching students who may be more likely to hinder their value-added<sup>85</sup> they “seek safer [grade level, subject area, classroom, or school] assignments, where they can avoid the risk of low VAMS scores.”<sup>86</sup> Of course, the flip side of this, teachers avoid challenging assignments or leave the profession all together.<sup>87</sup> Third, and most troubling perhaps, is the dehumanization that high-stakes testing causes. Indeed, under such regimes, teachers view and react to students as “potential score increasers or score compressors,” not children.<sup>88</sup>

---

<sup>80</sup> Messick, *supra* note 47.

<sup>81</sup> See, e.g., Susan Moore Johnson, *Will VAMS Reinforce the Walls of the Egg-Crate School?*, 44 EDUC. RES. 117–26 (2015).

<sup>82</sup> *Id.* at 120.

<sup>83</sup> *Id.*

<sup>84</sup> *Id.*

<sup>85</sup> *Id.*

<sup>86</sup> *Id.*

<sup>87</sup> *Id.*

<sup>88</sup> Hewitt, *supra* note 61, at 32.

### III. The Cases

This section discusses cases where the central issue was the role VAMs played in adverse employment actions. It first traces those cases related to arguments grounded in the substantive Due Process and Equal Protection clauses of the U.S. Constitution. It then highlights the series of cases where plaintiffs challenged the use of VAMs on jurisdictional grounds (i.e., that a particular government agency superseded its authority or other statutes in requiring the use of VAMs). The final subsection assesses the cases where process arguments have been advanced by the plaintiffs.

#### **A. Federal Substantive Due Process Rights & Equal Protection Arguments: VAMs May Be Unwise But Still Constitutional**

##### 1. *Cook v. Bennett*

In 2015, a group of teachers challenged Florida's use of student test scores to evaluate their job performance.<sup>89</sup> As part of that state's application for Race to the Top funds, the state legislature enacted a new teacher performance evaluation regimen in their law of teacher evaluation.<sup>90</sup> Specifically, the legislature required that at least 50% of a teacher's performance evaluation be based on student growth on state standardized tests in math and English (the Florida Comprehensive Assessment Test, or FCAT).<sup>91</sup> The remaining portion of the teacher's evaluation was

---

<sup>89</sup> *Cook v. Bennett*, 792 F.3d 1294 (11th Cir. 2015).

<sup>90</sup> FLA. STAT. ANN. § 1012.34 (West 2011).

<sup>91</sup> *Id.* A teacher's final evaluation was based on the student test growth (the VAM rating) on the FCAT (50%) and a VAM rating based on the school's contribution to a student's growth. *Cook*, 792 F.3d at 1297.

calculated based on a school-wide VAM rating.<sup>92</sup> Not all students take the math and English tests. In fact, students took the English FCAT exam in grades 3 through 10 and the mathematics FCAT exam in grades 3 through 8.

Under the evaluation law, Florida teachers fell under one of three types of categories.<sup>93</sup> “Type A” teachers were those that taught the tested subjects (math and English) in the years that the FCAT was administered for those subjects. In effect, as the Eleventh Circuit Court of Appeals noted, the model adopted by the state education commissioner only worked as designed in evaluating teachers of English in grades 4 through 10 and math in grades 4 through 8.<sup>94</sup> The rest of Florida’s public school teachers fell into two groups. “Type B” teachers taught students in grades 4 through 10, but in subjects other than English or math.<sup>95</sup> “Type C” teachers taught students in grades below 4 or above 10 or their students did not take standardized tests (e.g., art).<sup>96</sup>

The thrust of the legal problem, according to the teachers challenging the evaluation scheme, related to the evaluation of Type B and C teachers. As a practical matter, school districts evaluated Type B teachers using student FCAT scores for math and English, notwithstanding the fact that those teachers did not instruct the students in those subjects.<sup>97</sup> Type C teachers’ VAM scores were calculated based on school-wide FCAT scores derived from student scores in subjects they did not teach.<sup>98</sup> Under this scenario, for example, a second

---

<sup>92</sup> *Id.*

<sup>93</sup> The district court designated the classification set forth in this discussion and, for ease of reference, the appeals court adopted it in its analysis.

<sup>94</sup> *Cook*, 792 F.3d at 1297.

<sup>95</sup> *Id.*

<sup>96</sup> *Id.*

<sup>97</sup> *Id.*

<sup>98</sup> *Id.* at 1298.

grade art teacher's VAM rating could be calculated based on a 3rd grade student's math and English test growth.

The plaintiff-teachers argued that the evaluation laws violated the Substantive Due Process and Equal Protection clauses of the Fourteenth Amendment.<sup>99</sup> Because no fundamental right was at issue, the court applied the rational basis test to determine whether the government's actions had a legitimate purpose and whether the chosen methods were rationally related to that purpose.<sup>100</sup> Ultimately, the court sided with the government, finding that there was a legitimate interest which was to "increas[e] student academic performance by improving the quality of instructional, administrative, and supervisory services in the public schools of the state."<sup>101</sup>

The court also concluded that there was a rational relationship between this purpose and the use of the FCAT VAMs.<sup>102</sup> The court concluded—and the plaintiffs conceded at oral argument—that the government "could have reasonably believed that (1) a teacher can improve student performance through his or her presence in a

---

<sup>99</sup> U.S. CONST. amend. XIV provides, in relevant part, that: "No state shall . . . deprive any person of life, liberty, or property, without due process of law; nor deny to any person within its jurisdiction the equal protection of the laws."

<sup>100</sup> *Cook*, 792 F.3d at 1300 (citing *Fresenius Med. Care Holdings, Inc. v. Tucker*, 704 F.3d 935, 945 (11th Cir. 2013); *FCC v. Beach Comm'ns, Inc.*, 508 U.S. 307, 314 n.6 (1993)).

<sup>101</sup> *Id.* at 1301 (citing FLA. STAT. § 1012.34(1)(a) (2013)); *see also* *Houston Fed'n of Teachers, Local 2415 v. Houston Indep. Sch. Dist.*, 251 F. Supp. 3d 1168, 1182 (S.D. Tex. 2017) (concluding that plaintiff's substantive due process claims failed because "[e]ven accepting plaintiffs' criticisms at face value, the loose constitutional standard of rationality allows governments to use blunt tools which may produce only marginal results."). The *Houston* court, however, ruled that the plaintiff's allegations of procedural due process violations survived summary judgment dismissal. *Id.* at 1183.

<sup>102</sup> *Cook*, 792 F.3d at 1301.

school and (2) the FCAT VAM can measure those school-wide performance improvements, even if the model was not designed to do so.”<sup>103</sup> To be sure, both the appellate and district courts criticized the chosen model.<sup>104</sup>

The court similarly applied the rational basis review to dismiss the equal protection claims.<sup>105</sup> Under this claim, the teachers argued that the evaluation law created a separate class of teachers: “those whose evaluations are based on student growth data for students assigned to the teacher in the subjects taught by the teacher, and those whose evaluations are based on student growth data for students and/or subjects they do not teach.”<sup>106</sup> However, because this classification did not implicate a suspect class (e.g., race, gender) rational basis applied and, under the same line of reasoning of the substantive due process claim, the equal protection claim was dismissed.<sup>107</sup>

---

<sup>103</sup> *Id.*

<sup>104</sup> *Id.* at 1301 (noting that “[w]hile the FCAT VAM may not be the best method—or may even be a poor one—for achieving this goal, it is still rational to think that the challenged evaluation procedures would advance the government’s stated purpose.”). The district court in finding for the government concluded, in dicta, that “[t]he unfairness of the evaluation system as implemented is not lost on this Court” and that “this Court would be hard-pressed to find anyone who would find this evaluation system fair to non-FCAT teachers, let alone be willing to submit to a similar evaluation system.” *Cook v. Stewart*, 28 F. Supp. 3d 1207, 1215–16 (N.D. Fla. 2014), *aff’d sub nom.* *Cook v. Bennett*, 792 F.3d 1294 (11th Cir. 2015).

<sup>105</sup> *Cook*, 792 F.3d at 1301.

<sup>106</sup> *Stewart*, 28 F. Supp. 3d at 1213.

<sup>107</sup> *Cook*, 792 F.3d at 1301 (citing *City of Cleburne v. Cleburne Living Ctr.*, 473 U.S. 432, 440 (1985) (internal citations omitted)).

2. Trout v. Knox County Board of Education

Plaintiff teachers in *Trout v. Knox County Board of Education* brought substantive and procedural due process claims based on their evaluations that used VAMs for purposes of teacher evaluations.<sup>108</sup> In *Trout*, the teachers challenged the use of Tennessee's VAM rating (the EVAAS). Specifically, two teachers (one a math teacher and the other a science teacher) were denied bonuses based on their VAM rating.<sup>109</sup>

Both teachers involved (*Trout* and *Taylor*, respectively) argued that the use of the VAMs was arbitrary and capricious and, therefore, could not be sustained under the rational basis test. Echoing criticisms of the reliability and validity of VAMs,<sup>110</sup> the plaintiffs argued that the VAMs were too imprecise to be used to assess their effectiveness<sup>111</sup> and therefore violated substantive due process rights.

The federal district court ruled in favor of the government. It began its analysis by noting that the plaintiffs failed to state a substantive due process claim.<sup>112</sup> By way of background, a substantive due process claim requires that there be some property interest at stake. Here, under an analysis of property interest rights in the Sixth Circuit Court of Appeals, the court concluded that the plaintiffs did not have an interest in bonuses.<sup>113</sup>

For sake of argument, however, the court went on to apply the rational test and found that the government's use of the VAMs in this case satisfied that

---

<sup>108</sup> *Trout v. Knox Cty. Bd. of Educ.*, 163 F. Supp. 3d 492, 494 (E.D. Tenn. 2016).

<sup>109</sup> *Id.*

<sup>110</sup> *See supra* Part I.

<sup>111</sup> *Trout*, 163 F.Supp. 3d at 500.

<sup>112</sup> *Id.*

<sup>113</sup> *Id.* at 501.

test.<sup>114</sup> The use of VAMs to identify and support instruction to lead to increased student achievement was not in dispute as a legitimate government interest.<sup>115</sup> The plaintiffs, similar to *Cook v. Bennett*,<sup>116</sup> argued that various statistical infirmities made reliance on VAMs irrational, however.<sup>117</sup> In rejecting these arguments, the district court noted, among other things, that there was no legal authority requiring the court to apply a standard with respect to the confidence level of a test.<sup>118</sup>

To be sure, the *Trout* court was sympathetic to the plaintiffs' complaints regarding the statistical inadequacy of the VAMs.<sup>119</sup> Yet, at bottom, there was no legal authority that required the court to apply a certain level of statistical confidence with respect to the government's chosen method for purposes of measuring teacher effectiveness.<sup>120</sup>

---

<sup>114</sup> *Id.*

<sup>115</sup> *Id.* at 503.

<sup>116</sup> *Cook*, 792 F.3d at 1297.

<sup>117</sup> For example, the plaintiffs took issue with the confidence level of the statistical test (68%). *Trout*, 163 F.Supp. 3d at 503.

<sup>118</sup> *Id.*

<sup>119</sup> *Id.* at 504 (writing that the Court notes that Plaintiffs' criticisms of the statistical methods of TVAAS are not unfounded.)

<sup>120</sup> *Id.* at 504–05. The court wrote that while “[p]laintiffs bemoan the statistical imprecision of TVAAS,” no legal authority “support[s] the proposition that the United States Constitution requires legislative decision making regarding the use of statistics to require ‘statistically significant’ results. Absent controlling authority to the contrary, this Court refuses to extend the rational basis test this far—where no suspect class or fundamental right is at issue, the Constitution requires a rational basis, not a statistically significant basis, for the law in question.” *Id.*

### 3. *Wagner v. Haslam*

Another set of teachers in Tennessee challenged the use of VAMs in *Wagner v. Haslam*.<sup>121</sup> Pursuant to state and district evaluation policies, teachers of non-tested subjects were evaluated based on school wide data of student performance on test subjects.<sup>122</sup> Similar to *Cook v. Bennett*, the teachers claimed that this practice violated the substantive Due Process and Equal Protection clauses of the U.S. Constitution.<sup>123</sup>

The federal court, however, echoing the decisions of other federal courts assessing similar claims, rejected the teachers' arguments. With respect to the substantive due process claim, the court enumerated several reasons why the policies at issue passed constitutional muster. It noted that "the State Board could rationally believe that a school-wide score provides some measure (*albeit a crude one*) of evaluating an individual teacher's performance."<sup>124</sup> The court also added that the legislature had continued to amend its teacher evaluation laws to address some of the concerns raised by the plaintiffs.<sup>125</sup>

While the *Wagner* court concluded that the use of VAMs was constitutional, it expressed concerns over fairness similar to those found in *Cook* and *Trout*. Indeed, the *Wagner* court wrote that although the current evaluation processes may produce "unfair results" for certain teachers, it did not rise to the level of being irrational.<sup>126</sup> At the same time, the court was explicit about its use of judicial restraint, especially with respect to education policy questions. Indeed, subject to limited

---

<sup>121</sup> 112 F. Supp. 3d 673 (M.D. Tenn. 2015).

<sup>122</sup> *Id.*

<sup>123</sup> *See Cook*, 792 F.3d at 1297.

<sup>124</sup> *Wagner*, 112 F. Supp. 3d at 694 (emphasis added).

<sup>125</sup> *Id.*

<sup>126</sup> *Id.* at 695.

exceptions,<sup>127</sup> the states have “unfettered”<sup>128</sup> discretion to regulate education, and state legislators can make both “excellent decisions and terrible decisions,” so long as there is some “modicum of rationality.”<sup>129</sup> Put another way, a court may disagree with the policy choice of a governing body, but it is not the role of the courts to second-guess policy judgments of elected officials.<sup>130</sup>

#### 4. *Matter of Lederman v. King*

The one extant case that succeeded in demonstrating the government’s use of VAMs rose to the high bar of arbitrary and capricious is found in *Matter of Lederman v. King*.<sup>131</sup> In this case, a well-regarded veteran teacher who had previously had positive evaluations received an “ineffective” review under New York’s new evaluation system.<sup>132</sup> This new system required the use of VAMs. The teacher, Sheryl Lederman, submitted “overwhelming” and ample evidence from experts in the field that the court concluded satisfied her burden in the record before the court.<sup>133</sup>

In contrast, the court noted that state defendants left numerous statistical issues unaddressed, including the potential VAM biases against teachers with high-

---

<sup>127</sup> Some exceptions, of course, would include the use of race to segregate schools. *See generally* *Brown v. Bd. of Educ.*, 373 U.S. 483 (1954).

<sup>128</sup> *Id.* at 692.

<sup>129</sup> *Id.* at 693.

<sup>130</sup> *But see* PAIGE, *supra* note 9 (arguing that for scholars of educational policy the appropriate question is determining which institutions—courts, legislatures, or markets—have the capacity to best address a particular policy need in education, like teacher evaluation).

<sup>131</sup> *Lederman v. King*, 54 Misc. 3d 886 (N.Y. Sup. Ct. 2016).

<sup>132</sup> *Id.* at 888.

<sup>133</sup> *Id.* at 897–98.

performing students.<sup>134</sup>Critically, how Mrs. Lederman's scores swung so wildly from the second-highest level of effective all the way to the lowest level of ineffective in a single year with statistically similar scoring students, among others.<sup>135</sup> In sum, the court was constrained to the record before it and, on that evidence, found Ms. Lederman satisfied her burden.<sup>136</sup>

## **B. Legislative State Agency Authority Questioned**

Litigants have also challenged the use of VAMs in teacher evaluation on jurisdictional grounds. In these cases, organizations (typically unions) have argued that a legislative or executive agency exceeded their respective authority in requiring VAMs for purposes of evaluation or high-stakes employment decisions. These cases are discussed below.

### 1. *Leff v. Clark County School District*

At issue in *Leff v. Clark County School District* was the constitutionality of changes made to state laws governing teacher evaluation and post-probationary (or continuing contract) status.<sup>137</sup> By way of background, up until 2011, a teacher who completed a probationary period of employment (three years) and was subsequently rehired by a school district received post-probationary status.<sup>138</sup> Post-probationary status conferred to a teacher certain procedural protections should they face termination and required that termination be “for

---

<sup>134</sup> *Id.*

<sup>135</sup> *Id.*

<sup>136</sup> *Id.* at 898.

<sup>137</sup> *Leff v. Clark Cnty. Sch. Dist.*, 210 F. Supp. 3d 1242, 1244–45 (D. Nev. 2016).

<sup>138</sup> *Id.* at 1245.

cause.”<sup>139</sup> In contrast, probationary teachers could be non-renewed without cause and did not have similar procedural protections.

In 2011, the Nevada legislature changed its teacher evaluation and post-probationary statutes. In particular, it required that VAMs be used as part of teacher evaluations. The legislature also required that if a post-probationary teacher achieved two negative evaluations, they would revert back to probationary teacher status.<sup>140</sup> Put another way, a teacher could lose the protections (e.g., a teacher’s termination could only be for “cause”) because of the changes to the state statutes.

Teachers contested the changes based on the federal Constitution’s Contracts Clause.<sup>141</sup> That clause, in relevant part, reads as follows: “No State shall . . . pass any . . . Law impairing the Obligation of Contracts[.]”<sup>142</sup> In essence, the post-probationary teachers claimed that they had a binding contract with the state once they achieved post-probationary status. In exchange for meeting the demands of satisfactory performance, the state had agreed to give them procedural protections and the only grounds for termination were cause. By passing the 2011 amendment that tied teacher contract status to teacher evaluations (that incorporated VAMs), the state breached the contract, something not permitted under the U.S. Constitution.

The federal court declined to adopt the teachers’ position and held that the statute prior to 2011 did not create a contractual obligation between the state and teachers. In its analysis, the court determined that there is a strong presumption in law *against* the idea that a

---

<sup>139</sup> *Id.*

<sup>140</sup> *Id.*

<sup>141</sup> *Id.* at 1244.

<sup>142</sup> U.S. CONST. art. I, § 10.

legislative action creates a private contract.<sup>143</sup> Absent any expression of the legislature that they were creating a contract, it is generally assumed that typical legislative activity simply reflects a policy determination that can be changed.<sup>144</sup> Accordingly, the teachers' claim that the state legislature exceeded its authority with the statutory amendments failed.

## 2. Stapleton v. Skandera

In *Stapleton v. Skandera*, teachers challenged the use of VAMs in teacher evaluation on several jurisdictional grounds related to statutory and agency authority.<sup>145</sup> By way of brief background, the New Mexico legislature attempted—but failed—to make several amendments to its existing teacher evaluation laws in 2012. Notwithstanding this, the New Mexico Department of Education Secretary (through the Department) promulgated new regulations relative to the evaluation of teachers.<sup>146</sup> The teachers sought judicial relief in that the court would suspend the use of the regulations.<sup>147</sup>

The teachers argued that the Secretary exceeded her authority—that, in effect, she acted in a legislative capacity. They raised particular objection to the incorporation of VAMs in teacher evaluation, arguing that such a move could only be done by way of legislative action because it represented a shift in public policy under exclusive legislative purview.<sup>148</sup> However, the New Mexico Court of Appeals sided with the Department on

---

<sup>143</sup> *Leff*, 210 F. Supp. 3d at 1246–47 (citing *Nat'l R.R. Passenger Corp. v. Atchinson Topeka and Santa Fe Ry. Co.*, 470 U.S. 451, 465–66.)

<sup>144</sup> *Id.*

<sup>145</sup> *Stapleton v. Skandera*, 346 P.3d 1191, 1194 (N.M. App. 2015).

<sup>146</sup> *Id.* at 1193 (citing N.M. CODE R. § 6.69.8).

<sup>147</sup> *Id.*

<sup>148</sup> *Id.* at 1194.

this issue. It noted that the enabling statute required only that the Department enact evaluation regulations that were “uniform statewide” and “highly objective.”<sup>149</sup> Accordingly, the legislature left the Secretary “broad authority” to enact regulations reflecting these requirements and, in the view of the court, including VAMs in teacher evaluation protocol did not exceed her authority.<sup>150</sup>

The teachers in *Stapleton* raised other claims related to agency authority. In particular, they raised two additional objections. They contended the new departmental regulations permitted “assistant principals” to observe teachers which violated the state evaluation law that only gave such authority to “principals.”<sup>151</sup> Similarly, they argued that the provisions in the regulations that exempted charter schools from coverage of the evaluations violated the state law requirement that the Department enact a system of “uniform” evaluation.<sup>152</sup>

The court of appeals rejected both of the arguments. With respect to the first claim (that only principals could observe teachers), the court read the state statute as allowing others to observe teachers, including assistant principals. The court wrote, “We agree with the district court that the regulation does not necessarily conflict with the statute because the statute ‘mandates the participation of school principals [but] does not limit the persons who *may* [also] observe [teachers].’”<sup>153</sup> Regarding the claim that the regulations inappropriately exempted charter schools, the state court of appeals noted that the state Charter School Act specifically allowed the Department to waive certain

---

<sup>149</sup> *Id.* at 1195 (citing N.M. STAT. ANN. § 22-10A-19(A) (1978)).

<sup>150</sup> *Id.*

<sup>151</sup> *Id.* at 1196.

<sup>152</sup> *Id.*

<sup>153</sup> *Id.* (alterations in original).

regulations normally applicable to public schools.<sup>154</sup> Because the teachers could not cite to any other legal authority that suggested the waiver was not permitted under the Charter School Act, this theory was also rejected.<sup>155</sup>

### 3. Louisiana Federation of Teachers v. State

In *Louisiana Federation of Teachers v. State*, a teacher's union challenged Louisiana's enactment, amendment, and repeal of multiple state laws related to public education, including those related to teacher evaluation requirements.<sup>156</sup> During the 2012 legislative sessions, the state legislature amended and re-enacted nine different statutes, enacted two new distinct statutes, and repealed twenty-eight statutes all related to education.<sup>157</sup>

The plaintiffs alleged that these actions, which all occurred through one legislative act, violated the state constitution's "single object" requirement.<sup>158</sup> That requirement stipulates that the legislature enacts bills that have "one object" and that various pieces of a bill must have a relationship to one another.<sup>159</sup> The teachers argued that the bill contained unrelated subjects, such as the changes to teacher evaluation, reduction in force issues, rules governing contracts with superintendents, among others.<sup>160</sup>

Louisiana's supreme court rejected the plaintiffs' arguments.<sup>161</sup> The court began its assessment by noting

---

<sup>154</sup> *Id.*

<sup>155</sup> *Id.* at 1196–97.

<sup>156</sup> *La. Fed'n of Teachers v. State*, 171 So. 3d 835, 841 (La. 2014).

<sup>157</sup> *Id.*

<sup>158</sup> *Id.* at 838.

<sup>159</sup> *Id.* at 841.

<sup>160</sup> *Id.* at 842.

<sup>161</sup> *Id.* at 851.

that there is a general presumption that a legislature's acts satisfy the "one object" rule.<sup>162</sup> It also noted that the purpose of the rule was to prevent "logrolling," or the practice of packaging many measures into one bill because any of those measures, alone, would not pass the legislature.<sup>163</sup> The court noted that under such a "grave and palpable" scenario, the legislature would violate the single object rule.<sup>164</sup> Yet, in this case, the court concluded that the object of the act at issue "is improving elementary and secondary education through tenure reform and performance standards based on effectiveness."<sup>165</sup> The court concluded that various components of that bill could be broadly related to this objective.<sup>166</sup>

#### 4. *Robinson v. Stewart*

Another Florida case, *Robinson v. Stewart*,<sup>167</sup> also involved a challenge to the authority of the state Board of Education to implement teacher evaluation regulations using VAMs.<sup>168</sup> In *Robinson*, the plaintiffs sought to declare the 2011 Student Success Act unconstitutional on the grounds that it impermissibly delegated legislative control over public education to the executive branch.<sup>169</sup> The act revised teacher evaluation procedures and required the use of "student learning growth measures" (or VAMs) to evaluate teachers and make significant employment decisions, such as tenure.<sup>170</sup> The act left it to the Department of Education

---

<sup>162</sup> *Id.* at 845.

<sup>163</sup> *Id.* at 845–46.

<sup>164</sup> *Id.* at 851.

<sup>165</sup> *Id.* at 850.

<sup>166</sup> *Id.*

<sup>167</sup> 161 So. 3d 589 (Fla. Dist. Ct. App. 2015).

<sup>168</sup> *Id.* at 590–91.

<sup>169</sup> *Id.*

<sup>170</sup> *Id.* at 591.

Commissioner (the executive branch) to develop the formula to achieve these goals<sup>171</sup> and required the use of standardized test scores.<sup>172</sup>

The Florida District Court of Appeals rejected the plaintiffs' argument that the legislature, in requiring the Commissioner to develop the formula, violated the non-delegability doctrine of the state constitution that ensures a separation of powers.<sup>173</sup> Its analysis noted that the plaintiffs carried a high burden of proof: that the legislature's action violated the doctrine "beyond a reasonable doubt," the highest standard of proof under the law.<sup>174</sup> The court further interpreted the act as simply requiring the Commissioner to provide technical implementation support, as opposed to allowing the executive to make policy determinations.<sup>175</sup>

#### 5. Filed but not Adjudicated

Another case that deserves some attention as it also related to a claim that a state agency exceeded its authority by incorporating VAMs in evaluating teachers. In *Texas Teachers Association v. Texas Education Agency*, the Texas Department of Education adopted teacher

---

<sup>171</sup> *Id.*

<sup>172</sup> *Id.* at 592.

<sup>173</sup> *Id.* at 590–91.

<sup>174</sup> *Id.* at 591.

<sup>175</sup> *Id.* at 592. *But see id.* at 597 (Benton, J., dissenting) (noting that the legislature "has conferred on the State Board of Education power to designate some of them—perhaps nearly all of them—professionally 'unsatisfactory,' and therefore, among other things, subject to being laid off, for reasons that are so unclear and indefinite that the Legislature has abandoned its responsibility to set public policy in this important area, and delegated legislative authority it should have exercised itself to the State Board of Education, an executive branch agency.")

evaluation regulations requiring the use of VAMs.<sup>176</sup> Numerous plaintiffs, including teachers' unions, sought to enjoin the use of VAMs on the grounds that the regulations exceeded the power vested in the state Department of Education.<sup>177</sup> The case settled and the state ultimately agreed to eliminate the required use of VAMs in teacher evaluation regulations.<sup>178</sup>

In *New Mexico ex rel Stewart v. New Mexico Public Education Department*, a group of plaintiffs consisting of legislators, unions, and teachers filed a complaint on the grounds that the state Department of Education improperly infringed other state laws when it promulgated its teacher evaluation regulations.<sup>179</sup> Plaintiffs argued that the School Personnel Act provides for the processes associated with teacher evaluation and termination.<sup>180</sup>

Similarly, plaintiffs allege that the Department's regulation conflicts with New Mexico's Public

---

<sup>176</sup> Sean Collins Walsh, *Union Sues to Block Texas Teacher Evaluation Change*, AUSTIN AM.-STATESMAN, (Aug. 13, 2016), <https://www.statesman.com/news/20160813/union-sues-to-block-texas-teacher-evaluation-change> [https://perma.cc/MQ C2-FATW].

<sup>177</sup> *Id.*

<sup>178</sup> Melissa B. Taboada, *Lawsuit Settled: Texas Teacher Appraisals Won't Be Tied to STAAR Scores*, AUSTIN AM.-STATESMAN (last updated Sept. 25, 2018), <https://www.statesman.com/news/20170504/lawsuit-settled-texas-teacher-appraisals-wont-be-tied-to-staar-scores> [https://perma.cc/XP3C-H2WB].

<sup>179</sup> Complaint, State *ex rel Stewart v. N.M. Pub. Educ. Dep't*, No. D-101-CV-2015-00409 (N.M. 1st Jud. Dist. Feb. 13, 2015), [https://www.aft.org/sites/default/files/nm-complaint-teacherevals\\_1114.pdf](https://www.aft.org/sites/default/files/nm-complaint-teacherevals_1114.pdf) [https://perma.cc/7T99-FG89]. The plaintiffs also claim substantive and procedural due process violations.

<sup>180</sup> See *e.g.*, N.M. STAT. ANN. § 22-10A-19(D) (2010) (providing that evaluations should be determined in part by how well professional development was carried out).

Employment Bargaining Law (the state's enabling collective bargaining statute) that governs "the terms and conditions of employment."<sup>181</sup> More specifically, that law provides that local school districts must negotiate terms and conditions of employment with the representative union.<sup>182</sup> The case is pending with various motions before the court.<sup>183</sup>

### C. Process & "Fundamental Fairness" Cases

#### 1. Houston Federation of Teachers

A group of Houston teachers sought declaratory and injunctive relief in the case of *Houston Federation of Teachers v. Houston Independent School District*.<sup>184</sup> At issue for the court was the constitutional protections afforded teachers in the instance where the Houston public school districts used VAMs to rate and make employment decisions for its teachers.<sup>185</sup> The Houston Independent School District (HISD) had contracted with a third-party vendor who had created certain algorithms to classify and rate teachers based on their students' test performance.<sup>186</sup> This third party vendor, citing trade secrecy, refused to reveal the algorithms when they were requested for review by the teachers.<sup>187</sup> Therefore, teachers who faced adverse employment consequences

---

<sup>181</sup> Complaint at 31, *Stewart*, No. D-101-CV-2015-00409.

<sup>182</sup> See generally N.M. STAT. ANN. § 10-7E-17 (New Mexico's Public Employment Labor Relations Statute).

<sup>183</sup> See *Motion for Summary Judgment Filed in New Mexico Teacher Evaluation Lawsuit*, (Feb. 13, 2018), <http://www.krwg.org/post/motion-summary-judgment-filed-new-mexico-teacher-evaluation-lawsuit> [<https://perma.cc/R8CU-DYHN>].

<sup>184</sup> *Houston Fed'n of Teachers, Local 2415 v. Houston Indep. Sch. Dist.*, 251 F. Supp. 3d 1168, 1174 (S.D. Tex. 2017).

<sup>185</sup> *Id.* at 1171.

<sup>186</sup> *Id.*

<sup>187</sup> *Id.* at 1172.

could not review the underlying formulas that contributed to these decisions.<sup>188</sup>

The teachers claimed that the use of the value added models constituted violation of the substantive and procedural due process clauses of the Constitution.<sup>189</sup> Repeating a line of reasoning in *Cook v. Bennett*, and other cases, the federal district court ruled that the district's use of VAMs did not amount to a substantive due process violation.<sup>190</sup> The court concluded the following: "Even accepting plaintiffs' criticisms at face value, the loose constitutional standard of rationality allows governments to use blunt tools which may produce only marginal results. HISD's motion for summary judgment on this substantive due process claim is granted."<sup>191</sup>

Yet the court found in favor of the plaintiffs' procedural due process claims.<sup>192</sup> The court's analysis is instructive because it relied heavily on procedural due process as ensuring fundamental fairness.<sup>193</sup> The court wrote:

"[The] purpose of procedural due process is to convey to the individual a feeling that the government has dealt with him fairly, *as well as to minimize the risk of mistaken deprivations of protected interests.*" [] In short, due process is designed to foster government decision-making that is both fair and accurate.<sup>194</sup>

---

<sup>188</sup> *Id.* at 1172–73.

<sup>189</sup> *Id.*

<sup>190</sup> *Id.* at 1181–82.

<sup>191</sup> *Id.* at 1182.

<sup>192</sup> *Id.* at 1180.

<sup>193</sup> *Id.*

<sup>194</sup> *Id.* at 1176 (alteration in original) (quoting *Carey v. Piphus*, 435 U.S. 247, 262 (1978)).

The court then listed the factors required for procedural due process to be satisfied in the case of a teacher termination in Texas.<sup>195</sup> Of particular note was that a teacher facing termination must “be advised of the cause for his termination in sufficient detail so as to enable him to show any error that may exist.”<sup>196</sup>

Teachers contended—and the court agreed—that they were not being afforded due process protections because the school district violated the requirement that afforded a teacher “sufficient detail” to show that there may be an error in the government’s decision.<sup>197</sup> Because the district’s third party vendor would not release the underlying formulas, teachers could not possibly assess the accuracy of the district’s value-added rating.<sup>198</sup>

The court listed numerous potential errors that could be revealed if inspection of the formulas was permitted.<sup>199</sup> As the court stated: “The [] score “might be erroneously calculated for any number of reasons, ranging from data-entry mistakes to glitches in the computer code itself. . . . HISD has acknowledged that mistakes can occur in calculating a teacher’s EVAAS score . . . .”<sup>200</sup> The court was troubled by the district’s stipulation that it could not correct a single teacher’s score, even if an error was found, because correcting one score would alter the results of all other teachers.<sup>201</sup>

---

<sup>195</sup> *Id.*

<sup>196</sup> *Id.* The court also noted that a teacher facing termination must be afforded: “the names and testimony of the witnesses against him; [] a meaningful opportunity to be heard in his own defense within a reasonable time; [] and a hearing before a tribunal that possesses some academic expertise and an apparent impartiality toward the charges.” *Id.* (citing *Ferguson v. Thomas*, 430 F.2d 852, 856 (5th Cir. 1970)).

<sup>197</sup> *Id.* at 1176–77 (citing *Levitt v. Univ. of Tex. at El Paso*, 759 F.2d 1224, 1228 (5th Cir. 1985)).

<sup>198</sup> *Id.*

<sup>199</sup> *Id.* at 1177.

<sup>200</sup> *Id.*

<sup>201</sup> *Id.* at 1178.

Indeed, it is worth recalling that value added scores are comparative in nature, assessing one teacher against others.<sup>202</sup> This means that, if one teacher's score is adjusted for an error, it alters all others.<sup>203</sup> The court characterized the underlying foundation of the VAM ratings as built upon a "house of cards."<sup>204</sup> Accordingly, it denied the school district's summary judgment claim with respect to procedural due process.<sup>205</sup>

2. *Washington Teachers' Union v. D.C. Public Schools*

The collective bargaining forum has also been another forum wherein teachers have successfully appealed the use of VAMs in teacher evaluations. By way of background, collective bargaining agreements (CBAs) provide for a process (grievance arbitration), to redress violations of the contract. This arbitration process can be important, especially when a contract calls for certain specifications concerning how teacher evaluations can be conducted. Indeed, districts' decisions to non-renew or terminate a teacher for performance have been called into question because a district fails to follow contractually mandated processes.<sup>206</sup> With some limited

---

<sup>202</sup> *Id.* at 1172.

<sup>203</sup> *Id.* at 1177.

<sup>204</sup> *Id.* at 1178.

<sup>205</sup> *Id.* at 1180. To be sure, procedural due process claims made in *Wagner v. Haslam*, see *supra* notes 121129 and accompanying discussion, did not survive. However, at issue in that case was whether the teachers' bonuses could be linked to their VAM scores. *Wagner v. Haslam*, 112 F. Supp. 3d 673, 688 (M.D. Tenn. 2015). In that context, the court concluded that bonuses were not a property interest sufficient to trigger due process protections. *Id.* at 698.

<sup>206</sup> See, e.g., *Dennis Yarmouth Teachers v. Dennis Yarmouth Reg'l Sch. Dist.*, 360 N.E.3d 883, 884–885 (1977) (reversing a school district's decision to non-renew a probationary teacher

exceptions, scholarship has omitted consideration of the value and importance of collective bargaining agreements in relation to legal challenges to the use of VAMs in teacher evaluations.<sup>207</sup>

Cases emerging from Washington, D.C., illustrate this theme. In Washington, a teacher's union grieved the public district's performance ratings based on VAMs of hundreds of teachers. As an initial matter, the school district challenged whether the issue could, in fact, be subject to the grievance arbitration procedures in the contract. Indeed, as a general matter, disputes are subject to the grievance process only if both parties agreed to arbitrate the dispute under the CBA.<sup>208</sup>

In *Washington Teachers' Union*, a lower court had concluded that the district's final evaluation decisions made under the evaluation systems were not arbitrable but the district's use of evaluation procedures under the collective bargaining was, in fact, arbitrable.<sup>209</sup> Put another way, the parties did not, under the CBA, agree to arbitrate disputes over the judgment of the teachers' final performance, but they did agree to arbitrate whether or not the evaluation procedures outlined were

---

because school district violated terms of the collective bargaining agreement that specified evaluation processes).

<sup>207</sup> *But see* PAIGE, *supra* note 9, at 63–73 (arguing the use of VAMs is susceptible to the grievance arbitration process and the failures of VAMs to accurately assess teacher effectiveness could be remedied through the collective bargaining process.); *see also* Mark A. Paige, *Applying the Paradox Theory: A Law and Policy Analysis of Collective Bargaining Rights and Teacher Evaluation Reform From Selected States*, 2013 BYU EDUC. & L.J. 21, 41–42 (highlighting the benefits of a more collaborative collective bargaining process understood as “interest-based” bargaining particularly with respect to teacher evaluation).

<sup>208</sup> *Wash. Teachers' Union v. D.C. Pub. Schs.*, 77 A.3d 441 (D.C. Cir. 2013)

<sup>209</sup> *Id.* at 444.

followed.<sup>210</sup> On appeal, the District of Columbia Court of Appeals upheld the decision that the district's final judgments were not arbitrable. However, the school district did not challenge the lower court's determination that the issue of whether the district followed evaluation procedures was subject to evaluation.<sup>211</sup>

In at least one other well-publicized case, the Washington Teachers' Union succeeded in frustrating the D.C. Public Schools use of the IMPACT evaluation system.<sup>212</sup> In this case, the union alleged that the school district violated various evaluation procedures when they terminated a seventeen year veteran teacher, Thomas O'Rourke, under the district's evaluation procedures.<sup>213</sup> As noted above, the controlling courts in the District of Columbia have concluded that "process arguments" under the collective bargaining agreement are arbitrable, although the school district's final judgment with respect to evaluation categorization (e.g., ineffective, satisfactory, etc.) is not.

In the *District of Columbia Public Schools* matter, the arbitrator found that the district violated evaluation procedures governing the length of observation visits, which, according to the contract, should be "at least 30 minutes."<sup>214</sup> In this case, the administrators evaluating the teacher exceeded that length by substantial amounts (e.g., observations lasted 80 minutes), which, in the eyes of the arbitrator, amounted to a procedural violation of evaluation processes.<sup>215</sup> Importantly, the arbitrator noted

---

<sup>210</sup> *Id.*

<sup>211</sup> *Id.*

<sup>212</sup> D.C. Pub. Sch. v. Wash. Teachers Union, Local 6, AAA No. 16-20-1300-0499 AVH (Feigenbaum, Arb.); see also Perry Stein, *Teachers Union Touts Victory in Evaluation Fight* WASH. POST (Apr. 5, 2016), <https://www.washingtonpost.com/news/education/wp/2016/04/05/teachers-union-touts-victory-in-evaluation-fight/> [<https://perma.cc/P7RU-PSP7>].

<sup>213</sup> *D.C. Pub. Schs.*, AAA No. 16-20-1300-0499 AVH.

<sup>214</sup> *Id.* at 26–28.

<sup>215</sup> *Id.* at 18.

two other significant factual findings to his decision. He concluded that the administrator evaluating the teacher had a reputation of using the observation system to penalize teachers “he did not like.”<sup>216</sup> A school district administrator, as well, testified that an observation that exceeded or did not meet the thirty minute threshold would amount to a process violation.<sup>217</sup> In sum, and under these circumstances, therefore, procedural violations could be seen as simply pretext for terminating a teacher.<sup>218</sup>

In arbitration cases, the remedy for a bargaining violation can be a contested issue. In Washington, D.C., an arbitrator cannot issue a remedy in the form of recategorizing a teacher’s evaluation from ineffective to effective.<sup>219</sup> Reinstatement and back pay, however, are typical arbitration remedies,<sup>220</sup> and these were, in fact, used in the case.

#### IV. Current Policy Landscape in Wake of ESSA

This section discusses the current policy landscape following the reauthorization of the Elementary and Secondary Education Act of 1965 by Congressional passage of the Every Student Succeeds Act (ESSA) of 2015. It illustrates that the ESSA reauthorization allowed for more state-level flexibility with regards to VAM use. It then highlights how the new policies have essentially shifted the emphasis from VAMs

---

<sup>216</sup> *Id.* at 19.

<sup>217</sup> *Id.* at 7.

<sup>218</sup> *Id.* at 19.

<sup>219</sup> *Wash. Teachers’ Union v. D.C. Pub. Sch.*, 77 A.3d 441, 458 (D.C. Cir. 2013).

<sup>220</sup> *See e.g.*, DISCIPLINE AND DISCHARGE IN ARBITRATION ch. 13.I.A. (Norman Brand & Melissa Birens, eds., 3d ed.) (noting that back-pay and reinstatement are two essential remedies for making an employee whole).

in high stakes decision making to, perhaps, other ways of measurement.

### A. ESSA Reauthorization

In 2015, Congress passed a reauthorization of the Elementary and Secondary Education Act under a new name, the Every Student Succeeds Act.<sup>221</sup> In general, ESSA reduced some federal mandates and incentives tied to accountability system effectively limiting some of the federal control promoted by RttT and other waiver requirements.<sup>222</sup> Specifically, ESSA allowed state departments of education two main changes: (a) ESSA gave state departments leniency to interpret key terms like, “including, as a significant factor, data on student growth for all students,” and (b) ESSA gave state departments more control to determine state goals and measures for success with a federal framework.<sup>223</sup> Put simply, ESSA allowed more flexibility.

To break down the policy changes further, the first main change, allowing states to interpret “data on student growth” differently, allowed state departments of education to step back from the statistically-based measures of student growth such as VAMs. ESSA allowed states to use some measures which could include qualitative measures as data showing student growth, such as student learning objectives (SLOs), which are objectives for the growth of students developed at the beginning of the year by teachers (sometimes in conjunction with others).<sup>224</sup>

SLOs still rely on evidence which can still include VAM scores, but the evidence can also include course

---

<sup>221</sup> Every Students Succeeds Act of 2015, Pub. L. No. 114-95 § 114 Stat. 1177 (2015).

<sup>222</sup> Race to the Top Act of 2011, S. Res. 844, 112th Cong. (2011).

<sup>223</sup> *ESEA Flexibility*, U.S. DEPT EDUC. (2012), <https://www.ed.gov/esea/flexibility> [<https://perma.cc/95A7-FLFA>].

<sup>224</sup> CLOSE ET AL., *supra* note 35, at 18.

exams, performance demonstrations, and other types of evidence. In short, ESSA allowed states to incorporate more nuanced and qualitative measures of student growth without removing the requirement that states must use evidence of student growth. The distinction is small but significant. It signals a redefinition of “data” to include information beyond large standardized testing (although, importantly, it can still include these test scores).

The second main change, allowing states to set their own goals and measures for success, marks a backing away from the strict adequate yearly progress (AYP) goals established by NCLB. Although states still must meet AYP for certain subgroups of students, the consequences and the interventions that must be imposed can be decided by the states themselves. Essentially, ESSA removes the punitive bite demonstrated previously by NCLB, the bite that encouraged many states to apply for waivers and adopt VAMs in the first place, and replaces it with flexibility. States choose their own bite now. The standards remain, but the consequence, the type of intervention required for a failure to meet AYP, is decided by state departments of education.

These two changes, though small, rolled back some of the features that encouraged, or forced, states to use large standardized statewide systems that leaned on VAM results to measure teacher achievement.<sup>225</sup> The new policy meant states did not need to create large-scale comparable data about teacher achievement. States no longer needed to structure their systems top-down and could allow for more bottom-up control, essentially handing more control to local educational authorities such as school districts. ESSA marked a shift of power. The federal government loosened reigns on state

---

<sup>225</sup> Cindy Long, *Six Ways ESSA Will Improve Assessments*, NEATODAY (2016), <http://neatoday.org/2016/03/10/essa-assessments/> [<https://perma.cc/92AW-UC6A>].

departments of education, who, in turn, had the freedom to deviate from establishing one-size-fits-all teacher evaluation systems across their state, handing more of the power to make decisions to local educational authorities, such as districts.

## B. State Plans

Though ESSA allowed for many of the changes stated above, it did not require or guarantee these changes. The work of exercising the flexibility was for the states, not the federal government. Hence, this section on state plans reveals how state teacher evaluation plans changed as a whole after the passage of ESSA through state legislative and regulatory action. The changes, as expected, trend toward less use of VAMs in high-stakes decision making, though the trend is somewhat muted.

In general, less states are currently using growth models or VAMs for teacher evaluation. The percentage dropped from 42% in 2014 to 30% in 2018.<sup>226</sup> However, that percentage drop fails to highlight the magnitude of change. The study showing that the percentage decreased measured whether some states currently use or, importantly, endorse statewide use of VAMs. Some of these states endorse VAMs but allow for local educational authorities to avoid VAMs completely. For example, Maine, encourages the use of VAMs, but offers two models from which local education authorities can choose, one of which measures student growth with SLOs, not VAMs.<sup>227</sup> In this case, VAMs play a role in the state's teacher evaluation process, but, ultimately, the choice is made locally. This represents a major departure from the trend of heavy-handed state teacher evaluation systems before the passage of ESSA.

Additionally, some states have maintained their VAMs but use them in novel ways. North Carolina still

---

<sup>226</sup> CLOSE ET AL., *supra* note 35, at 12.

<sup>227</sup> *Id.* at 13.

uses a VAM, called EVAAS, which featured heavily in many of the lawsuits.<sup>228</sup> However, the state does not use the results to make high-stakes decisions. Rather, North Carolina uses and reports the scores to foster professional development.<sup>229</sup> In other words, the state does not shy from using VAM data as a part of their system, but they do shy from using VAMs for consequential decisions such as tenure decisions and others.

Additionally, and of note, recent state plans demonstrate increased focus on formative feedback practices compared to state plans collected in 2012, with 31 of 51 education plans stating that their evaluation systems use formative data.<sup>230</sup> This shift indicates a significant change in the stated values present in this new set of state documents.

## V. Conclusions

Quite apart from what education scholars and policymakers believe with respect to the merits of added models, all would likely agree that their introduction has had significant consequences. Of course, there is widespread disagreement with respect to *how* these statistical models should be used. Teachers and unions seeking to block the use of VAMs in high-stakes employment decisions have sought judicial relief with mixed success. That said, while courts may uphold the use of VAMs under a rational basis test, they are suspect about the wisdom of using VAMs to make significant decisions with respect to teacher employment status.

But that does not mean that VAMs should be relegated to the dustbin of educational policy history. They may have important contributions to improving teacher quality. They may be important “flags” for

---

<sup>228</sup> See Hewitt, *supra* note 61, at 32.

<sup>229</sup> CLOSE ET AL., *supra* note 35, at 14.

<sup>230</sup> *Id.*

teachers, alerting them to investigate their practice a bit further. VAMs may, someday, play an important role in helping teachers.

Importantly, however, the use of VAMs must be judicious, especially in light of their severe limitations. VAMs cannot tell a teacher *what* causes a particular result (the type of robust and actionable feedback a teacher would want) and they are highly sensitive to demographics and variables outside of a teacher's control. Yet, because VAMs were incorporated in high-stakes decisions with such haste, especially with the impetus of the Race to the Top, they were brought to scale, warts and all.

Thankfully, states have a rare opportunity in educational policy to take a bit more control over their destiny under the Every Student Succeeds Act. They can—and are—placing VAMs as a *piece* of a puzzle to solve teacher quality issues. Many are beginning to adopt laws and policies that minimize or eliminate their use in high-stakes employment. That is a step in the right direction, one that recognizes a relative value to VAMs in the larger quest to improve public education.