

2017

# Report: Tying Teacher Evaluation to Student Test Performance in New York State

Anthony Ciaccio

*Maurice A. Deane School of Law at Hofstra University*

Richard DeMaio

*Maurice A. Deane School of Law at Hofstra University*

Ashley Flynn

*Maurice A. Deane School of Law at Hofstra University*

Sean Hanssler

*Maurice A. Deane School of Law at Hofstra University*

Michelle Malone

*Maurice A. Deane School of Law at Hofstra University*

*See next page for additional authors*

Follow this and additional works at: [http://scholarlycommons.law.hofstra.edu/hofstra\\_law\\_student\\_works](http://scholarlycommons.law.hofstra.edu/hofstra_law_student_works)



Part of the [Education Law Commons](#)

---

## Recommended Citation

Ciaccio, Anthony; DeMaio, Richard; Flynn, Ashley; Hanssler, Sean; Malone, Michelle; Mare, Steven; Santiago, George; and Short, Victoria, "Report: Tying Teacher Evaluation to Student Test Performance in New York State" (2017). *Hofstra Law Student Works*. 11. [http://scholarlycommons.law.hofstra.edu/hofstra\\_law\\_student\\_works/11](http://scholarlycommons.law.hofstra.edu/hofstra_law_student_works/11)

This Report is brought to you for free and open access by Scholarly Commons at Hofstra Law. It has been accepted for inclusion in Hofstra Law Student Works by an authorized administrator of Scholarly Commons at Hofstra Law. For more information, please contact [lawcls@hofstra.edu](mailto:lawcls@hofstra.edu).

---

**Authors**

Anthony Ciaccio, Richard DeMaio, Ashley Flynn, Sean Hanssler, Michelle Malone, Steven Mare, George Santiago, and Victoria Short

## **REPORT:**

### **Tying Teacher Evaluation to Student Test Performance in**

#### **New York State**

#### **Executive Summary**

- Tying teacher evaluation to student test performance evolved as a result of federal legislation
- Arguments in favor of tying teacher evaluation to student performance:
  - Reliable measure of teacher effectiveness
  - Useful when analyzing student test scores over time
  - Objective measure of teacher accountability
  - Proven results in practice
- Arguments against tying teacher evaluations to student performance:
  - Unreliable measure of teacher effectiveness
  - Unintended consequences
    - Financial burden on schools
    - Demoralization of teachers and reduced professional autonomy
    - Narrowly tailored curriculum according to test subjects
    - Increased pressure on students
    - Students opting out of standardized testing
    - Teachers' avoidance of students with special needs
    - Manipulation of and fraudulent tampering with test results
- Alternatives for reform:
  - Eliminate student test performance from teacher evaluations
  - Lower the weight of student test performance on teacher evaluations
  - Use student test scores as a teacher development tool rather than punitively
- Recommendation for reform
  - Change locus of decision-making from State Legislature to State Education Department and Board of Regents

## I. INTRODUCTION

This Report, authored by a small group of third-year law students under the guidance of their Professor,\* focuses on the controversial issue of tying teacher evaluations to student performance on state assessments, specifically, as this practice has been applied under New York State law. First, we provide a brief history of the federal and state laws that have resulted in the implementation of this practice. We then examine the arguments both for and against using student performance on state assessments as a measure of teacher effectiveness, assess all options for amending or abolishing the practice, and propose one procedural recommendation.

## II. HISTORY & BACKGROUND OF TEACHER EVALUATION IN NEW YORK

Originally, each New York public school district was required to have a development committee, consisting of faculty and administrators, which would submit a professional development plan to the State Education Department or have one on file.<sup>1</sup> The plan included a process for evaluating teachers, which generally consisted of administrators conducting teacher observations.<sup>2</sup>

In 2002, President George W. Bush secured passage of the No Child Left Behind Act (“NCLBA”), aimed at improving performance in primary and secondary schools.<sup>3</sup> The NCLBA required states to increase accountability for student performance through state chosen assessments in the fields of mathematics and reading for grades 3-8.<sup>4</sup> Scores were broken down by various student characteristics, including “poverty, race, ethnicity, disability, and limited English proficiency.”<sup>5</sup> States that implemented the NCLBA’s requirements received federal funding.<sup>6</sup> Teachers hired with federal funds linked to the NCLBA were required to meet a “highly qualified” level of effectiveness.<sup>7</sup>

In 2007, New York revamped its teacher evaluation system by implementing Education Law section 3012-b.<sup>8</sup> Section 3012-b required three factors to be considered when evaluating a teacher:<sup>9</sup> (1) the teacher’s use of available student data when providing instruction, (2) peer review, and (3) an assessment of the teacher’s performance by the teacher’s building principal or other building administrator.<sup>10</sup> Section 3012-b was New York’s first step in developing a teacher evaluation system that linked teacher accountability to student performance, as it mandated that

---

\*The following students at the Maurice A. Deane School of Law at Hofstra University (J.D. Candidates, 2017) collectively researched, wrote, and edited this paper as part of a third-year seminar on Law Reform: Anthony Ciaccio, Richard DeMaio, Ashley Flynn, Sean Hanssler, Michelle Malone, Steven Mare, George Santiago, and Victoria Short, under the guidance and instruction of Robin Charlow, Professor of Law.

<sup>1</sup> 8 NYCRR § 100.2(dd) (2016).

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

<sup>3</sup> Sabrina R. Moldt, *Making the Grade: A Ground-Level Analysis of New York State’s Teacher Performance Review Under the APPR*, 2016 B.Y.U. EDUC. & L.J. 217, 220 (2016).

<sup>4</sup> See *NCLB Executive Summary*, U.S. DEP’T OF EDUCATION, <http://www2.ed.gov/nclb/overview/intro/execsumm.html> (last visited Sept. 22, 2016).

<sup>5</sup> *Id.*

<sup>6</sup> 20 U.S.C. § 6311(a)–(b).

<sup>7</sup> 20 U.S.C. § 7801.

<sup>8</sup> N.Y. EDUC. LAW § 3012-b.

<sup>9</sup> *Id.*

<sup>10</sup> *Id.*

teacher evaluations be based on analysis of student data<sup>11</sup> and required a statewide evaluation system that linked teacher accountability to student performance.<sup>12</sup>

Following the expiration of the NCLBA, in 2009, President Obama used the American Recovery and Reinvestment Act to “reshape policy for elementary and secondary schools” through a program known as Race to the Top (“RTTT”).<sup>13</sup> To qualify for a significant new federal grant, RTTT requires a state to submit a federally approved plan that demonstrates its commitment to, among other things, student achievement, teacher success and accountability, and improving “weak” schools.<sup>14</sup> Under the RTTT, students are still subjected to standardized testing from grades 3-8 and student test scores must be a significant factor in evaluating teachers.<sup>15</sup>

In hopes of qualifying for RTTT funds, in 2010, New York State enacted education reform legislation that included the Annual Professional Performance Review (“APPR”) under Education Law section 3012-c.<sup>16</sup> This section, which replaced section 3012-b, required that teachers be evaluated based on three categories, the sum of which would determine an individual’s evaluation: (1) sixty percent based on classroom observations,<sup>17</sup> (2) fifteen to twenty percent determined by some other measure instituted by the district,<sup>18</sup> and (3) the final twenty to twenty-five percent based on state standardized examinations that measured student growth.<sup>19</sup> This final component was determined by student growth and improvement in English and Math for grades 4-8.<sup>20</sup> Based on these scores, each teacher would receive an overall evaluation of highly effective, effective, developing, or ineffective (also known as the “HEDI” system).<sup>21</sup>

In April 2015, the APPR was amended and will eventually be replaced by section 3012-d.<sup>22</sup> The newer system requires that fifty percent of a teacher’s overall score be based on testing and HEDI scores, and the remaining fifty percent on classroom observations.<sup>23</sup> The classroom observation component now requires teachers to undergo two to three classroom observations.<sup>24</sup>

---

<sup>11</sup> *Id.*

<sup>12</sup> Compare N.Y. EDUC. LAW § 3012-b, with N.Y. EDUC. LAW § 3012-c(2)(b)(1).

<sup>13</sup> See U.S. DEP’T OF EDUC., RACE TO THE TOP EXECUTIVE SUMMARY 2 (2009), <http://www2.ed.gov/programs/racetothetop/executive-summary.pdf>.

<sup>14</sup> *Id.* RTTT proposes four reform objectives: “(1) . . . ensure students are prepared to achieve in ‘college and the workplace . . .’; (2) . . . test student growth and success rates . . . ; (3) ‘recruiting, developing, rewarding, and retaining “effective” teachers and principals’; and (4) improving the weakest schools.” *Id.*

<sup>15</sup> *Id.* Principals are also evaluated under the APPR, but for purposes of this Report, the focus is on teachers.

<sup>16</sup> See *id.*; see also N.Y. EDUC. LAW § 3012-c.

<sup>17</sup> N.Y. EDUC. LAW § 3012-c(2)(h). It was mandated that each teacher undergo two observations, and districts were to choose from a list of rubrics to evaluate each teacher.

<sup>18</sup> *Id.* § 3012-c(2)(a)(1). Many districts opted to use SLOs (student learning objectives) in which teachers would distribute some type of pre- and post-test that would earn all points available if a certain percentage of students improved on the post-test.

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

<sup>20</sup> *Id.* § 3012-c(2)(b)(1).

<sup>21</sup> *Id.* § 3012-c(2)(a)(1).

<sup>22</sup> *Id.* § 3012-d.

<sup>23</sup> *Id.* New York’s Governor Andrew Cuomo decided teacher evaluations should be based fifty percent on student standardized test scores. He said the increase was to simplify and standardize the evaluation system, rather than having it be locally negotiated. Diane Ravitch, *Cuomo Wants Test Scores to Count for 50% of Teacher Evaluation*, DIANE RAVITCH’S BLOG (Jan. 21 2005), <https://dianeravitch.net/2015/01/21/cuomo-wants-test-scores-to-count-for-50-of-teacher-evaluation/comment-page-1>.

<sup>24</sup> N.Y. EDUC. LAW § 3012-d(4)(b). One observation must be an announced formal principal evaluation, which comprises eighty percent of the teacher’s observation score, and one is to be done by an unannounced, independent observer, making up the other twenty percent. See *id.*

This last amendment, particularly the increased weight placed on student test scores, was met with widespread pushback. Significant numbers of parents opted their children out of state standardized tests in fear of their children being under too much pressure and stress.<sup>25</sup> In response, the New York State Department of Education (“DOE”) instituted a transition period that will last until 2019, in which schools are still required to abide by section 3012-d while the Legislature decides how better to deal with teacher evaluations.<sup>26</sup> During this current period, state standardized test scores, resulting in HEDI scores, are merely “advisory”, and districts can replace these scores with their own student learning objectives (“SLOs”).<sup>27</sup>

In December 2015, Congress enacted the Every Student Succeeds Act (“ESSA”).<sup>28</sup> ESSA, which takes effect at the start of the 2017-18 school year, gives some teacher evaluation authority back to the states.<sup>29</sup> The controversial student performance tie-in to teacher evaluations required by RTTT is not a requirement under the new ESSA.<sup>30</sup> In essence, ESSA allows states to eliminate the student performance tie-in and still receive federal funding under RTTT<sup>31</sup>—which some states have begun to do.<sup>32</sup>

### III. ARGUMENTS IN FAVOR OF TYING TEACHER EVALUATION TO STUDENT PERFORMANCE

A Value Added (“VA”) approach is a system used to evaluate teachers based on their impact on students’ test scores.<sup>33</sup> Supporters of the VA approach have argued that it has proven to be a reliable measure of teacher effectiveness,<sup>34</sup> can be useful when looking at trends in

---

<sup>25</sup> See Valerie Strauss, *New York’s Costly Experiment in Test-Based Educator Evaluation is Crashing*, WASH. POST (Nov. 17, 2015), <https://www.washingtonpost.com/news/answer-sheet/wp/2015/11/17/new-yorks-costly-experiment-in-test-based-educator-evaluation-is-crashing/>.

<sup>26</sup> N.Y. COMP. CODES R. & REGS. tit. 8, § 30-3.17 (2015)

<sup>27</sup> *Id.*

<sup>28</sup> Every Student Succeeds Act, Pub.L. 114–95 (2015); see also Alyson Klein, *The Every Student Succeeds Act: An ESSA Overview*, EDUCATION WEEK (Mar. 31, 2016), <https://www.edweek.org/ew/issues/every-student-succeeds-act/>. The Act can be found at <http://www2.ed.gov/documents/essa-act-of-1965.pdf>.

<sup>29</sup> See *id.*

<sup>30</sup> Audrey Amrein-Beardsley, *The “Every Student Succeeds Act” (ESSA) To Replace “No Child Left Behind” (NCLB)*, VAMBOOZLED (Dec. 10, 2015), <http://vamboozled.com/the-every-student-succeeds-act-essa-replaces-no-child-left-behind-nclb/>.

<sup>31</sup> See Klein, *supra* note 28.

<sup>32</sup> Audrey Amrein-Beardsley, *Massachusetts Also Moving to Remove Growth Measures from State’s Teacher Evaluation Systems*, VAMBOOZLED (June 2, 2016), <http://vamboozled.com/massachusetts-also-moving-to-remove-growth-measures-from-states-teacher-evaluation-systems/>.

<sup>33</sup> Raj Chetty et al., *Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood*, 104 AM. ECON. REV. 2633, 2633 (2014).

<sup>34</sup> See *id.*; BILL & MELINDA GATES FOUND., LEARNING ABOUT TEACHING: INITIAL FINDINGS FROM THE MEASURES OF EFFECTIVE TEACHING PROJECT (2013), [http://k12education.gatesfoundation.org/wp-content/uploads/2015/12/Preliminary\\_Finding-Policy\\_Brief.pdf](http://k12education.gatesfoundation.org/wp-content/uploads/2015/12/Preliminary_Finding-Policy_Brief.pdf); Thomas Kane, *Should Student Test Scores Be Used to Evaluate Teachers?*, WALL ST. J. (June 24, 2012), <http://www.wsj.com/articles/SB10001424052702304723304577366023832205042>.

students' test scores over time,<sup>35</sup> achieves an objective measure of teacher accountability, and has proven to be effective in at least one major troubled school system.<sup>36</sup>

Some researchers have established that individual teachers have an impact on their students' test scores. For example, Harvard University economist Raj Chetty and colleagues observed teachers' track records for improving student achievement, as measured by gains in test scores, and what effect there was on students when teachers either left or joined a school.<sup>37</sup> Chetty found that "when the teachers with strong track records left, student achievement in that grade level fell," and "when they joined a school, achievement rose."<sup>38</sup> Chetty also noted that student achievement was otherwise unaffected, except in the grade or subject which the departing teacher taught.<sup>39</sup> While these findings are not dispositive, they demonstrate that a gain in student test scores may play an important role in gauging teacher effectiveness.<sup>40</sup> The study authors conclude that student test scores may prove to be useful when combined with classroom observation and student surveys rating their teachers.<sup>41</sup>

Detractors of the VA approach have argued that teachers evaluated using this approach will simply coach students to perform better on tests at the expense of true learning.<sup>42</sup> However, a study conducted by the Gates Foundation yielded results favoring the use of the VA approach in this regard.<sup>43</sup> The Gates study found that "the students with the largest gains on the state tests also tended to have larger gains on other tests which probed students' conceptual understanding."<sup>44</sup> These findings suggest that teachers are not necessarily sacrificing learning for test preparation. In fact, Gates researchers found that teachers who focused heavily on test preparation rarely showed the highest value-added on state tests.<sup>45</sup>

In addition to the positive findings in the Chetty and Gates Foundation studies, supporters of the VA approach have touted the model as being beneficial when looking at trends in students' test scores over time. In a letter to teachers and principals, Joel Klein, the former Chancellor of the New York City public school system, maintained that "where teachers have performed consistently toward the top or the bottom, year after year, these data surely tell us something very important."<sup>46</sup> The underlying premise is that if students are consistently performing poorly on state exams in a certain teacher's class, this pattern may be indicative of teacher ineffectiveness.<sup>47</sup>

---

<sup>35</sup> Anna Phillips, *Klein: ratings are useful for the worst and best teachers*, CHALK BEAT (Oct. 25, 2010), <http://www.chalkbeat.org/posts/ny/2010/10/25/klein-ratings-are-useful-for-the-worst-and-best-teachers/#.WB4fUXeZnp8>.

<sup>36</sup> SUSAN HEADDEN, *INSIDE IMPACT: D.C.'S MODEL TEACHER EVALUATION SYSTEM*, EDUCATOR SECTOR REPORTS 2 (2011) <http://www.nnstoy.org/download/evaluation/Impact%20Report%20Release.pdf>.

<sup>37</sup> Kane, *supra* note 34; Chetty et al., *supra* note 33.

<sup>38</sup> Kane, *supra* note 34.

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

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

<sup>41</sup> *Id.*

<sup>42</sup> *See infra* Part IV.

<sup>43</sup> BILL & MELINDA GATES FOUND., *supra* note 33, at 4.

<sup>44</sup> Kane, *supra* note 34.

<sup>45</sup> BILL & MELINDA GATES FOUND., *supra* note 33, at 4-5.

<sup>46</sup> Phillips, *supra* note 35. Dr. Edward Melnick, Superintendent of The North Shore School District on Long Island, who otherwise opposes the VA approach, echoed Klein's sentiments, and said that students' test scores could be useful to look at longer term patterns in performance. Doctor Edward Melnick, Address at Hofstra University Special Topics Seminar (September 19, 2016).

<sup>47</sup> Nevertheless, even supporters of the VA approach agree that a more robust picture is needed, one that looks at multiple measures of effectiveness. Kane, *supra* note 34.

Supporters have also argued that the VA approach achieves an objective measure of accountability. Prior to the implementation of the VA approach in the APPR, teachers in New York were evaluated based on observations conducted by school principals a couple times a year, often referred to as “drive-bys.”<sup>48</sup> At the end of each year, a principal would rate the teacher as either “satisfactory” or “unsatisfactory,” and there were no structured rules governing how principals assigned these ratings.<sup>49</sup> Subjective systems like “drive-bys” may not provide accurate information on how teachers are performing because they are subject to the potential bias inherent in human judgment. Moreover, Raymond Pecheone, co-director of the School Redesign Network at Stanford University and an expert on teacher evaluation, warns of another danger of subjectivity, that evaluators may not focus on the most meaningful aspects of teaching, including its actual effectiveness. He offers the example of “[a] teacher [who] might get a ‘satisfactory’ check under ‘using visuals’ by hanging up a mobile of the planets in the Earth’s solar system, even though [his/her] students could walk out of their science class with no knowledge of the sun’s role in the solar system or other key concepts.”<sup>50</sup> Further, supporters might argue that the VA approach achieves objective accountability because all students in a certain grade or subject take the same test, so the impact of teachers on student test scores are all estimated uniformly by the same criteria.<sup>51</sup> Thus, using a VA approach, all teachers are treated the same, and human judgment is absent.

Despite criticism of the VA approach, the District of Columbia public school system experienced some success using the VA model. Prompted by the prospect of receiving federal funds granted by RTTT, the D.C. schools implemented the IMPACT system, which employed the VA approach by using student test score improvement to account for 50% of a teacher’s overall effectiveness score.<sup>52</sup> The District then experienced some of the greatest student improvement of all major U.S. cities.<sup>53</sup> The establishment of IMPACT coincided with an increase in high school graduation rates and a significant rise in math and reading test scores.<sup>54</sup>

---

<sup>48</sup> See Kate Taylor, *New York City Teachers Score Highly Under New Evaluation System*, N.Y. TIMES (Dec. 16, 2014), <http://www.nytimes.com/2014/12/17/nyregion/new-york-city-teachers-score-highly-under-new-evaluation-system.html>; Sarah Butrymowicz & Sarah Garland, *How New York City’s value-added model compares to what other districts, states are doing*, THE HECHINGER REPORT (Mar. 1, 2012), <http://hechingerreport.org/how-new-york-citys-value-added-model-compares-to-what-other-districts-states-are-doing/>.

<sup>49</sup> Taylor, *supra* note 48.

<sup>50</sup> THOMAS TOCH & ROBERT ROTHMAN, RUSH TO JUDGMENT: TEACHER EVALUATION IN PUBLIC EDUCATION (2008), [http://educationpolicy.air.org/sites/default/files/publications/RushToJudgment\\_ES\\_Jan08.pdf](http://educationpolicy.air.org/sites/default/files/publications/RushToJudgment_ES_Jan08.pdf).

<sup>51</sup> Derek W. Black, *The Constitutional Challenge to Teacher Tenure*, 104 CAL. L. REV. 75, 131 (2016).

<sup>52</sup> HEADDEN, *supra* note 36, at 2.

<sup>53</sup> *Id.*

<sup>54</sup> Scott Sargrad et al., *Investing in Educator Capacity: An Analysis of State Race to the Top Spending*, CNTR. FOR AMERICAN PROGRESS (Dec. 2015), <https://www.scribd.com/document/293556292/Investing-in-Educator-Capacity-An-Analysis-of-State-Race-to-the-Top-Spending>. The District’s fourth and eighth grade averages in reading scores increased by eleven and five points, respectively, between 2009 and 2015, and students also saw significant improvement in their math scores. See John Davis, *DC Public Schools Continues Momentum as the Fastest Improving Urban School District in the Country*, DISTRICT OF COLUMBIA PUBLIC SCHOOLS, chart 2 in article (Oct. 28, 2015) <http://dcps.dc.gov/release/dc-public-schools-continues-momentum-fastest-improving-urban-school-district-country>. D.C. fourth grade students increased their scores by twelve points from 2009-2015 and eighth graders saw their scores jump by seven points during the same time period. *Id.* In order to adjust to Common Core aligned tests, D.C. suspended the inclusion of student test scores in evaluations, but test scores are expected to be re-introduced in the upcoming school year. Alexandria Neason, *Why Do Schools Keep Changing the Way They Grade Teachers*, SLATE (Feb, 17, 2016, 3:14 P.M.), [http://www.slate.com/blogs/schooled/2016/02/17/dcps\\_changed\\_its\\_teacher\\_evaluation\\_system\\_again\\_why\\_can\\_t\\_schools\\_settle.html](http://www.slate.com/blogs/schooled/2016/02/17/dcps_changed_its_teacher_evaluation_system_again_why_can_t_schools_settle.html).



#### IV. ARGUMENTS AGAINST TYING TEACHER EVALUATIONS TO STUDENT TEST PERFORMANCE

Opposition to the VA model used under the APPR is well documented. Specifically, detractors have argued that the VA approach is an unreliable measure of teacher effectiveness and has significant unintended consequences, including financial effects, that negatively impact teachers and students alike.

VA reliability has been rejected by a significant number of educators and researchers.<sup>55</sup> Unreliability is strongly attributed to the fact that the VA model currently being used is applied on an individual basis and over brief periods, as opposed to in the aggregate over a long period.<sup>56</sup> Scores are subject to significant fluctuation when they are specifically and narrowly applied to each individual student because variables in individual personal characteristics are much harder to account for than those among a broad, general population.<sup>57</sup> When models are applied in the aggregate, such as across an entire school district, generalizations and predictive factors are less vulnerable to instability. Likewise, if a set of scores are apportioned to individual teachers on an annual basis for a long term of years, the aggregate data as applied to that teacher are less vulnerable to outliers. Thus, the greater the data set, the more stable and reliable the models can be.<sup>58</sup>

Studies have offered support for the conclusion that VA models applied on an individual basis are unreliable.<sup>59</sup> One representative study followed the data of a set of teachers from five separate school districts who fell within the lower 20% of the rating scale.<sup>60</sup> The following year, only 20-30% of those teachers had similarly low ratings, and 25-45% of those teachers actually moved to the very top of the ratings scale.<sup>61</sup> Therefore, if an evaluation used only those two years of data, about 70% of the low-scoring teachers would have conflicting score sets from one year to the next, confounding the question of which score is truly representative.<sup>62</sup>

---

<sup>55</sup> NEWTON, X. ET. AL, *Value-Added Modeling of teacher effectiveness: An exploration of stability across models and contexts* in EDUCATIONAL POLICY ANALYSIS ARCHIVES, at 23 (2010), <http://epaa.asu.edu/ojs/article/view/810>.

<sup>56</sup> Stephen J. Caldas, Value-Added: The Emperor with No Clothes, November 2012, Volume 70, Number 3, Teacher Evaluation: What's Fair? What's Effective?, [http://www.ascd.org/publications/educational\\_leadership/nov12/vol70/num03/Value-Added@\\_The\\_Emperor\\_with\\_No\\_Clothes.aspx](http://www.ascd.org/publications/educational_leadership/nov12/vol70/num03/Value-Added@_The_Emperor_with_No_Clothes.aspx).

<sup>57</sup> *Id.*

<sup>58</sup> NEWTON, X. ET. AL, *supra* note 55.

<sup>59</sup> Many studies support findings of similar irregularities as the study referenced here. *See, e.g.*, Sean Corcoran, *Can Teachers be Evaluated by Student Test Scores?*, <http://files.eric.ed.gov/fulltext/ED522164.pdf> (last visited Nov. 27, 2016) (evaluating the VAM in Houston and New York finding a margin of error so large that a teacher at the 43rd percentile (average) might actually be at the 15th percentile (below average) or the 71st percentile (above average), the following year); Caldas, S. J., & Bankston, C. L., *The effect of school population socioeconomic status on individual student academic achievement*, 90 J. OF EDUCATIONAL RES. 269–277 (1997); Caldas, S. J., & Bankston, C. L., *The inequality of separation: Racial composition of schools and academic achievement*, 34 EDUCATIONAL ADMINISTRATION QUARTERLY 533–557 (1998) (finding control models that typically only explained around 20 percent of the total causes of student test scores).

<sup>60</sup> American Education Research Association & National Academy of Education, *Getting Teacher Evaluation Right: A Brief for Policymakers*, <https://edpolicy.stanford.edu/sites/default/files/publications/getting-teacher-evaluation-right-challenge-policy-makers.pdf> (last visited Nov. 27, 2016).

<sup>61</sup> *Id.*

<sup>62</sup> *Id.*

In addition to being an unreliable and unfair measure of teacher effectiveness as it is currently utilized, tying student performance to teacher evaluations also has deeply problematic unintended consequences on the schools, teachers, and students in the system. The primary consequences include: (1) an increased financial burden on schools; (2) demoralization of teachers and reduced professional autonomy; (3) a curriculum narrowly tailored to test subjects; (4) unreasonable pressure imposed on students; (5) parents refusing to have their children participate in standardized testing; (6) a two-tier “gaming of the system”, when teachers limit time investment in students who do not show room for large score improvements and school districts manipulate data to ensure all teacher evaluations come out “effective”; and (7) fraudulent tampering with test results.

First, significant financial burdens are imposed on schools to implement and administer programs in order to become compliant under the APPR. A recent report surveying members of the New York State Council of School Superintendents<sup>63</sup> indicated that 81% were worried that they would not have the funds or time to soundly implement the APPR evaluation system. This was attributable in part to inconsistencies in State government actions, such as placing new demands on schools while at the same time cutting state aid, limiting local tax increases, and doing little to relieve costly state mandates. One contributing factor to this concern was the need for a rapid turnaround in test scores so they can be included in end-of-year teacher evaluations, coupled with an increased emphasis on exam security, resulting in significantly increased costs for districts.<sup>64</sup> Over time, these financial burdens easily outweigh the original RTTT grant for which this system was created.

Second, superintendents have reported that the high-stakes consequences associated with linking teacher evaluations to student scores on state assessments have caused a great deal of fear, anxiety, stress, demotivation and demoralization among educators.<sup>65</sup> Teacher advocates argue that the APPR fails to recognize that teachers may be in the best position to understand how their students learn, and to know the most effective teaching methods for enhancing student performance. Teachers feel that under the current APPR system they are “denied the same respect given to other professionals.”<sup>66</sup>

Third, linking teacher evaluations to student performance on tests has resulted in a “teach to the test” approach, which has altered not only which subjects are taught, but how subjects are taught as well. Using standardized testing to evaluate teacher effectiveness results in classrooms focused on test preparation rather than on educating students with the skills necessary to become engaged members of the global economy.<sup>67</sup> When a teacher’s livelihood is directly impacted by his or her students’ performance on state assessments, the focus of daily classroom activities shifts to test preparation, causing the curriculum to narrow in an effort to raise student

---

<sup>63</sup> THE COUNCIL OF SCHOOL SUPERINTENDENTS, AT THE EDGE: A SURVEY OF NEW YORK STATE SCHOOL SUPERINTENDENTS ON FISCAL MATTERS (2011), <http://www.nyscoss.org/pdf/upload/AttheEdgeSurveyReportFINAL.pdf>.

<sup>64</sup> New York State Principals et. al, *An Open Letter of Concern Regarding New York State’s APPR Legislation for the Evaluation of Teachers and Principals*, [www.newyorkprincipals.org](http://www.newyorkprincipals.org).

<sup>65</sup> Cheryl H. Champ, EdD, *Measuring Teacher Effectiveness: The Impact of institutional Culture on initial Implementation of New York’s Annual Professional Performance Review*, 12 AASA J. OF SCHOLARSHIP & PRACT. 34, 38 (2015).

<sup>66</sup> *Id.* at 241.

<sup>67</sup> Moldt, *supra* note 3, at 245 (noting such a result “directly counter[s] the goals articulated under RTTT and the APPR”).

scores.<sup>68</sup>This narrowed curriculum targets the tested subjects of math and reading and reduces the amount of time teachers spend instructing students in non-accountability subjects, including the arts, music, history, and the sciences.<sup>69</sup> This is particularly true in elementary schools, where teachers are responsible for teaching most, if not all, curricular areas. The curriculum within even the tested subjects of math and reading is then narrowed even further to ensure students score high on the tests, as teachers anticipate which topics are most likely to appear on the tests and thus focus their instruction on those topics.

Fourth, due to the high-stakes nature of the state assessments, an immense amount of pressure is placed on students to do well. Teachers have observed a shift in their student's excitement to learn as a result of this "teach to the test" attitude many feel is encouraged by the APPR's "tie-in" of student scores to teacher evaluation.<sup>70</sup> For example, one teacher noted that her second grade students said "they don't want to go to [third] grade because of all the testing."<sup>71</sup>

Fifth, the pressure on students to perform well has resulted in many parents deciding it is no longer in their children's best interests to take these state assessments, and thus "opting-out" by refusing to have their children sit for the standardized tests. Parents argue that, rather than being used to help students learn or teachers teach, the test results are largely punitive, used to punish teachers who are not performing well.<sup>72</sup> The opt-out movement continues to grow each year. In 2015, twenty percent of students in New York refused to take the state assessments.<sup>73</sup> The following year, in 2016, approximately 230,000 third through eighth grade students in New York refused to take the state standardized math and English exams, up one percent from the previous year.<sup>74</sup>

Sixth, an emphasis on growth measured in student test scores creates concern that teachers and schools alike will attempt to "game the system," which can take several forms. First, schools will have an incentive to keep struggling students in lower-level classes that do not require standardized assessments.<sup>75</sup> Schools may also attempt to keep students out of challenging classes, such as Advanced Regents and Advanced Placement courses, for fear that

---

<sup>68</sup> *Id.* at 246.

<sup>69</sup> *Id.*; see Champ, *supra* note 65, at 38; Eva L. Baker et al., *Problems with the Use of Student Test Scores to Evaluate Teachers*, Briefing Paper #278, Economic Policy Institute at 16 (Aug. 29, 2010), [http://epi.3cdn.net/b9667271ee6c154195\\_t9m6iij8k.pdf](http://epi.3cdn.net/b9667271ee6c154195_t9m6iij8k.pdf).

<sup>70</sup> Moldt, *supra* note 3, at 246.

<sup>71</sup> *Id.* at 245 (internal quotations omitted).

<sup>72</sup> Parents have also voiced concern that the standardized tests are forcing teachers to "teach to the test," as discussed previously, thus spending more time on test preparation than actual learning.

<sup>73</sup> Valerie Strauss, *The testing opt-out movement is growing, despite government efforts to kill it*, WASH. POST (Jan. 31, 2016), <https://www.washingtonpost.com/news/answer-sheet/wp/2016/01/31/the-testing-opt-out-movement-is-growing-despite-government-efforts-to-kill-it/>. At least a dozen other states followed New York's lead. In response, the U.S. Education Department sent states letters warning that federal funds could be withheld from states if at least ninety-five percent of their public school students did not take the standardized assessments in Spring 2016. Valerie Strauss, *U.S. Education Department threatens to sanction states over test opt-outs*, WASH. POST (Jan. 28, 2016), [https://www.washingtonpost.com/news/answer-sheet/wp/2016/01/28/u-s-education-department-threatens-to-sanction-states-over-test-opt-outs/?tid=a\\_inl](https://www.washingtonpost.com/news/answer-sheet/wp/2016/01/28/u-s-education-department-threatens-to-sanction-states-over-test-opt-outs/?tid=a_inl).

<sup>74</sup> Bill Mahoney & Keshia Clukey, *Suffolk leads state in opting out of standardized tests*, POLITICO (Aug. 3, 2016), <http://www.politico.com/states/new-york/albany/story/2016/08/suffolk-leads-state-in-opting-out-of-standardized-tests-104447#ixzz4OoukOjAE>

<sup>75</sup> *Id.*

poor performance on those examinations might result in teachers being unfairly penalized.<sup>76</sup> Next, there is concern that teachers will attempt to avoid students with disabilities or health and/or emotional issues that could challenge their learning or growth, ultimately having a negative impact on the teacher's evaluation and effectiveness rating.<sup>77</sup> One teacher in Upstate New York noted that, prior to the APPR, she worked with her school to take on low performing and emotionally disturbed students, but they did not progress at the level of other students in the same grade, thus negatively impacting her evaluations.<sup>78</sup> As teachers actively avoid students who are already struggling due to the potential impact it may have on their effectiveness rating, these struggling students become at risk of falling through the system's gaps.<sup>79</sup> Also, it has been reported that many school districts have now realized how it is possible to manipulate the system so that all teachers manage to come out with evaluations of "effective", thus frustrating the objective of the entire enterprise.<sup>80</sup>

Finally, there is an increasing concern that high-stakes testing will lead educators to cheat by tampering with or illegally boosting student test scores. There have been a number of recent scandals across the country involving educator cheating: in Galena Park, Texas, a school distributed test-specific study guides to students after illegally reviewing that year's state science test; in Norfolk, Virginia, a principal pressured teachers to show students the answers for a state reading assessment after finding a leaked copy of the test; and, in Springfield, Massachusetts, a principal told teachers to point out wrong answers to students as they took the tests.<sup>81</sup>

## V. ANALYSIS AND RECOMMENDATION FOR REFORM

This section discusses several alternatives and one recommendation for changing the APPR as it is currently constructed. The first, and most extreme, alternative would be to eliminate the APPR altogether and return to a mode of teacher evaluation not based on students' test scores. The second alternative would lower the weight that test scores have in teacher evaluations. The third alternative calls for utilizing the APPR as a teacher development tool rather than as a means of punishing under-performing teachers. Whichever of these alternatives is adopted, we recommend changing the locus of decision-making on these matters from the State Legislature to the State Education Department (SED) and the Board of Regents (BOR) on an ongoing basis.

First, while the APPR's student performance test score tie-in is well-intentioned, its consequences are overwhelming. One possibility is to undue it, abolishing the tie-in altogether.<sup>82</sup> This would obviate all the issues generated by the tie-in and it would create less stress on teachers and students.<sup>83</sup> Teachers would no longer be teaching to the tests nor hindering creativity in the classroom. They could focus more on educating a class based on its needs, rather

---

<sup>76</sup> *Id.* at 247; Carol Burris, *What the 'thoughtless' N.Y. government just did to teachers*, WASH. POST (Apr. 3, 2015), <https://www.washingtonpost.com/news/answer-sheet/wp/2015/04/03/what-the-thoughtless-n-y-government-just-did-to-teachers/>.

<sup>77</sup> Moldt, *supra* note 3, at 247.

<sup>78</sup> *Id.* (explaining "[m]y scores may not reflect the growth that teachers who do not teach the lowest children have. Why would I (or anyone) want to have a classroom with low students in it?").

<sup>79</sup> *Id.* at 248.

<sup>80</sup> Melnick, *supra* note 46.

<sup>81</sup> Moldt, *supra* note 3, at 248.

<sup>82</sup> See A09182, Gen. Assemb., Reg. Sess. (N.Y. 2016).

<sup>83</sup> See Moldt, *supra* note 3, at 250.

than teaching with an eye toward being rated anything other than Ineffective.<sup>84</sup> Another advantage of abolishing the tie-in is saving money and resources,<sup>85</sup> as many schools have spent more to implement the principle than they received under RTTT, and the results for many have not been worth the cost.<sup>86</sup> Thus, the APPR's tie-in is more problematic than it was anticipated to be, a change seems advisable,<sup>87</sup> and one potential change is eliminating the tie-in from the APPR entirely.<sup>88</sup> Conversely, however, eliminating the tie-in leaves us with no objective measure to evaluate a teacher's effectiveness in the classroom.<sup>89</sup> Without using student performance on standardized tests, there is no other objective and efficient way to weed out teachers who really are ineffective.<sup>90</sup> Also, all the money and resources already expended on the APPR will have been wasted, and definitely would not be reimbursed, though one could argue this would be the case whether the tie-in is retained or abolished.<sup>91</sup>

Another alternative option is to lower the weight of student test scores in evaluating teachers. As an example, in March 2016, Georgia lowered the weight of student scores in teacher evaluations from 50% to 30% to address many of the concerns outlined above,<sup>92</sup> particularly the loss of discouraged professional teachers.<sup>93</sup> By lowering the tie-in percentage, Georgia State Superintendent Richard Woods stated that the new law "will provide a better tool for teachers and also help recruit and retain the best teachers."<sup>94</sup>

The final alternative for addressing the APPR's tie-in would be to continue to collect student performance data, but use it as part of a teacher training and development system rather than as a high stakes tool leading to teacher termination in the short term. Instead of adding student growth performance into the score for annual teacher evaluations, the testing data would be collected solely to gain a better picture of performance over a longer period of time.<sup>95</sup> Under

---

<sup>84</sup> *Id.* at 240-41.

<sup>85</sup> *See id.* at 258.

<sup>86</sup> *See id.* The money could have, and possibly should have, been spent on in-class resources and experiences, instead. *Id.*

<sup>87</sup> *See id.* at 258-60.

<sup>88</sup> *See* A09182, Gen. Assemb., Reg. Sess. (N.Y. 2016).

<sup>89</sup> Moldt, *supra* note 3, at 239. This article also offers that, although the tie-in can be used as an effective, objective measure, the APPR, as presently constructed, does not emphasize support for professional growth for those teachers who are labeled ineffective. *Id.*

<sup>90</sup> *Id.*

<sup>91</sup> *See id.*

<sup>92</sup> *See* Richard Woods, *Georgia's Teacher Keys Effectiveness System: Implementation Handbook*, Georgia Department of Education, at 3 (July 2016) <http://www.gadoe.org/School-Improvement/Teacher-and-Leader-Effectiveness/Documents/Finalized%20TKES%20Handbook%20with%20district%20feedback%20%202016-2017.pdf>. In an effort to increase teacher support, Georgia modified its Teacher Keys Effectiveness System ("TKES") to reflect a lower value for student test scores in teacher evaluations.

<sup>93</sup> *See* Ty Tagami, *School Testing Overhaul Legislation Goes to the Georgia Governor*, *The Atlanta Journal-Constitution* (March 24, 2016) <http://www.myajc.com/news/news/local-education/school-testing-overhaul-legislation-goes-to-the-ge/nqrzj/>

<sup>94</sup> Maureen Downey, *Two Senate bills side with teachers: Student growth on tests should not count for half their rating*, GET SCHOOLED (Feb. 10, 2016), <http://getschooled.blog.myajc.com/2016/02/10/two-senate-bills-side-with-teachers-student-growth-on-tests-should-not-count-for-half-their-rating/>. According to Woods, the passage of SB 364 was in response to over-testing and classroom preparation geared towards testing. Further, since TKES was enacted, 44 percent of newly hired teachers left the profession within five years, blaming the method for their evaluation and the number of state-mandated tests as the main reason.<sup>94</sup>

<sup>95</sup> Studies have indicated that while VA scores are unreliable on an individually applied short term basis, they become increasingly more reliable as applied broadly on a longer term basis. *See supra* Part IV, notes 56 and 57. *See also* Stephen Caldras, *Value-Added: The Emperor with No Clothes in Teacher Evaluation: What's Fair? What's*

this proposal, schools could use the annual performance scores to identify those teachers who consistently underperform, and create an opportunity to focus on their training and development. This allows schools to have an objective evaluation tool to reference when addressing underperforming teachers, without creating concern among teachers that every set of scores can put their job in jeopardy. This proposal would not result in a complete discard of the current system; however, it would remedy some of the chilling effects the current system has placed on students and teachers.<sup>96</sup> If scores were used as a training and development instrument, teachers could feel less pressured, have more autonomy, and place less emphasis on solely getting students to attain a certain score on an exam.

This proposal is not without its drawbacks, and it does not completely obviate many of the issues arising out of the current APPR system. First, it will not alleviate the financial and other burdens of standardized testing, as that would remain unchanged.<sup>97</sup> Second, it does not provide a response to the need for an objective standard of measurement of teacher effectiveness.<sup>98</sup> Third, the data itself is not useful unless it is used in conjunction with a training and development program structured around the statistical feedback received by the VA model, arguably adding a new and significant burden on schools. However, this proposal does address many of the APPR tie-in issues while continuing to promote the best interests of the schools, teachers, and students.

Regardless of which alternative is adopted, including maintaining the status quo as an option, we recommend shifting the locus of decision making on the matter of teacher evaluation from the Legislature to the SED and the BOR.<sup>99</sup> By making this shift, the issue would be in the hands of an agency that has the most expertise in this area and is thus best situated to effectively formulate a plan for evaluating teachers, one which takes into account the significant body of ongoing research and knowledge developed in the relevant subject matter.<sup>100</sup> Further, because the SED and the BOR have smaller agendas, they have the ability to focus more attention than the Legislature can on the minutia of the teacher evaluation process.<sup>101</sup> Additionally, they can more readily make adaptations to the system, when necessary, as additional pertinent information is revealed over time, because the regulatory process is simpler than the legislative process, and may be less politically sensitive.<sup>102</sup> Also, the regulatory process can include outside input

---

*Effective?* 70 EDUC. LEADERSHIP, Vol. 3 (Nov. 2012), [http://www.ascd.org/publications/educational\\_leadership/nov12/vol70/num03/Value-Added@\\_The\\_Emperor\\_with\\_No\\_Clothes.aspx](http://www.ascd.org/publications/educational_leadership/nov12/vol70/num03/Value-Added@_The_Emperor_with_No_Clothes.aspx) (“In the aggregate, these models can indeed help us better understand how student, classroom, and school characteristics influence education outcomes.”).

<sup>96</sup> See *supra* Part IV.

<sup>97</sup> See THE COUNCIL OF SCHOOL SUPERINTENDENTS, WHAT LEADERS THINK ABOUT COMMON CORE AND STATE ASSESSMENTS 15 (Nov. 2015) (responses noting that testing is too long for students and scores need to be turned over faster).

<sup>98</sup> See *supra* Part III.

<sup>99</sup> See A09182, Gen. Assemb., Reg. Sess. (N.Y. 2016) for a bill, sponsored by Assemblyman Ed Ra et al., which appears to contain this recommendation.

<sup>100</sup> The commissioner of the SED is appointed by the Board of Regents, so there is a direct connection between the two agencies. See, e.g., *MaryEllen Elia Appointed New Commissioner of the State Education Department*, NEW YORK STATE EDUC. DEP’T (May 26, 2015), <http://www.nysed.gov/Press/MaryEllen-Elia-Appointed-New-Commissioner-of-the-State-Education-Department>.

<sup>101</sup> See generally Juliet Squire, *The State Education Agency: At the Helm, Not the Oar*, THOMAS B. FORDHAM INSTITUTE (Apr. 21, 2014), <https://edexcellence.net/publications/the-state-education-agency-at-the-helm-not-the-oar>

<sup>102</sup> *Rule Making Process*, DEP’T OF STATE, DIVISION OF ADMIN. RULES, <http://www.dos.ny.gov/info/ruledialog.html> (last visited Nov. 23, 2016).

through the notice and comment rulemaking procedure, thereby creating greater sensitivity to public concerns and, hopefully, an increased public “buy-in” to any regulatory plan.<sup>103</sup>

Conversely, moving the focus to the SED could potentially have setbacks. Some question whether the agency is adequately staffed and funded.<sup>104</sup> Furthermore, as an unelected body, the SED is less politically accountable than the Legislature, which means agency officials might be beholden to those who appoint them and not particularly responsive to public concerns.<sup>105</sup> On the other hand, some degree of political independence might serve as a strength, because it could neutralize the influence of the many constituencies that have a vested interest in the matter of teacher evaluations.

While we take no position on the many substantive issues previously identified and analyzed, we believe that, as a procedural matter, it makes more sense to place primary decision making authority in the administrative body going forward, at least in theory. If the agency proves to be insufficiently effective as it takes on this role, the Legislature, as always, retains the power to control administrative discretion should it see the need to do so in the future.

---

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

<sup>104</sup> For example, Senator John Flanagan, former education committee chair, recently stated that “the [New York State Education Department] is underfunded and needs more state aid.” Jessica Bakeman, *As Race to the Top Ends, Controversy Continues*, POLITICO (July 16, 2015), <http://www.politico.com/states/new-york/albany/story/2015/07/as-race-to-the-top-ends-controversy-continues-023795>. Ira Schwartz, assistant education commissioner, similarly relayed frustrations about the SED’s underfunding, stating in response to Governor Cuomo’s criticism, “You’re simultaneously criticizing us for the job that we’re doing and yet we specifically identified the need for resources that we didn’t receive.” See Jessica Bakeman, *In Turnabout, Lawmakers Back Power Increase for State Ed Officials*, POLITICO (Aug. 11, 2015), <http://www.politico.com/states/new-york/albany/story/2015/08/in-turnabout-lawmakers-back-power-increase-for-state-ed-officials-024423>. (“Cosimo Tangorra, a suburban Albany superintendent who previously served as deputy education commissioner, said state leaders need to increase funding or else limit the department’s responsibilities, because it can’t keep doing more with less.”)

<sup>105</sup> It should be noted that, relevant to the concerns of lack of SED responsiveness, the authors were unable to get anyone at the SED to respond to our requests for information.