August 17, 2012

Of Testing and Tablespoons: Evaluating the Use of Student Test Scores for Teacher Assessment

Daniel Straw

Available at: https://works.bepress.com/daniel_straw/1/
Of Testing and Tablespoons: Evaluating the Use of Student Standardized Test Scores for Teacher Assessment

Employing standardized achievement tests to ascertain educational quality is like measuring temperature with a tablespoon.

W. James Popham

I. Introduction

The news was disturbing, to say the least. A Los Angeles Unified School District (LAUSD) fifth grade teacher, Rigoberto Ruelas, committed suicide after a Los Angeles Times (“L.A. Times”) teacher-rating database labeled him as “less effective” than his peers. Ruelas was a teacher in one of Los Angeles’s most difficult neighborhoods for fourteen years. While Ruelas did not leave a note, family members indicated that Ruelas was depressed following the publication of the database in August 2010. The L.A. Times analyzed the test scores of LAUSD students and then published the results to “estimate the effectiveness of L.A. teachers.” While there are numerous caveats to the use of student test scores to evaluate teachers, these warning are routinely unheeded, often by those issuing

---

1 W. James Popham, Why Standardized Tests Don’t Measure Education Quality, Educ. Leadership, March 1999, at 8, 10. Dr. Popham, an Emeritus Professor at UCLA with over 30 years of teaching graduate students about assessment, also put it thusly:

I’m sure you recognize that if you tried to measure a child’s temperature with a tablespoon instead of a thermometer most folks would think you’re rather confused. Unfortunately, in education there is a lot of confused behavior . . . because people are judging the effectiveness of our schools using . . . tools that are absolutely wrong for such a purpose.


3 Id.

them. The L.A. Times did this when it published its controversial teacher database. While Ruelas’s suicide is an extreme reaction to the data, serious concerns exist surrounding the use of student test scores to evaluate teachers. To be clear, the issue is not that teachers do not want evaluations. Rather, teachers desire evaluation systems that are fair and allow them to become better professionals.

Several current federal programs such as Race to the Top (RTTT) and the Teacher Incentive Fund (TIF) require that states competing for funds tie student test scores to teacher evaluation. The Obama Administration’s Elementary and Secondary Education Act (ESEA) Waiver plan, which allows states to apply for a waiver from some of the requirements of No Child Left Behind (NCLB), also requires the tying of student test scores to teacher evaluation.

On October 20, 2011, the United States Senate Committee on Health, Education, Labor & Pensions (“HELP Committee”) passed out of committee the Elementary and Secondary Education Reauthorization Act (“ESEA Reauthorization”). While the bill sent to the Senate does not require states to tie teacher assessment to student achievement, it would codify RTTT and the TIF.

---

6 The L.A. Times article states, “No one suggests using [test scores] as the sole measure of a teacher.” However, the article proceeds to evaluate teachers solely on student test scores, classifying the teacher’s on a five-point scale. See Felch, supra note 4.
9 Id.
Further, both the Senate and the House of Representatives are likely to debate the issue of teacher assessment as reauthorization of the ESEA proceeds. In any event, teacher evaluation systems based on student test scores are in place in some states, with others looking to implement similar programs in the future.\textsuperscript{15}

As the debate over the reauthorization of NCLB gathers steam, now is the time to evaluate what role, if any, student standardized test scores should play in teacher evaluation. While teachers are the most important in-school factor contributing to student learning, teachers only account for roughly 10 to 15 percent of variability in student achievement.\textsuperscript{16} Approximately 60 percent of variability is attributable to out-of-school factors such as student and family background characteristics.\textsuperscript{17} Much of the focus on teacher accountability in school reform is misplaced and potentially damaging. Emphasis on teacher evaluations invests time and energy on an area that will not greatly increase student performance and detracts from discussion of more important areas such as inequitable resource allocation and poverty.\textsuperscript{18}

This note argues that the use of student test scores as a significant part of teacher evaluations has no rational basis in law, and therefore the government should instead focus on performance-based assessments and take steps to elevate the status of teaching as a profession. Just as a tablespoon is not the right

\textsuperscript{14} Id. At §§2301-05.
\textsuperscript{16} Matthew Di Carlo, Teachers Matter, But So Do Words, Shanker Blog (July 14, 2010), http://shankerblog.org/?p=74.
\textsuperscript{17} Id.
\textsuperscript{18} Id.
instrument to take a child’s temperature, a standardized test is not the right instrument to measure the effectiveness of a teacher. Research does not support the use of student test scores to evaluate teachers. Further, there is a serious risk that states will invest needed funds in untried programs and that states will become mired in litigation as teachers seek to prevent the use of these evaluation systems. Part II summarizes the role of teacher assessment in federal programs. Part III critiques the use of student test scores to assess teachers in order to obtain federal funding and argues such programs violate teachers’ substantive due process rights. Part IV suggests some alternative approaches to teacher assessment.

II. Teacher Evaluation in Federal Policy and Statutes

A. Elementary and Secondary Education Act of 1965

The ESEA is the United States’ flagship education law. At the time of its passage, the ESEA primarily focused on providing funding to schools serving low-income families. While the ESEA provided funding for qualified schools, it expressly prohibited the Federal government from “exercis[ing] any direction, supervision or control over the curriculum, program of instruction, administration, or personnel of any educational institution or school system.” Subsequent reauthorizations of the ESEA maintained this hands-off model, though some states implemented “minimum competency” standards for students in the 1970’s.

---

20 Elementary and Secondary Education Act of 1965, Pub. L. No. 89-10, 79 Stat. 27 (1965). This funding is commonly referred to as Title I funding. Superfine, supra note 19.
21 Elementary and Secondary Education Act of 1965 § 604.
Congress established the Department of Education in 1979.\textsuperscript{23} The Department of Education Organization Act contains nearly identical language prohibiting the involvement of the federal government in school control.\textsuperscript{24}

Federal hesitancy to get involved in state and local control over education extended to the judiciary, as demonstrated by \textit{San Antonio Independent School District v. Rodriguez}\textsuperscript{25} in 1973. In \textit{Rodriguez}, the U.S. Supreme Court determined that education is not a fundamental right under the United States Constitution. The Court also expressed its hesitancy to get involved in “persistent and difficult questions of educational policy.”\textsuperscript{26} The Court stated it did not have the requisite “specialized knowledge and experience” to interfere with educational policy decisions made at the local level.\textsuperscript{27}

An increased federal role in education can be tied to the publication of \textit{A Nation at Risk: the Imperative for Education Reform} (hereinafter “\textit{A Nation at Risk}”) by the Department of Education.\textsuperscript{28} \textit{A Nation at Risk} argued that the United States was falling behind other nations because of the failure of American schools.\textsuperscript{29} Reform was necessary if the United States was to maintain its global dominance, particularly in the realms of science and technology.\textsuperscript{30} As to teachers, \textit{A Nation at

\textsuperscript{24} 20 U.S.C. § 3403(b). It is interesting to note that the prohibition found in the ESEA is at the conclusion of the statute, while it is prominently placed very near the beginning of the Department of Education Authorization Act.
\textsuperscript{25} 411 U.S. 1 (1973).
\textsuperscript{26} Id. at 42
\textsuperscript{27} Id.
\textsuperscript{29} Id. at 9.
\textsuperscript{30} Id. at 10.
Risk found teachers were severely underpaid\(^{31}\) and that there was a severe shortage of teachers qualified to teach math and science.\(^{32}\) While the federal government created no new federal programs because of A Nation at Risk, the report served to preserve the Department of Education, which President Reagan vowed to abolish two years earlier.\(^{33}\) In the wake of A Nation at Risk, states began overhauling their education programs by adding science classes, extending the school year, increasing teacher salaries and implementing state educational standards.\(^{34}\)

The federal government encouraged the movement towards the creation of state standards when Congress passed the grant program Goals 2000 in 1994.\(^{35}\) Also passed in 1994 was the Improving America’s Schools Act (IASA), which incorporated the Goals 2000 framework.\(^{36}\) This reauthorization of the ESEA conditioned the receipt of Title I funds on the development and implementation of standards and assessments.\(^{37}\) However, the statute had very little in the way of measures to address poorly performing schools.\(^{38}\) To the extent that these statutes discussed teachers, they highlighted the need for teachers to be involved in developing educational policies, earmarked funds for professional development, and encouraged states to invest in teacher preparation programs.\(^{39}\)

---

\(^{31}\) The study found the average pay after twelve years was $17,000 and that many teachers had to supplement their income with part time jobs and summer employment. Id. at 20.

\(^{32}\) Id.

\(^{33}\) Kress, supra note 22, at 190.

\(^{34}\) Id.

\(^{35}\) Superfine, supra note 19, at 89.

\(^{36}\) Id.

\(^{37}\) Id.

\(^{38}\) Kress, supra note 22, at 192.

B. No Child Left Behind

NCLB is the current incarnation of the ESEA.\(^{40}\) Signed into law by President Bush in 2002, NCLB represented an intensification of the federal role in education.\(^{41}\)

NCLB conditions receipt of Title I funds on the satisfaction of several federal requirements.\(^{42}\) States must implement standards for math, science, and reading.\(^{43}\) While the states determine which standards to adopt, the standards must be “challenging academic standards.”\(^{44}\) In addition, the states must test students on these standards using a “high quality” assessment.\(^{45}\) The states must then determine how much progress, as measured by the assessments, is required annually.\(^{46}\) This measure is known as Adequate Yearly Progress (AYP).\(^{47}\) Schools failing to meet AYP face increasing penalties.\(^{48}\) If a school fails to meet AYP for five years, it is then subject to one of five “alternative governance arrangements,” including “replacing all or most of the school staff who are relevant to the failure to make adequate yearly progress.”\(^{49}\) The ultimate goal of NCLB is to have all students meet or exceed the states’ proficiency requirements by 2014.

---

\(^{41}\) Superfine, supra note 19, at 89.
\(^{42}\) Id.
\(^{44}\) Id.
\(^{46}\) 20 U.S.C. §6311(b)(2)(B) (2006). Students must be tested annually in grades 3-8, and at least once in high school. Id.
\(^{47}\) Id.
\(^{48}\) AYP is to include objectives for the achievement of all public school elementary and secondary students, as well as separate objectives for economically disadvantaged students, minority students, students with disabilities, and students with limited English proficiency. 20 U.S.C. § 6311 (b)(2)(C) (2006).
\(^{49}\) The other alternatives are to reopen the school as a charter school, contract with a private entity to run the school, turn the operation of the school over to the state, or any other major restructuring of the school’s governance. Id.
Some have criticized the AYP requirement as some states set low requirements for proficiency. For example, 26 percent of fourth grade students in both South Carolina and Georgia achieved proficiency in reading as measured by the National Assessment of Educational Progress (NAEP). However, per their respective state requirements, 36 percent of South Carolina’s fourth graders are proficient in reading compared to 87 percent of Georgia’s fourth graders. Similarly, 51 percent of fourth graders in Massachusetts are rated proficient per the NAEP, compared to 23 percent in Louisiana. Yet, 81 percent of Massachusetts schools have failed to make AYP, compared to only 22 percent of Louisiana schools. Overall, the Center for Education Policy estimates that half of all public schools in the United State will be labeled as failing as of 2012.

NCLB set a goal of having a “highly qualified teacher” (HQT) in every public school classroom by the end of the 2005-2006 school year. The HQT standard requires that teachers obtain a state certification or pass a state certification test and not have any “licensure requirements waived on an emergency, temporary or provisional basis.” The HQT standard also requires that all teachers have at least a bachelor’s degree and pass a “rigorous state test” that demonstrates proficiency in the teacher’s subject area. However, states have found it difficult to satisfy the

51 Id.
53 Id.
54 Id.
57 Id.
HQT requirement, citing low salaries, lack of incentive for teachers to work in low-income areas, and teacher shortages in high need areas such as math and science. 58

C. American Recovery and Reinvestment Act of 2009

The American Recovery and Reinvestment Act of 2009 (ARRA) was primarily concerned with stabilizing and stimulating the American economy. 59 However, part of the ARRA included provisions to preserve jobs in education and to lay the groundwork for the Obama Administration’s educational policy priorities. 60 Two components of the ARRA address teacher accountability: RTTT and the TIF. While both are voluntary grant programs, they provide an indication of where the Administration would like to go with teacher accountability. 61 Both programs tie teacher compensation and accountability to the performance of students on state standardized tests.

1. Race to the Top Fund

RTTT made $4.35 billion available for investment in education. 62 RTTT focused on four main reform areas: 1) to “increase teacher effectiveness and address inequities in the distribution of highly qualified teachers”; 2) to establish data systems to track progress from pre-k to college and career readiness in order to improve education; 3) to make progress towards college and career readiness standards; and 4) to implement turnaround strategies for failing schools. 63

59 Superfine, supra note 19, at 97.
60 Id.
61 Id. at 99.
62 Id. at 101.
63 Id. at 100.
Of the four reform areas, increasing teacher effectiveness and developing data systems have direct implications for teacher evaluation. Increasing teacher effectiveness requires that states develop evaluation systems that include student growth as a significant factor. States must use the evaluations to inform professional development decisions, provide incentives for highly effective teachers, determine tenure, and remove ineffective tenured and untenured teachers. The data systems developed must track individual student performance and link that data to individual teachers. The availability of this data allows states to apply the data in teacher evaluations.

Competing states earn points based on their plan to improve each of these areas and their ability to implement their proposals. Increasing teacher effectiveness carries the most weight, making up 28 percent of all available points. Data system development comprises 9 percent. Overall, changes to teacher evaluation systems make up over one-third of all points available to competing states. As of 2012, there have been twelve RTTT winners. Of those, all but two require that teacher evaluations include objective evidence of student learning as

---

64 RTTT defines student growth as change in a student’s test scores over time. Id. at 122.
65 Id. at 123.
66 Id. at 122.
68 Id.
measured by test scores.\textsuperscript{70} Seven require that student growth function as the preponderant criterion in teacher evaluations.\textsuperscript{71}

The RTTT competition has had a national impact on teacher evaluation as states made changes in an effort to compete for the funds.\textsuperscript{72} Two years ago, only sixteen states used test scores in teacher evaluation.\textsuperscript{73} Today twenty-three states and the District of Columbia use test scores in some manner in their teacher evaluations.\textsuperscript{74}

2. Teacher Incentive Fund

The Departments of Labor, Health and Human Services, and Education, and Related Agencies Appropriations Act of 2006 originally authorized the TIF.\textsuperscript{75} The ARRA earmarks $200 million for projects that create financial incentives for effective teachers via the TIF.\textsuperscript{76} The purpose of the TIF is to “support projects that develop and implement PBCs [performance-based compensation systems] for teachers, principals and other personnel…” to improve student achievement in high need schools.\textsuperscript{77} High need schools are defined as those that have 50 percent or

\textsuperscript{70} The two exceptions are Georgia and Hawaii. Hawaii in particular is having difficulty in this area and is threatened with losing its funding. In January 2012, 67 percent of teachers represented by the Hawaii State Teachers Association rejected a contract that would have brought Hawaii into compliance with the RTTT requirements. Valerie Strauss, \textit{Hawaii Teachers Reject Contract in 'Blow' to Race to the Top}, \textit{Washington Post}, January 12, 2012, [online].

\textsuperscript{71} Tennessee, Rhode Island, Ohio, Maryland, Washington DC, and Delaware. See, Stephanie Banchero, \textit{Nearly Half of States Link Teacher Evaluations to Tests}, \textit{Wall Street Journal}, October 26, 2011, [online].

\textsuperscript{72} \textit{Id.}

\textsuperscript{73} \textit{Id.}

\textsuperscript{74} \textit{Id.}


\textsuperscript{76} Teacher Incentive Fund: Notice of Final Priorities, Requirements, Definitions, and Selection Criteria, 75 Fed. Reg. 28,714 (May 21, 2010).

\textsuperscript{77} \textit{Id.}
more of their enrollment comprised of low-income families.\textsuperscript{78} Student achievement is defined by performance on the state’s ESEA assessment as well as other performance measures that are “rigorous and comparable across schools.”\textsuperscript{79}

The TIF has three absolute priorities,\textsuperscript{80} and three competitive preferences.\textsuperscript{81} The first absolute priority is differentiated compensation to school employees based on effectiveness. Effectiveness must give “significant weight to student growth,” must include observations over the course of the year and may include other measures that increase the effectiveness of teachers.\textsuperscript{82} The amount of incentive must be enough to serve as an incentive for teachers and principals.\textsuperscript{83}

\section*{D. ESEA Waiver Plan}

By addressing education in ARRA and in publishing the “Blueprint for Reform,” the Obama Administration attempted to encourage Congress to implement new educational policy and outlined its priorities in education.\textsuperscript{84} Not only have many criticized NCLB for years, NCLB is four years overdue for an overhaul.\textsuperscript{85} In spite of this, states are still held to NCLB’s requirements to keep their Title I funds. States with increasing numbers of failing schools have essentially entered into a

\begin{footnotesize}
\begin{enumerate}
\item Id. at 28,736.
\item Id. at 28,732.
\item Id. at 28,732.
\item Use of value added measures of student achievement, increased recruitment and retention for high need schools and subject areas, and new applicants. Id. at 28,732-33.
\item Id. Other measures that increase the effectiveness of teachers include evidence of leadership roles. Id.
\item Id. Also particularly relevant to the use of test scores in teacher evaluation is the first competitive preference: the use of value added measure of student achievement. See discussion of valued added infra pp. 28–39.
\end{enumerate}
\end{footnotesize}
game of “chicken” with the federal government.\textsuperscript{86} Faced with stagnation in Congress and NCLB’s 2014 deadline rapidly approaching, the Obama Administration utilized a provision of NCLB that empowers the Secretary of Education to grant waivers to the strictures of NCLB if certain threshold requirements are met.\textsuperscript{87}

The ESEA waiver plan waives NCLB’s 2014 AYP proficiency requirement.\textsuperscript{88} Instead, states may set new “ambitious but achievable” goals.\textsuperscript{89} Additionally, states may design their own interventions for failing schools, rather than be limited to those defined in NCLB.\textsuperscript{90} Finally, schools that have not satisfied the HQT requirement may use funds to develop “more meaningful evaluation and support systems.”\textsuperscript{91}

In exchange, states have to adopt college and career readiness standards, focus on the 15 percent of most troubled schools, and implement teacher and principal evaluation systems.\textsuperscript{92} The teacher and principal evaluation systems must contain three categories of evaluation.\textsuperscript{93} Evaluations must:

\begin{itemize}
  \item use multiple valid measures in determining performance levels,
  \item including as a significant factor data on student growth for all students . . . , and other measures of professional practice (which may be
\end{itemize}


\textsuperscript{89} Id., at 4.

\textsuperscript{90} Id., at 1.

\textsuperscript{91} Id., at 2.

\textsuperscript{92} Id., at 3-5.

\textsuperscript{93} Id., at 5.
gathered through multiple formats and sources, such as observations based on rigorous teacher performance standards, teacher portfolios, and student and parent surveys.\textsuperscript{94}

Additionally, evaluations must be given on a regular basis, and must be used to “inform personnel decisions.”\textsuperscript{95}

As of December 2011, eleven states applied for waivers.\textsuperscript{96} Twenty-six more states and Washington D.C. applied for waivers before the February 21, 2012 deadline.\textsuperscript{97}

E. Variations of Teacher Evaluation in the ESEA Reauthorization Process

1. The Senate HELP Committee Draft

The ESEA Reauthorization drafted by the Senate HELP Committee keeps NCLB’s HQT requirement, but does not require states to include testing data in teacher evaluation.\textsuperscript{98} However, assessments must “measure individual academic achievement” and individual student growth.\textsuperscript{99} An early version of the bill contained a provision mandating the inclusion of student achievement in teacher evaluations.\textsuperscript{100} However, the HELP Committee removed the provision because

\textsuperscript{94} Id., emphasis added.
\textsuperscript{95} Id.
\textsuperscript{96} Id.
\textsuperscript{97} Dep’t of Education, ESEA Flexibility, http://www.ed.gov/esea/flexibility (last visited March 5, 2012.)
\textsuperscript{98} Id. at § 2141(Comm. Print 2011) available at http://www.help senate.gov/imo/media/doc/ROM117523.pdf.
\textsuperscript{99} Id. at §1112(2)(A)(II)
Senate Republicans protested that the bill was intruding on an area of traditional state and local control.101

In an unusual move, the HELP Committee held a round table hearing on the ESEA Reauthorization after voting the bill out of committee.102 During the hearing, members of the round table discussed teacher evaluation.103 However, while many of the speakers lauded the use of student test scores to evaluate teachers, the primary focus was on teacher evaluation as a matter for local authorities.104

In addition, the HELP Committee ESEA Reauthorization codifies the TIF and RTTT.105 These programs remain voluntary competitive grant programs. However, given the impact these programs have already had on teacher evaluation, codifying the programs in the ESEA will continue, if not increase, the pressure on states to use student test scores in teacher evaluations.

2. The House Education and the Workforce Committee Draft

The Republican leadership of the House Education and Workforce Committee (“House Committee”) passed out of committee two proposals on February 28, 2012.106 Rather than propose and approve one large bill, the House of Representatives has approached education renewal in a more piecemeal form.107 The Student Success Act and the Encouraging Innovation and Effective Teachers Act (“Effective Teachers Act”) would work together to scrap the HQT requirement of

101 Id.
103 Id.
104 Id.
107 See id.
NCLB. While reducing the federal role in almost every other respect, the Effective Teachers Act would require states to tie teacher evaluations to student test performance.

III. Critique of Using Student Test Scores to Evaluate Teachers

A. The Use of Standardized Test Scores to Evaluate Teachers is Not Supported by Research

1. The Use of Test Based Incentives in Education

For over ten years now, NCLB has attempted to increase student achievement using test results to reward successful schools and punish failing ones. However, a report by the National Research Council of the National Academies of Science (NRC) concluded that test based incentives have not been effective in education. The report reviewed the results of numerous studies of various test-based incentive programs. The report provided three recommendations. First, the NRC recommended further research on test-based incentives. The report noted, “[d]espite using them for several decades, policy makers and educators do not yet know how to use test-based incentives to consistently generate positive effects on achievement and to improve education.” Second, new test-based incentive program design should help develop potential alternatives for incentives and

108 See id.
109 Id.
111 The broad scope covered high stakes high school exit exams, AYP requirements of NCLB, teacher evaluation programs, and international models as well. Id. at 87, tbl.4-1A.
112 Id. at 92
accountability. Finally, research should extend beyond the information provided by the tests administered to evaluate factors that make test-based incentive programs more effective.\textsuperscript{114}

2. The Use of Test-Based Incentives in Teacher Assessment

In order to measure student growth to evaluate teachers, a statistical model must be used to assess the data.\textsuperscript{115} One model looks at groups of students from year-to-year.\textsuperscript{116} This “cohort to cohort” model evaluates students in a particular grade and their performance over time.\textsuperscript{117} This model is a component of NCLB’s safe harbor provision, which allows credit to schools for improving a cohort’s percentage of proficient students, even if the school has not yet achieved the state’s target.\textsuperscript{118}

Another model, the “growth model,” looks at the individual data of student test scores from multiple standardized tests over time.\textsuperscript{119} One difficulty with using growth model data is that many factors that are beyond a teacher's influence have a significant impact on test scores.\textsuperscript{120} A specific type of growth model, known as a “value added model” (VAM), attempts to control statistically for differences between students in order to quantify teacher and school effectiveness.\textsuperscript{121} The use of VAMs is gaining in popularity and supported by the Obama Administration.\textsuperscript{122}
The L.A. Times used a VAM when it analyzed the test scores of LAUSD students. The basic idea behind VAMs is to look at the changes in a student’s test score from year-to-year, focusing on the changes in the test score rather than the score itself. A predicted score is calculated using certain factors. This predicted score is then compared to the obtained score. The difference between the scores (the "residual") is determined to be the measure of the teacher's effectiveness. If the score is positive, then the teacher is considered to have increased student learning. If it is negative, then the teacher has not improved student learning.

There are some serious concerns about the use of VAMs to analyze student test scores and apply those results to teacher evaluations. Research has shown VAMs to be highly unstable, fluctuating wildly from year to year. VAMs incorrectly assume that schools assign students to teachers randomly. VAMs cannot account for what external factors influence a student’s score. Finally, teaching to the test can skew VAMs.

a. VAMs are highly unstable

One concern raised by the use of VAMs to evaluate teachers is the wild fluctuations in a teacher's effectiveness rating that may occur. This variability occurs from "class to class and from year to year, as well as from one statistical

123 Id.
124 Ewing, supra note 5.
125 Briggs, supra note 15.
126 Id.
127 Id.
128 Id.
129 Id.
131 Haertel, supra note 130.
model to the next. . .”132 Wide variability is also present when different tests are used, even in the same content area.133

A study of data from five school districts showed that of teachers that scored in the bottom 20 percent one year, only 20 to 30 percent of them had similar ratings the next year, while 25 to 45 percent of them scored well above average.134 Another study demonstrated that from year to year 74 to 95 percent of teachers’ effectiveness ratings changed by at least one decile, 45 to 63 percent changed by at least two deciles, and 19 to 41 percent changed by at least three deciles.135

An evaluation of the L.A. Times VAM demonstrated significant fluctuations in ratings if other legitimate variables were considered.136 By using the additional variables,137 only 46.4 percent of teachers maintained the same reading outcomes and 60.8 percent maintained their math outcomes.138 In addition, 8.1 percent of teachers classified as "more" or "most" effective in the alternative model would shift to "less" or "least effective" under the L.A. Times model.139 Additionally, 12.6 percent would shift in the other direction from "less" or "least effective" to "more" or "most effective."140 While this does not demonstrate that it is correct to use one set of variables as opposed to another or that one model is more valid than another, it is important to note how the choice of variables factored have significant consequences.

132 Id. (emphasis in original).
133 Id. at 3.
134 Id. at 2.
135 Id.
137 Subject specific test scores in grades two and three, mean of fourth grade test scores in fifth grade, and an indicator for the schools placement in California’s School Similarity Rank. Id.
138 Id. at 14.
139 Id.
140 Id.
for the results. This is particularly true when the results will have high stake consequences.

b. Students are not randomly assigned to teachers

To be effective, "VAMs require students to be assigned randomly."[^141] To determine whether the differences between the predicted and obtained scores are a result of a teacher's effectiveness, any VAM must control for other variables that may contribute to a student's test score.[^142] However, statistical models cannot totally compensate for teachers that have a disproportionate number of students with "greater challenges"[^143] or those whose performance on traditional tests may not "accurately reflect learning."[^144]

If students are assigned to certain teachers based on variables that are not accounted for by the model, then the results are questionable and will result in biased results.[^145] One way to test for this bias is by looking at a teacher's effects on test scores of students in grade levels before they have the student.[^146] It makes sense, for example, that a fifth grade teacher should have no impact on a student's third grade test score. Yet, when this analysis is conducted, "studies . . . have shown large 'effects' . . ."[^147] on a number of VAMs.[^148] A review of the L.A. Times VAM found that the impact of fourth and fifth grade teachers on third grade students was just

[^141]: Haertel, supra note 130, at 3.
[^143]: E.g., poor attendance, homelessness, or severe home problems. Haertel, supra note 130, at 3.
[^144]: Such as ELL students or students with special needs. Id.
[^146]: Id.; Haertel, supra note 130, at 3.
[^147]: Haertel, supra note 130, at 3.
[^148]: Id.
as great their impact on their present classes. This result "raises a red flag against drawing the conclusions that the [L.A. Times VAM] is producing unbiased estimates of the causal effects of teachers on their students."150

Another study looked at correlations between teacher effectiveness ratings and various student factors such as ethnicity, language background, parent education, and income.151 Even after controlling for prior student achievement and student characteristics, a high correlation was found between these factors and teacher effectiveness ratings.152 If high stakes incentives are attached to teacher evaluations based on test scores, there is a very real risk that teachers will be discouraged from serving students with high needs.153 This could also increase already existing inequalities in schools as teachers would seek to avoid schools with needy student populations, and would seek to prevent certain students from being in their classes.154

c. It is impossible to separate out all the influences on student learning

It is impracticable to sort out the various in-school effects on student progress in order to arrive at a true measure of a teacher's effectiveness.155 The influence of prior teachers, teachers of other subjects, tutoring, and special support provided by a school with abundant resources all contribute to a student's learning.156 The costs

149 Briggs, supra note 15, at 11.
150 Id. at 12.
151 Haertel, supra note 130.
152 In one extreme example, an English teacher jumped from the lowest decile, to the highest decile the next year. The class he taught in the second year contained fewer English learners, Hispanic students and low-income students. The second year class also contained many more students with well-educated parents. Id. at 4.
153 Id.
154 Id.
155 Id.
156 Id.
and benefits of these factors to a child's education may serve to make a teacher appear more or less effective than another.\textsuperscript{157} Research into the use of VAMs to make decisions about bonuses and teacher dismissals indicates not only are the teacher effectiveness ratings impacted by the above factors, but also drastically differ with the observations of skilled observers.\textsuperscript{158}

Houston Independent School District is currently using a VAM to determine merit pay and termination decisions.\textsuperscript{159} One teacher terminated because of her VAM rating was named "Teacher of the Year" in 2008 and had been awarded "Teacher of the Month" in 2010.\textsuperscript{160} Prior to the implementation of the VAM system, she was consistently rated as "exceeding expectations."\textsuperscript{161} Another teacher was also consistently rated as "exceeding expectations" or "proficient," but her rating made a significant drop when a large number of English language learners were transitioned into her class.\textsuperscript{162}

The receipt of bonuses appears so arbitrary that teachers liken it to "winning the lottery."\textsuperscript{163} Teachers are unable determine the relationship between their instructional practices and their value added scores.\textsuperscript{164} Teachers reported

\footnotesize

\textsuperscript{157} Id.
\textsuperscript{158} Id. at 5.
\textsuperscript{159} Id.
\textsuperscript{160} Id.
\textsuperscript{161} Id.
\textsuperscript{162} Id.
\textsuperscript{164} Haertel, \textit{supra} note 130, at 5.
attempting to avoid teaching certain subjects and students in order to boost their scores, and reported they were "confused and demoralized by the system." \[165\]

d. Teaching to the test

Standardized tests measure only a subset of a whole domain that is measured. \[166\] That is, any given test can measure only a fraction of the subject material it is supposed to assess. \[167\] In addition, content that requires higher cognitive skills and application is often omitted from standardized testing due to the difficulty in designing items that would survive pilot testing. \[168\] The “missing segment would typically be the portion of the curriculum that deals with higher levels of cognitive functioning and application of knowledge and skills.” \[169\] Thus, teacher evaluation that relies on student test scores provides a very limited measure of that teacher’s ability to teach application of knowledge and skill, as well as engage students’ higher level cognitive functioning.

In addition, score inflation may occur when overly narrow instruction occurs in order to raise test scores. This so-called teaching to the test occurs often to the detriment of the untested material. \[170\] The fundamental problem of score inflation that occurs when teachers teach to the test is a result of incentives that are

\[165\] Id.
\[166\] Nat’l Research Council, supra note 110, at 39.
\[167\] For example, an eleventh grade social studies standard is as follows: “Discuss the significant domestic policy speeches of Truman, Eisenhower, Kennedy, Johnson, Nixon, Carter, Reagan, Bush and Clinton.” The only test item assessing that standard gives a brief quote from President Johnson, then asks, “The program President Lyndon B. Johnson created to wage his unconditional war on poverty was the (A) Alliance for Progress; (B) Fair Deal; (C) Great Society; (D) New Deal.” Haertel, supra note 163, at 24–25.
\[168\] Id.
\[169\] Id.
\[170\] Id. at 40.
attached to a test that does not fully represent the outcomes that are desired.\textsuperscript{171} In other words, if incentives are tied to a test score, teachers may teach to the test in order to obtain the incentive, while not improving a child’s education.

B. The Use of Student Test Scores is a Violation of Substantive Due Process

The Due Process Clause of the Fourteenth Amendment declares that no state may “deprive any person of life, liberty, or property, without due process of law.”\textsuperscript{172} A party alleging a deprivation of substantive due process must establish that state action is depriving the party of life, liberty, or property. If life, liberty, or property is at stake, the court must determine if the right interfered with by the state action is a fundamental one.\textsuperscript{173} If the right is not fundamental, then the court will review the state action to determine if it is rationally related to a legitimate state interest.\textsuperscript{174} The state action will not be upheld under rational basis if the claimant can demonstrate the state action is arbitrary and capricious.\textsuperscript{175}

The analysis below aims to provide a broad framework under which a teacher may claim a violation of substantive due process rights. In application, this is a fact intensive analysis that relies on the specific agreements between the parties, i.e. the employment contract, school district policies and procedures, and relevant state statutes.

\textsuperscript{171} Id. at 42.
\textsuperscript{172} U.S. Const. amend. XIV, § 1.
\textsuperscript{175} Moore v. Warwick Pub. Sch. Dist. No. 29, 794 F.2d 322, 329 (8th Cir. 1986).
According the United States Supreme Court, “the meaning of ‘liberty’ must be broad indeed.”\textsuperscript{176} Liberty covers “not merely freedom from bodily restraint but also the right of the individual to contract [and] engage in any of the common occupations of life.”\textsuperscript{177} Liberty, therefore, is at issue when an individual’s “good name, reputation, honor, or integrity” is at stake.\textsuperscript{178} Liberty is at stake when state action imposes a stigma or disability that eliminates the individual’s ability to pursue other opportunities for employment.\textsuperscript{179} Courts have also found a liberty interest in the right to engage in a profession free from unreasonable government interference.\textsuperscript{180}

Property interests are created by statute or contract.\textsuperscript{181} A teacher has a property interest in continued employment during the term of a contract.\textsuperscript{182} A teacher also has a property right in continued employment pursuant to a written contract with a tenure provision.\textsuperscript{183} Absent an explicit tenure agreement, a teacher may have a property interest in continued employment because of a reasonable expectation of continued employment based on customs, practice, or mutual understanding.\textsuperscript{184} Teachers also have a property right in their teaching certificates.\textsuperscript{185}

\textsuperscript{176} Bd. of Regents of State Colleges v. Roth, 408 U.S. 564, 572, (1972).
\textsuperscript{177} Id.
\textsuperscript{178} Id. at 573.
\textsuperscript{179} Id.
\textsuperscript{180} United States v. LULAC, 793 F.2d 636, 647 (5th Cir. 1986)
\textsuperscript{181} Bd. of Regents of State Colleges, 408 U.S. at 577.
\textsuperscript{182} Id. at 576.
\textsuperscript{183} Perry v. Sinderman, 408 U.S. 593, 601 (1972).
\textsuperscript{184} Id.
Few cases have directly addressed the use of student test scores in teacher evaluation.\textsuperscript{186} In \textit{Scheelhaase v. Woodbury Central Community School District},\textsuperscript{187} a teacher was not offered a new contract because of the low test scores of her students. The teacher argued that the failure to offer a new contract for that reason was arbitrary and capricious.\textsuperscript{188} The trial court found that a “teacher's professional competence [can] not be determined solely on the basis of her students' achievements on the [test].”\textsuperscript{189} The trial court then found a property interest in her contract renewal and ordered reinstatement and damages.\textsuperscript{190} The 8th Circuit reversed, essentially finding she had no property right to a renewed contract. As such, the school board had the discretion to “decline to employ or to re-employ any teacher for any reason or for no reason,” as long as it was not constitutionally impermissible.\textsuperscript{191} \textit{St. Louis Teachers Union v. Board of Education}\textsuperscript{192} held that teachers stated a claim for a deprivation of their due process rights where they were rated “unsatisfactory” because of their students’ test scores.\textsuperscript{193} The teachers successfully alleged that the use of student test scores in their evaluations was arbitrary and capricious.\textsuperscript{194} Both cases support the argument that the use of student test scores to evaluate teachers may be arbitrary and capricious. The


\textsuperscript{188} \textit{Id.} at 241.

\textsuperscript{189} \textit{Id.} at 239.

\textsuperscript{189} \textit{Id.}

\textsuperscript{190} \textit{Id.}

\textsuperscript{191} \textit{Id.}

\textsuperscript{192} 652 F.Supp. 425 (E.D.Mo. 1987).

\textsuperscript{193} \textit{Id.} at 435.

\textsuperscript{194} \textit{Id.}
instability in a teacher’s VAM score as well as the arguments as to why the test scores do not really measure a teacher’s effectiveness\textsuperscript{195} lend strength to the proposition that such policies are arbitrary and capricious.

Teachers may challenge the use of test scores as arbitrary and capricious if they can establish that the test covers material that has not been taught.\textsuperscript{196} In \textit{Debra P. v. Turlington}, Florida enacted a statute that made students’ promotion from one grade to the next and receipt of a diploma contingent on test scores.\textsuperscript{197} The court found the Florida Department of Education had not sufficiently determined whether the “test administered measure what was actually taught in the schools of Florida.”\textsuperscript{198} The court stated that under due process, “the state is obligated to avoid action which is arbitrary and capricious, does not achieve or even frustrates a legitimate state interest, or is fundamentally unfair.”\textsuperscript{199} The court found the test to be “fundamentally unfair in that it \textit{may} have covered matters not taught in the schools of the state.”\textsuperscript{200}

Similarly, it would be unfair to evaluate teachers based on student test scores on tests not aligned with what is actually taught. With the recent adoption by most states of the Common Core Standards, states will have to develop new tests, and

\textsuperscript{195} See supra Part III.A.  
\textsuperscript{196} Debra P v. Turlington, 644 F.2d 397 (1981).  
\textsuperscript{197} Id. at 400.  
\textsuperscript{198} Id. at 405.  
\textsuperscript{199} Id. at 404.  
\textsuperscript{200} Id. (emphasis in original).
subsequently be sure they are valid in that they actually measure what they claim to test. Tests, when selected, must accurately reflect what is actually taught.\textsuperscript{201}

The widespread allegations of cheating add an additional element of arbitrariness. A teacher may receive a poor rating based on a VAM score not as a result of ineffective teaching, but rather, because the teacher in a previous year cheated, thereby providing an inflated score not reflective of the students’ abilities and setting a standard that a subsequent teacher could not satisfy.

In sum, given the research that currently exists, there seem adequate grounds for teachers to challenge the use of student test scores in their evaluations. The use of test scores may appear rationally related to the legitimate state interest in having qualified educators. However, as statutes mandate that the test scores overwhelm all other potential measures of a teacher’s effectiveness, their use results in increasingly arbitrary employment decisions and violations of substantive due process.

States have already encountered litigation because of their efforts to compete for RTTT money.\textsuperscript{202} For example, litigation in New York arose because of the weight student test scores were to be given in teacher evaluation.\textsuperscript{203} The regulations propounded by the state to comply with the RTTT required 40 percent of a teacher’s

\begin{footnotesize}
\begin{enumerate}
\item In Buffalo, New York, schools stand to lose millions of dollars because the Buffalo Teachers Federation voted to exclude test results from students that have missed 20 percent of the school year. Kendra Eaglin, \textit{Buffalo Teachers Vote Down Absentee Item}. \textit{Buffalo Business First}. March 8, 2012 http://www.bizjournals.com/buffalo/blog/morning_roundup/2012/03/buffalo-teachers-vote-down-absentee-item.html. Students that miss nearly two months of the school year will almost certainly be tested on material they have not been taught. To hold teachers accountable for student attendance, which is obviously out of their control, seems the kind of patent unfairness substantive due process would address.

\item The Georgia Supreme Court invalidated a state law enacted to create charter schools in order to comply with RTTT because it violated Georgia’s constitution. Gwinnett County Sch. Dist. v. Cox, 710 S.E.2d 773, 775 (2011), \textit{reconsideration denied} (June 13, 2011).

\end{enumerate}
\end{footnotesize}
evaluation to be measured by student test scores. At issue was interpretation of a New York statute that required 20 percent of a teacher’s evaluation be comprised of student test scores and 20 percent by another measure. The regulations propounded by the Board of Regents allowed districts, at a local level, to use test scores as another measure, meaning that up to 40 percent of a teacher’s evaluation could be based on a test score. Given the point allocation the Board of Regents developed, the court found that “a teacher or principal could be deemed ‘ineffective’ on the basis of a single standardized state test.” This violated the statute’s requirement that a teacher’s evaluation consist of multiple measures.

C. The Use of Student Test Scores to Evaluate Teachers Will Result in Increased Instances of Cheating

Campbell’s law states:

The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor.

The events in Atlanta, Georgia, are a prime example. One hundred seventy-eight teachers and principals at forty-four Atlanta, Georgia, public schools were

---

204 Id. at 702.
205 Id.
206 Id. at 706. An evaluation score of sixty-four out of a possible 100 points would result in a teacher rating of “ineffective.” Given that 0-4 points would be earned if a teacher’s test scores failed to improve and sixty points could be earned if the teacher scored perfectly on all other evaluation criteria, a teacher could receive a rating of ineffective based solely on the test results. Id.
207 Id.
implicated in cheating on the state’s standardized tests. Teachers in Atlanta are evaluated by the performance of their students on standardized tests. Employees at Atlanta schools with the highest test scores may receive a $2,000.00 bonus. Contrariwise, teachers with low scoring students may be put under intense pressure to raise those scores or else risk losing their jobs. Similar allegations on a smaller scale have been made in Washington, D.C., Baltimore, Pennsylvania, and New Jersey. The L.A. Times recently reported of cheating allegations against nearly forty California teachers. Obviously, cheating invalidates any given set of scores for use in teacher evaluation. Further, honest teachers will be penalized because they will be required to raise test scores that are artificially inflated because a previous teacher or administrator cheated.

IV. Alternatives to the Use of Student Test Scores to Evaluate Teachers

A. Implement Performance Assessments

The National Board Certification (NBC) established by the National Board for Professional Teaching Standards (NBPTS) uses performance assessments to assess experienced teaching candidates. Performance assessments generally

210 Id.
211 Id.
212 Id.
require teachers to document lesson plans, videotape the lesson, and collect and evaluate evidence of student learning. All of the components are compiled to create a portfolio for assessment. Assessment of the portfolio is based on teaching standards developed by NBPTS. The NBC performance assessments identify teachers that are more effective at raising student achievement than others.

Working off the NBC framework, states have developed assessment systems for use with new teachers. For example, California has developed the Performance Assessment for California Teachers (PACT). Currently, thirty-three California teacher education programs use PACT. PACT has been found to be “reliable, valid, and a strong lever for improving both teacher competence and program quality.”

Utilizing performance assessments both at the licensing stage and then over the course of a teacher’s career would provide a solid basis for teacher evaluation. Performance measures provide concrete evidence in the form of portfolios and allow a more nuanced assessment of a teacher’s effectiveness than does a test score. In addition, performance evaluations allow teachers to reflect and grow by virtue of the evaluation process.

217 Id.
219 Darling-Hammond, supra note 216, at 8. Specifically, studies in North Carolina and Florida found that students with NBC teachers made greater gains in achievement. A similar study in LAUSD also found significant positive effects. Id.
220 Id.
221 Id. at 10. PACT requires the intern or student teacher to plan a week long unit based on state standards: assess each daily lesson and make needed revisions for the next day; analyze videotapes of them teaching, reflect on what was successful, what was not, and why; and set out what they would do differently in a future unit. Id. at 14.
222 Id.
B. Look to International Teaching Models

The Programs for International Student Assessment (PISA) is an assessment administered every three years.\textsuperscript{223} In 2009, nine nations scored higher than the United States on the reading literacy section.\textsuperscript{224} Twenty-three scored higher than the United States in mathematics literacy,\textsuperscript{225} and eighteen scored higher in science literacy.\textsuperscript{226} Of the highest scoring nations, none have implemented any of the major educational reform ideas dominating reform discussions in the United States.\textsuperscript{227} Additionally,

The idea of grade-by-grade national testing has no takers in the top-performing countries. . . Typically, there are state or national tests only at the end of primary or lower secondary education, and at the end of upper secondary school. Schools and the teachers in them are expected to assess their students regularly as an indispensable aid to good teaching, but the assessments given between gateways are not used for accountability purposes, as the basis of teachers’ compensation or to stream or track students.\textsuperscript{228}

\textsuperscript{223} Department of Education, Highlights from PISA at iii (2010) available at http://nces.ed.gov/pubs2011/2011004.pdf. PISA tests reading, math, and science. The 2009 test was administered to 15 year olds in "60 nations and 5 other education systems." Id.

\textsuperscript{224} South Korea, Finland, Canada, New Zealand, Japan, Australia, Shanghai-China, Hong Kong-China, and Singapore. Id. Tbl.3 at 8–9.

\textsuperscript{225} Finland, South Korea, Canada, New Zealand, Japan, Estonia, Australia, Netherlands, Belgium, Iceland, Denmark, Norway, France, Slovak Republic, Germany, Switzerland, United Kingdom, Slovenia, Shanghai-China, Hong Kong-China, Singapore, Chinese Taipei, Lichtenstein, and Macao-China. Id. Tbl.7 at 18.

\textsuperscript{226} Finland, South Korea, Canada, New Zealand, Japan, Estonia, Australia, Netherlands, Germany, Switzerland, United Kingdom, Slovenia, Shanghai-China, Hong Kong-China, Singapore, Chinese Taipei, Lichtenstein, and Macao-China Id. Tbl.8 at 24.

\textsuperscript{227} This includes charter schools, vouchers, involvement of entrepreneurs in education, and the use of student test scores to reward and punish teachers and principals. Marc S. Tucker, Standing on the Shoulders of Giants 2 (National Center on Education and the Economy 2011).

\textsuperscript{228} Id. at 8.
To assure teacher quality, the highest performing school systems focus on teachers that have a “high level of general intelligence,” a firm grasp on their content area, and the ability to engage students in learning. The highest scoring nations use various tools to satisfy these requirements. All set high standards for admission into teacher preparation programs and offer teachers compensation that is competitive with other fields. Only high quality, major universities provide teacher education and study focuses on pedagogy and content. Mentoring continues after hiring and educators research and direct the improvement of the system. All of the above approaches improve the teacher workforce without implementing harsh measures based on student test scores. In fact, the obsession with test scores prevalent in the United States is conspicuous by its absence in other high performing nations.

C. Implement a Comprehensive Plan to Elevate Teaching as a Profession

Many cultural and economic differences make parroting another nation’s school system impractical and undesirable. However, there are key

---

229 In this case Shanghai, Finland, Singapore, Canada and Japan. Id. at 4.
230 Id. at 10. Tucker seems to be using the phrase “general intelligence” rather loosely here. In psychology, general intelligence is equated to IQ or some measureable level of “intelligence.” However, the definition of intelligence is not necessarily so clear-cut, nor does a high IQ necessarily indicate a great teacher. As Tucker adds, employers are looking for people with “drive, tenacity, and capacity for hard work,” as well as individuals who can learn tasks quickly and work independently, none of which necessarily equate to a high IQ.
231 Id.
232 Id. at 11.
233 Id. at 15.
234 Id. at 16.
235 For example, Finland requires all teachers to have a master’s degree in education. At the undergraduate level, teaching candidates are expected to obtain two minors in the content areas they will teach. “[T]he training of American medical doctors rests firmly on the very elements . . . described as the basis of training of Finnish and Shanghai educators.” Id. at 19.
236 Id. at 22.
concepts that would make the heavy-handed evaluation regimes propounded by various federal and state programs unnecessary.

The overall goal is to raise teaching to the level of a profession on par with doctors and lawyers. Several steps are needed to attract more qualified teaching candidates. Currently, elementary level teachers tend to have below average high school grade point averages and SAT scores. Encouraging prospective students with higher grade point averages and SAT scores may at least increase the perception that acceptance into an teacher’s education program carries with it a certain amount of prestige. States need to eliminate short-term emergency credentialing programs that give the impression that teaching is something that an individual can learn in a six-week course or with minimal effort. State licensing agencies could increase the baseline requirements to obtain a credential thereby elevating the prestige associated with a teaching credential and assuring qualified individuals become teachers.

The federal government should provide grant money to states that offer scholarships or stipends to highly qualified students seeking to become teachers. Teacher pay needs to be elevated to that of other comparable professions in order to encourage potential teachers that my otherwise choose other professions.

237 Stephen Sawchuk, Teacher Quality, Status Entwined Among Top Performing Nations, Education Week (January 9, 2012), http://www.edweek.org/ew/articles/2012/01/12/16teachers.h31.html?tkn=SPZFPdty3lWmLeb2QDOUFlogiKKAR8SPCIQ6&cm p=ENL-EU-NEWS1. It is important to note that SAT scores do not measure whether someone will be a good teacher. SAT scores merely attempt to predict a potential student’s performance in the first year of an undergraduate program.
More time and effort needs to be invested in professional development. Interestingly, teachers in Finland and South Korea spend about 600 hours a year in face-to-face student instruction, compared to 1,100 hours per year in the United States.\footnote{Id.} Finland, South Korea, and other top performing nations use that time planning lessons with colleagues.\footnote{Id.} Increased professional development could address issues with the current teacher workforce, as was done in Canada.\footnote{Tucker, supra note 198, at 43-44.} Federal and state programs could provide incentives for teachers seeking to obtain master’s degrees or bachelor’s degrees in high need areas such as math and science.

Finally, the United States needs to stop pushing ineffective test based reforms. There is no evidence that high stakes testing results in a better-educated population. To the contrary, high stakes testing has resulted in widespread cheating, narrowing of curriculum, the demoralization of our teaching workforce, and ultimately, a less competitive nation. Much educational reform, however, seeks to mire the nation even further in the morass of data driven accountability. The United States is investing in a “get rich quick scheme,” educationally speaking. Unfortunately, ten years of NCLB and the United States is no “richer.” In fact, the United States in on track to become much poorer, as the nation invests millions of dollars into teacher evaluation programs that do not make teachers any better or result in better education for our children.
V. Conclusion

Research indicates that there is simply not enough support behind the use of student test scores to evaluate teachers. As such, programs that do so are arbitrary and capricious. The implementation of such programs violates the substantive due process rights of American teachers. Additionally, it is irresponsible to evaluate teachers in this manner. The financial cost on already strained state budgets is enormous. The states face a potentially huge financial outlay with little or no projected benefit.

More important than the financial aspect, is the damage done to America’s educators and students. Students will be robbed of an education as desperate teachers focus on teaching test material in an effort to keep their jobs. Teachers will be faced with a system that rewards and punishes them arbitrarily. In the end, this will create a hostile work environment that, rather than encouraging America’s best and brightest to become educators, will instead drive individuals into other fields. There is the possibility that litigation ultimately will have to resolve the issue. It would be better for everyone involved if policy makers heed the warnings present in the research and turn from this path before the United States wastes more resources and further damages American education. Instead, programs should focus on providing meaningful assessment that treats teachers as professionals. In that way the Unites States can improve its educational system through the collaboration of parents, teachers, and the government.