Here comes the monkey! : best practices for using student test scores as part of the teacher evaluation process based on an analysis of legislation and litigation in early-adopting Race to the top states

Julie Dausey Oziemkowski

Follow this and additional works at: https://huskiecommons.lib.niu.edu/allgraduate-thesesdissertations

Recommended Citation
Oziemkowski, Julie Dausey, "Here comes the monkey! : best practices for using student test scores as part of the teacher evaluation process based on an analysis of legislation and litigation in early-adopting Race to the top states" (2018). Graduate Research Theses & Dissertations. 3364.
https://huskiecommons.lib.niu.edu/allgraduate-thesesdissertations/3364

This Dissertation/Thesis is brought to you for free and open access by the Graduate Research & Artistry at Huskie Commons. It has been accepted for inclusion in Graduate Research Theses & Dissertations by an authorized administrator of Huskie Commons. For more information, please contact jschumacher@niu.edu.
ABSTRACT

HERE COMES THE MONKEY! BEST PRACTICES FOR USING STUDENT TEST SCORES AS PART OF THE TEACHER EVALUATION PROCESS BASED ON AN ANALYSIS OF LEGISLATION AND LITIGATION IN EARLY-ADOPTING RACE TO THE TOP STATES

Julie D. Oziemkowski, Ed.D.
Department of Leadership, Educational Psychology, and Foundations
Northern Illinois University, 2017
Jon Crawford, Co-Director
Christine Kiracofe, Co-Director

Numerous lawsuits have been filed against school officials since the first Race to the Top (RTTT) monies were awarded in 2011, requiring school districts in participating states to incorporate student scores on standardized tests into their teacher evaluation processes. This study researched the history of standardized testing in the United States from Horace Mann to Common Core, the timeline of education reform in the United States, changes to state statutes in response to the federal Race to the Top program, and litigation filed in early-adopting RTTT states in response to the student growth component. It concludes with findings that provide school district administrators with a framework for avoiding litigation—and improving student learning—when applying student test scores to the teacher evaluation process.
HERE COMES THE MONKEY! BEST PRACTICES FOR USING STUDENT TEST SCORES AS PART OF THE TEACHER EVALUATION PROCESS BASED ON AN ANALYSIS OF LEGISLATION AND LITIGATION IN EARLY-ADOPTING RACE TO THE TOP STATES

BY

JULIE DAUSEY OZIEMKOWSKI
©2018 Julie Dausey Oziemkowski

A DISSERTATION SUBMITTED TO THE GRADUATE SCHOOL IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE DOCTOR OF EDUCATION

DEPARTMENT OF LEADERSHIP, EDUCATIONAL PSYCHOLOGY AND FOUNDATIONS

Doctoral Directors:
John Crawford
Christine Kiracofe
ACKNOWLEDGEMENTS

First and foremost, I would like to thank my husband, Jeff Oziemkowski. When several colleagues and I were pondering whether to begin our doctoral studies, you not only readily offered your support—which you have given so generously throughout the process—but went so far as to tell me I would be insane not to seize the opportunity. Heartfelt thanks go to my parents, Gary and Barbara Dausey. If you had not lovingly guided this now-reformed underachiever during my nursery school through high school years and sacrificed to put me through college, I would not be where I am today. Thanks are also due to my former principal Gene Sikorski, who looked at a twenty-something teacher with lava lamps and beanbags in her classroom and saw a future leader. Dr. Jon Crawford, Dr. Christine Kiracofe, and Dr. Kelly Summers, I cannot thank you enough for welcoming me to dissertation workshops as a “stowaway” just two months into coursework—not to mention the countless patient explanations, pages of feedback, and words of encouragement.

Colleagues Mary Davis, Derick Edwards, Jon Pilkington, and Chris Silagi, the doctoral journey has been easier with you along for the ride. Thank you, Faith Dahlquist, Dr. Jeff Schuler, and the Wheaton-Warrenville Community Unit School District 200 Board of Education, for providing us with this opportunity. I also appreciate the warm encouragement I received from friends each time I posted my progress on social media. Lastly, I have been inspired by my great-grandfather Edwin Salisbury Carman, an engineering industrialist who, I learned, also had a passion for the public schools. His words—which I encountered in early twentieth century...
documents during my research—were prescient: “A new day is upon us...We must launch out into the current of modern activities, or the current will strand us upon the reef.”
DEDICATION

This dissertation is dedicated to the staff of Wheaton Warrenville Community Unit School District 200—both my childhood teachers and my colleagues of twenty-five years. Long before test scores became part of your evaluations, you had an appetite for data, a drive for continuous improvement, and a heart for all students that make it a joy for me to come to work every day.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>LIST OF FIGURES</th>
<th>viii</th>
</tr>
</thead>
</table>

## Chapter

1. STANDARDIZED TESTING IN THE SPOTLIGHT ................................................................. 1
   - Problem Statement ........................................................................................................ 7
   - Research Questions ........................................................................................................ 8
   - Procedures ..................................................................................................................... 9
   - Significance of the Study .............................................................................................. 9
   - Delimitations .................................................................................................................. 11
   - Limitations .................................................................................................................... 12

2. TWO SUITS AND A SESQUICENTURY OF REFORM .......................................................... 13
   - Common School Movement ............................................................................................ 14
   - Scientific Management Era .......................................................................................... 23
   - 1930s to Mid-1950s: The Technocratic Meritocracy ..................................................... 41
   - Mid to Late 1950s: Era of Excellence ........................................................................... 65
   - Mid 1950s to Early 1960s: Technology, Teaching, & Testing Intersect ....................... 76
   - Mid 1960s: Era of Equity ............................................................................................... 89
   - Mid to Late 1960s: Equity Meets Accountability ......................................................... 106
   - Late 1960s through 1970s: Education Takes a Back Seat ............................................ 117
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980s: Excellence Returns</td>
<td>119</td>
</tr>
<tr>
<td>1990s: Lofty Goals, Unrequited</td>
<td>131</td>
</tr>
<tr>
<td>2000s: Equity Meets Excellence Via Bipartisan Centrism</td>
<td>142</td>
</tr>
<tr>
<td>New York’s Statutory Response to Race to the Top</td>
<td>175</td>
</tr>
<tr>
<td>Emergency Regulations in Response to Education Law 3012-c</td>
<td>177</td>
</tr>
<tr>
<td>New York Litigation in Response to Emergency Regulations</td>
<td>178</td>
</tr>
<tr>
<td>“Ground-Breaking Agreement”</td>
<td>183</td>
</tr>
<tr>
<td>New York Legislation to Address Unresolved Issues</td>
<td>184</td>
</tr>
<tr>
<td>Legislators Ask for Delay</td>
<td>193</td>
</tr>
<tr>
<td>Second Wave of Legislation Emerges</td>
<td>194</td>
</tr>
<tr>
<td>Florida’s Statutory Response to Race to the Top</td>
<td>202</td>
</tr>
<tr>
<td>Florida’s Value-Added Formula</td>
<td>206</td>
</tr>
<tr>
<td>Florida Local School District Evaluation Policies Under the New Statute</td>
<td>221</td>
</tr>
<tr>
<td>Florida Litigation in Response to the Statute and Policies</td>
<td>224</td>
</tr>
<tr>
<td>Changes to Florida State Statutes in Response to the Litigation</td>
<td>229</td>
</tr>
<tr>
<td>Florida Education Association Response to Revised Legislation</td>
<td>230</td>
</tr>
<tr>
<td>Court Grants Defendants’ Motion to Dismiss in Part</td>
<td>231</td>
</tr>
<tr>
<td>Court Order on Cross-Motions for Summary Judgment</td>
<td>236</td>
</tr>
<tr>
<td>Further Changes to Legislation</td>
<td>239</td>
</tr>
<tr>
<td>3. LEARNING FROM HISTORY AND LESSONS FROM LAWSUITS</td>
<td>242</td>
</tr>
<tr>
<td>Accountability is Not Poised to Pull a Vanishing Act</td>
<td>243</td>
</tr>
<tr>
<td>Inexorable Geographic Expansion of Reform and the “Testing Machine”</td>
<td>244</td>
</tr>
<tr>
<td>Chapter</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Shifting Locus of Responsibility: Students to Schools to Teachers</td>
<td>250</td>
</tr>
<tr>
<td>Media Criticism: Fact of Life since Letterpress Days</td>
<td>252</td>
</tr>
<tr>
<td>Prussians, Russians, and 170 Years of “Inferior” U.S. Schools</td>
<td>253</td>
</tr>
<tr>
<td>Private-Sector Involvement in Education: Business Has Never Minded Its Own Business</td>
<td>255</td>
</tr>
<tr>
<td>Causal Factors of Education Reform: It Is the Economy</td>
<td>259</td>
</tr>
<tr>
<td>College &amp; Career Readiness</td>
<td>262</td>
</tr>
<tr>
<td>Poor (and Immigrants, and Differing Abilities) Have Always Been Among Us</td>
<td>264</td>
</tr>
<tr>
<td>Litigation Slowed—But Did Not Stop—Race to the Top</td>
<td>266</td>
</tr>
<tr>
<td>Type of Test Matters</td>
<td>268</td>
</tr>
<tr>
<td>Validity of Value Added is in the Eye of the Beholder</td>
<td>275</td>
</tr>
<tr>
<td>Teachers Are More Than a Score: High Percentages May Mean Litigation</td>
<td>277</td>
</tr>
<tr>
<td>4. KEEPING THE MONKEY OFF A DISTRICT’ S BACK</td>
<td>320</td>
</tr>
<tr>
<td>Follow the Four Key Factors that Trigger Increased Accountability</td>
<td>321</td>
</tr>
<tr>
<td>Select the Right Assessments: Tailored, Timely, and Transparent</td>
<td>325</td>
</tr>
<tr>
<td>When Designing a Growth Model, Perceived Validity Trumps Statistical Validity</td>
<td>330</td>
</tr>
<tr>
<td>Take Care When Attributing Scores to Staff</td>
<td>331</td>
</tr>
<tr>
<td>Avoid Punitive Percentages if Possible</td>
<td>334</td>
</tr>
<tr>
<td>Areas for Further Study</td>
<td>336</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Changes in assessment standards.</td>
<td>250</td>
</tr>
</tbody>
</table>
CHAPTER 1

STANDARDIZED TESTING IN THE SPOTLIGHT

“Here comes the monkey!” declared television host John Oliver on his talk show Last Week Tonight. Oliver referred not to the latest comedy act, animal-themed cartoon, or misbehaving celebrity, but rather to his show’s parody of a mascot used by a Texas public school in a video. The video was created by the school to “get kids in the mood” for a standardized test, the scores from which are used as a part of teacher evaluation under the Obama administration’s Race to the Top (RTTT) initiative. The red monkey mascot shimmied his way across the soundstage wearing a t-shirt emblazoned with “I Love Testing.”

After the applause and laughter quieted, Oliver proceeded to describe to his 4.1 million viewers and millions of others who would eventually watch the video online the recent trends in standardized testing and the use of scores testing generates as part of teacher evaluations. He dedicated eighteen minutes of his hour-long show to current standardized tests such as

---


Partnership for Readiness for College and Careers (PARCC), addressed the role of the Pearson Corporation, and referenced the value added formulas used for teacher evaluation in early-adopting Race to the Top states such as Florida.\(^1\)

In addition to reviewing scholarly research, an astute school or district administrator would be wise to keep apprised of trends in education by monitoring popular media.\(^2\) When a comedy-oriented talk show with a large viewership dedicates a third of its airtime to canvassing a subject, it behooves district and school administrators to take heed. Something significant was obviously taking place in the worlds of standardized testing, data, and teacher evaluation, and John Oliver’s show was but one of many indicators.

According to research on the process used by editorial boards to select leading news stories, one can either believe the assertion of some journalists that “news reflects reality, pure and simple,”\(^3\) or that of sociologists that the social context, including a “preference for conflict or drama,”\(^4\) drives the lead stories in the news and other media (i.e., “if it bleeds, it leads”). Either way, recent popular media reflects the assertion that big data are “no longer the province of a few giant companies,”\(^5\) and the field of education has followed suit. “Quantitative data,”\(^6\) including


scores generated by standardized tests, have “gained enormous influence in education systems.”7

This became more readily apparent than ever after President Obama and his Secretary of Education, Arne Duncan, announced their first major education initiative in 2010. Race to the Top (RTTT), a competitive federal grant, required states to adopt rigorous standards, administer rigorous annual assessments, and use scores from the tests as a part of teacher evaluations.8

Articles about standardized testing and teacher evaluation, once typically the subject of back-page newspaper stories or education publications perused only by practitioners, were beginning to make their way to the front pages of major newspapers and news websites. The PARCC test, funded by the same RTTT initiative requiring states to use standardized test scores as part of teacher evaluation, has been mentioned in the front section—and even the front page—of major newspapers on several occasions.

One such article in the Chicago Tribune described Obama Secretary of Education Arne Duncan traversing a phalanx of sign-wielding protestors as he entered a school.9 Many of the protestors were affiliated with More Than a Score, a parent lobby that has not only established a

---


website, but even gone so far as to post printable links to a “Universal Opt-Out Letter”\textsuperscript{10} with instructions parents have been known to submit to schools, believing it is an official document.

Teacher evaluation, another topic previously only rarely discussed outside of education journals, also began to appear in major mainstream media stories after the first twelve states were awarded RTTT grants and scrambled to meet its requirements. New York, rejected in the first round, hastily introduced a raft of legislation to qualify for Phase Two of Race to the Top. This process was closely followed by the \textit{New York Times}.\textsuperscript{11} Plans in New York, as reported in the \textit{Times}, included the use of standardized test data—primarily that derived from state exams—as a significant percentage of teachers’ evaluation ratings. However, as noted in the same article, the plan impacted many teachers of “subjects or grades that do not have an annual state assessment.”\textsuperscript{12} State officials were quoted as saying that existing state tests “may have become too easy and predictable,”\textsuperscript{13} and yet “were not useless,”\textsuperscript{14} personifying the rushed atmosphere that pervaded the New York legislative process in response to RTTT.

“We believe that if done correctly this will change the landscape dramatically. This is not a gotcha system. This is about creating professional development that can really improve


education,”

state education commissioner David Steiner told the Times. The article clarified that while the legislation would not affect teacher compensation, “it could make it easier for schools to fire teachers deemed subpar.” A union official also offered his opinion to the reporter. “The concept of this has never been unacceptable,” claimed state union chief Richard Ianuzzi, “but doing it unilaterally or making evaluations solely dependent on students’ test scores were not options.” In the words of Commissioner Steiner when interviewed by the Times, “it remained unclear if the state was out of ‘choppy waters.’”

Florida also qualified for Phase Two of Race to the Top funds. The Orlando Sentinel ran a lengthy article on the new value-added model adopted across the state of Florida in response to RTTT, the formula for which talk show host John Oliver claimed on his show “looks like the kind of thing that aliens carve into an anti-Semite’s corn field.” The value-added formula, explained the Sentinel article, would analyze two years of scores from the Florida Comprehensive Assessment Test (FCAT) and measure a teacher’s impact on students’ growth by comparing the two scores while accounting for variables outside of the school’s control.

---


“It’s the engine for instructional improvement in the schools,” the article quoted Florida Department of Education staffer Kathy Hebda as proclaiming, “This is not their grandmother’s evaluation system.” The article quoted kindergarten teacher Grace Cabrera as saying, “It’s very complicated.” Regarding the inclusion of test scores in teacher evaluations, adding, “and I think that’s going to be very stressful for teachers.” Andrew Spar, president of the Volusia Teachers Organization, went a step further when interviewed by the *Sentinel*, declaring, “This is high-stakes now. I truly believe there will be a ton of lawsuits that come out of this.”

While litigation is not typically measured by units of weight such as the ton, popular media revealed that a number of lawsuits have been filed in early-adopting Race to the Top states, particularly as applied to the use of student achievement data as part of teacher evaluations. In a memorandum prepared for superintendents and principals in her state, Julie M. Slavens, Staff Attorney for the Indiana School Boards Association, noted recent legislation surrounding teacher evaluation in five states, including four early-adopting Race to the Top

---


states: Florida, New York, Tennessee, and Texas.\textsuperscript{26} This would seem to indicate that, much like the State of New York as it embarked on the implementation of Race to the Top, late-adopting states and their school officials may find themselves in “choppy waters”\textsuperscript{27} as they attempt to incorporate student achievement data into teacher evaluations.

**Problem Statement**

Eighteen of the fifty United States, along with the District of Columbia, accepted Race to the Top (RTTT) funds\textsuperscript{28} during Barack Obama’s presidential administration. Any state that accepted RTTT monies was required to apply student growth as measured by standardized tests to the teacher evaluation process. Aside from the mandate itself, little guidance was provided to states, and within many states, to school districts. In some cases, states or school districts must select assessments, designate or design a growth model, assign student scores to particular teachers or groups of teachers, and determine the extent to which test scores will impact teachers’ final evaluation ratings.

Since the first RTTT awards were announced in 2010, a number of lawsuits have been filed in early-adopting RTTT states, particularly as applied to the use of student achievement data as part of teacher evaluations required by the grant. In a memorandum prepared for superintendents and principals in her state, Julie M. Slavens, Staff Attorney for the Indiana


School Boards Association, noted recent legislation surrounding teacher evaluation in five states, including four early-adopting Race to the Top states: Florida, New York, Tennessee, and Texas. In light of these early filings, late-adopting states and their school officials may find themselves vulnerable to legal complaints as they follow the RTTT mandate to incorporate student achievement data into teacher evaluations.

Research Questions

As public school districts in Illinois and other late-adopting Race to the Top states wrestle with implementation of the student achievement component of teacher evaluations, it will help school leaders to find answers to the question, “What are best practices for incorporating student achievement and growth data into teacher classroom performance evaluation process, based upon litigation and legislation in early-adopting Race to the Top states?” Specific sub-questions pertaining to the general category of best practices include:

1. Should school districts assume that the use of standardized testing and student growth data are the unprecedented brainchild of a single administration and, therefore, likely to “go away,” or should they collaborate with principals and teachers to carefully plan the type of tests to be used and the attribution of scores to particular teachers?

2. Is it better to employ a relatively simple student academic growth model that may not be equitable but is easily understood by educators, or should school officials employ psychometricians to design a complex formula that takes into account some of the

---

many variables not in the classroom teacher’s direct control, even if this results in the
formula being incomprehensible to most staff?

3. To what extent should student growth data impact a teacher’s overall evaluation?

Procedures

Standard legal research methodology was applied to the analysis of litigation filed in
response to the use of student growth data for teacher evaluations in select early-adopting Race
to the Top states. The aforementioned litigation contains references to newly-minted legislation.
Ergo, historical state legislation and state legislation passed in response to Race to the Top were
also analyzed. Again, because this is a new legal issue lacking the rich history of, for example,
first amendment cases, a review of any cases cited by judges in the Race to the Top lawsuits will
also help inform analysis of the litigation. Lastly, to provide context to this newly-emerging area
of the law for which scant law review articles are currently available, the history of education
reform in the United States, particularly as pertains to standardized testing and accountability,
was analyzed in the preface to the legal analysis. The elucidation of history includes information
gleaned from original source documents whenever possible.

Significance of the Study

Following litigation can be instructive for leaders, as it makes an excellent cautionary
tale. A Google Internet search for “learn from lawsuits” returns 19,300 hits in less than a second.
However, while the linked material includes articles and posts about the police, medical
profession, and regulatory reform pertaining to hair braiders, there is no mention of education
until a solitary article—about bullying litigation—linked to the third page of hits. The current academic literature contains a plethora of articles pertaining to teacher evaluation in general. However, there is a dearth of literature—especially in the form of law reviews—pertaining to litigation in response to the use of student growth data as part of the evaluation process.

Amid this informational vacuum, seven Race to the Top Phase-Three states are beginning to incorporate test scores into teacher evaluations. Some of them, such as Illinois, grant local districts and even schools significant leeway in determining how student growth data is applied to teacher evaluations. While the recently-passed Every Student Succeeds Act does not require the use of student achievement data as part of teacher evaluations, it does not forbid it. In other words, at least eighteen states and Washington, D.C. operate under state statutes enacted in response to RTTT that could remain in place for years to come.

It is crucial that district and school administrators in these states understand and learn from trends observed in litigation filed in early-adopting RTTT states. This study is significant in that it is a multi-state study, limited not by geography or type of court, but rather by a focus on the use of test scores as part of teacher evaluations. The geographical breadth of coverage


applied to a relatively narrow topic affords the opportunity for the identification of more patterns than a narrower single-state or single-court level analysis.

This study will provide a sense of the history of standardized testing and school reform, the types of assessment most likely to trigger litigation, the relative value of validity versus simplicity in growth modeling, pitfalls to avoid when attributing students’ growth to a specific teacher, and the tipping point for weighting the impact of achievement data in calculating a final teacher evaluation rating.

Delimitations

This study is intended to analyze litigation in early-adopter Race to the Top states that elucidate unique problems pertaining to the use of student growth data as part of teacher evaluation. While it would be interesting to include litigation filed in other early-adopter Race to the Top states, such as Texas and Tennessee, at this point those cases do not appear to include any issues not already raised by the New York and Florida litigation. Ergo, the legal analysis considers only the latter two states. The roots of education reform and standardized testing in the United States—knowledge of which informs the analysis of the aforementioned litigation—reach back as far in history as 1837. Therefore, the historical overview is narrowly tailored to consider only the events most closely related to standardized testing and accountability.
Limitations

Only twelve states qualified to receive the initial two phases of Race to the Top funds. Remaining states typically received a waiver from the sanctions of No Child Left Behind and were given until the 2016-2017 school year to implement the use of student growth data as part of teacher evaluations. The 2016-2017 teacher evaluations—and any resulting litigation—in those states are still yet to be filed as of this writing. Also, the impact of Every Student Succeeds Act (ESSA), which lifted the requirement for but stopped short of forbidding the use of test scores in teacher evaluations, is not yet known as of this writing. Therefore, this dissertation includes the only early-adopting RTTT states of New York and Florida.

35 Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 177 (Cornell University Press 2012).

36 An Education in Politics: The Origins and Evolution of No Child Left Behind 181 (Cornell University Press 2012).


CHAPTER 2
TWO SUITS AND A SESQUICENTURY OF REFORM

After Horace Mann’s 1837 appointment to the Massachusetts State Board of Education, he needed evidence to support his education reform ideas, such as emphasizing the real world and relying less on rote recitation.\(^1\) At the turn of the 20\(^{th}\) Century, John Franklin Bobbitt became interested in the Industrial Revolution’s scientific efficiency studies and their potential application to education.\(^2\) In 1950, a Topeka, Kansas third grader, Linda Brown, tired of the long, hazardous commute to her segregated school, and her father attempted to enroll her in an all-white public school located just blocks from their home.\(^3\) Seven years later, the Soviet Union launched a metal contraption the size of a beach ball into space.\(^4\) The impact of these events eventually converged, bringing both political parties to a rare 21\(^{st}\) Century bipartisan consensus.

When introducing Race to the Top, Secretary of Education Arne Duncan made reference to a “perfect storm for reform”\(^5\) and prefaced his remarks with the phrase, “for the first time in

---


However, while some may view Race to the Top as a manifestation of the postmodern human capital philosophy, the initiative’s roots inhabit a 170 year history of standardized testing and school reform. This history can be traced through a standoff between a politician and his local schoolmasters, the impact of immigration and urbanization prior to the Civil War, the Industrial Revolution, the Cold War, the Civil Rights Movement, the dawn of the Information Age, and the onset of the 21st Century.

Common School Movement

The U.S. media sometimes claims that unfavorable comparisons to Europe and a demand for accountability through a standardized written test originated in the 1980s, with the *Nation at Risk Report*. Others cite the Elementary and Secondary Education Act (ESEA) or Sputnik as the proximal origin. In fact, Horace Mann, who implemented America’s first written standardized test during his tenure as the Secretary of the Massachusetts State Board of Education in the mid-nineteenth century envisioned public education as “the great equalizer of the conditions of men, the balance wheel of the social machinery.” This was an apt articulation of the Common School Movement’s primary goal, espousing for the first time the idea of free and universal public education.

---


Eventually realizing the potential of the aforementioned ideal, the Common School Movement originated in Massachusetts and spread across the nation. It is considered by many to be the nexus between the first American one-room schoolhouses and today’s comprehensive system of free public education.\(^5\) Typical of educational reform movements, lying beneath the Common School Movement’s lofty rhetoric and arguably beneficial outcomes lurked other motives related to political power and the economy. Unlike many future reform eras, however, the Common School Movement did not rise from the ashes of a floundering economy, but rather it found its impetus during a time when cities were expanding rapidly due to the emergence of a U.S. market economy.\(^6\)

In the mid-nineteenth century, the Whig Party was concerned the individualistic emphasis on gain fostered by market forces would undermine the sense of moral, social, and civic duty underpinning the American republic. Nevertheless the Whig Party did not want to stifle the economic growth generated by the market.\(^7\) Whigs considered government involvement to be essential to both the social good and the market economy.\(^8\) To further the latter, the Whigs supported government funding and oversight of infrastructure such as canals and turnpikes.\(^9\) To

---

5 David F. Labraree, Someone Has to Fail: The Zero-Sum game of Public Schooling 4 (Harvard University Press 2012).


7 David F. Labraree, Someone Has to Fail: The Zero-Sum game of Public Schooling 59 (Harvard University Press 2012).

8 William J. Reese, Testing Wars in the Public Schools: A Forgotten History 51 (Harvard University Press 2013).

9 David F. Labraree, Someone Has to Fail: The Zero-Sum game of Public Schooling 58 (Harvard University Press 2012).
advance the former, the Whig Party also advocated for a number of social solutions, including free public schools.  

Horace Mann, a staunch proponent of Whig ideals long before assuming his position on the Massachusetts Board of Education, argued education not only counteracted the class warfare emerging as the market economy prospered, but it also served to “disarm the poor of their hostility toward the rich; it prevents them being poor.” Mann and his Whig contemporaries in Massachusetts were proponents of school reforms such as promoting the proliferation of scholarly education journals, providing government funding for universal education, hiring of more female teachers, using statistics for evaluation and planning, improving the teacher hiring process as well as the training teachers received at Normal schools, hiring of principals and superintendents, and using written assessments to both select teachers and measure student learning.

The Whigs had opponents, not only in the Democratic Party that at the time as a whole abhorred centralized government control, but also more locally in the Massachusetts public schools. The Boston grammar school head teachers, known collectively as the Grammar Masters, had resisted reform for years. Hailing from prominent families and typically the

---


13 David F. Labraree, Someone Has to Fail: The Zero-Sum game of Public Schooling 58 (Harvard University Press 2012).

products of an Ivy League college education, the Grammar Masters used their connections on pre-Horace Mann School Committees to do as they wished with impunity.\(^{15}\)

Not only did the Grammar Masters persist in a pedagogy based almost entirely on lock-step adherence to textbooks, harsh discipline including corporal punishment, and short-term rote memorization,\(^{16}\) they also supplemented their already-generous salaries by requiring the use of textbooks they had written, selling supplies and materials to students, and even charging for extra tutoring conducted on school grounds in close proximity to school hours. In spite of this conduct, they were highly respected in the community. Accustomed to considerable approbation,\(^{17}\) the Grammar Masters did not readily concede to the Common School movement’s reforms.\(^{18}\)

Shortly after assuming his position, Horace Mann toured the Massachusetts schools. As a result of these tours Mann became concerned about the lack of accountability for schools, particularly schools managed by the Grammar Masters. He was especially appalled at the system of periodic exhibitions schools held for board members and the community. The exhibitions consisted primarily of students reciting from rote and singing for an audience. In addition to the absence of focus on higher-level thinking skills, the rigor of these exhibitions was not equitable. For example, Mann described one exhibition as the “harvest home of knowledge

\(^{15}\) William J. Reese, Testing Wars in the Public Schools: A Forgotten History 41 (Harvard University Press 2013).


\(^{17}\) William J. Reese, Testing Wars in the Public Schools: A Forgotten History 41 (Harvard University Press 2013).

and virtue,"\textsuperscript{19} but another as “mechanical movements got up as a mock representation of knowledge.”\textsuperscript{20} These observations led Mann to believe his vision of education as an equalizer was not being fulfilled simply by providing universal education.

In his First Annual Report to the Board in 1838, Mann identified two proximate causes of this inequity: parent involvement and teacher quality.\textsuperscript{21} By the Seventh Annual Report to the Board in 1844, Horace Mann had preceded the Trends in International Mathematics and Science Study (TIMSS) by more than 150 years in making a negative comparison between U.S. schools and schools in other nations—in this case, European countries. It is worth noting Mann also visited what were then known as “insane asylums” during his time in Europe, but found “none superior”\textsuperscript{22} to those in the United States.

Mann made his European tour during his honeymoon, accompanied not only by his wife but also by fellow Whig reformer Samuel G. Howe and Howe’s spouse, Julia Ward Howe. The latter recalled how during their time in Europe, Mann immersed himself in scholarly research on teaching and assessment while visiting countless schools. While this may not be a typical honeymoon, it served to firmly cement many of the ideals undergirding Mann’s implementation of Common School reforms.\textsuperscript{23} Citing the European—and particularly Prussian—emphasis on student engagement, real-world application, and higher-level thinking rather than rote

\textsuperscript{19} First Annual Report of the Board of Education, Massachusetts, U.S. 83 (Boston: Dutton and Wentworth 1838).


\textsuperscript{23} William J. Reese, Testing Wars in the Public Schools: A Forgotten History 57 (Harvard University Press 2013).
memorization,\textsuperscript{24} Mann made the “radical” assertion these foci should be among the priorities of schools both in Massachusetts and across the United States.\textsuperscript{25}

By contrast, in the same report, Mann addressed the pedagogy he had observed in the Boston Grammar Masters’ schools. Mann’s concerns included over dependence on textbooks, requiring students to memorize the technical vocabulary and definitions at the beginning of each chapter before they had “any practical idea of their meaning.”\textsuperscript{26} Generally teachers read the text to the students without contributing any additional content knowledge, let alone relating it “to other kindred subjects or the actual business of men and the affairs of life.”\textsuperscript{27} At the conclusion of these lessons, the students recited the definitions and questions from the text “from memory with suspicious fluency.”\textsuperscript{28} If the pupils were asked to connect or apply their knowledge to a real-life situation, they either sat in a silent stupor or replied with “some ridiculous answer, which at once disparages science and gratifies the ill-humor of some ignorant satirist.”\textsuperscript{29}

By contrast, the real-world application and higher-level thinking cited by Mann were evident when observing both instruction and assessment during his visits to Prussian schools. In


contrast to teaching lists of domain-specific vocabulary by rote before teaching a topic, the Prussian teacher, according to Mann, noted that the process “cumbers and darkens the subject with no technical phraseology.” Rather than mechanically following a text year after year, European teachers observed their students’ proficiency levels before beginning instruction and adjusted their teaching practices accordingly.

When Prussian teachers presented ideas, they did so in a way that caused students to ask questions and examine their previously held perceptions. Such was the extent of the connection of new knowledge to real life, hyperbolized Mann, that “should the most ignorant man or most destitute vagrant in society” ask a European school master if knowledge was worth attaining, the teacher could prove its importance to him. As opposed to their stultified American counterparts, Mann described how Prussian students were engaged to the point of being “delighted.”

Mann also found in his examinations of European student learning a greater emphasis on higher-level thinking and real-world application. As opposed to American schools where students were simply required to memorize measurement tables, Mann observed European students making conversions and simulating currency exchanges. During an electro-magnetism

---


exam, students demonstrated experiments on an apparatus. Another idea Mann borrowed from Europe for his testing plan was the idea of written examinations. The use of a standard set of written questions administered in a timed setting provided both uniformity and the opportunity for each student to share the full extent of his thinking in a shorter amount of time rather than dedicating a longer period of time to individual oral responses.

Ironically, while standardized testing is often associated with a traditional mindset, Horace Mann used mandatory exams to garner support for what were then considered radically progressive ideas about student engagement and real-life application of learning. Referring to the traditional system of oral exhibitions, a 19th Century scholarly journal of education defended the move to a standard written test, positing it was unfair to make any comparison of schools if students were assessed orally using questions created by different local school committees.

While the Grammar Masters continued to publically criticize and even ridicule Mann’s reforms and interest in statistics, his committees began work behind closed doors in January 1845, drafting standardized written exams to be administered in June. It is important to note that while the 19th Century Massachusetts tests were groundbreaking in that they were mandatory,

---


written, and uniform in terms of both directions and grading, they were not comparable to the
norm-referenced instruments currently mandated in U.S. public schools.\textsuperscript{39}

Mann formatted standardized test scores using tables and charts. This resulted in a
number of seismic shifts. For example, the position of the Grammar Masters—who at one time
felt sufficiently empowered to publish official retorts to Mann’s reports—was weakened, in part
because of the enthusiastic response of the news media to the salacious statistics.\textsuperscript{40} Some of the
Grammar Masters and their ushers (teachers) even found themselves dismissed from
employment.\textsuperscript{41} Other parallels to Race to the Top can be found in source documents emerging
from Mann’s efforts in the ensuing years. Mann’s Tenth Report referred to a comprehensive
assessment system that included a growth model of sorts, using test scores and observations to
determine “whether a teacher has carried the children successfully forward.”\textsuperscript{42} Mann
foreshadowed the Elementary and Secondary Education Act when he asserted that education
could resolve the social ills generally prevalent among poor children.\textsuperscript{43} However, a Common
School publication echoed the rationalization of many 21st Century school leaders when it
attributed a year of negative score trends in part to an increasing number of “foreign children
who labor under many disadvantages.”\textsuperscript{44} A scholarly journal hinted at a future practice of

\textsuperscript{39} U.S Congress, Office of Technology Assessment: Testing in American Schools: Asking the Right Questions,
http://govinfo.library.unt.edu/ota/Ota_1/DATA/1992/9236.PDF.

\textsuperscript{40} William J. Reese, Testing Wars in the Public Schools: A Forgotten History 130 (Harvard University Press 2013).

\textsuperscript{41} William J. Reese, Testing Wars in the Public Schools: A Forgotten History 140 (Harvard University Press 2013).

\textsuperscript{42} Report of an Educational Tour, Being Part of the Tenth Annual Report of Horace Mann Esq. to the Board of
Education, Massachusetts, U.S. 122 (Boston: Dutton and Wentworth 1849).


\textsuperscript{44} William B. Fowle, 13 The Common School Journal 243 (1851).
Professional Learning Communities when it recommended pre-testing high school students to save time previously spent reviewing what students already knew.\textsuperscript{45} Of course, none of these portentous ideas emerged immediately into the spotlight, but one thing was certain: after Mann victoriously fought the educational reform battle,\textsuperscript{46} it became “impossible to return fully to a time when impressions alone measured a school’s worth.”\textsuperscript{47}

Scientific Management Era

The urbanization and economic shifts propelling the Common School Movement continued throughout the 19\textsuperscript{th} and early 20\textsuperscript{th} centuries, as “the master craftsman running his own shop at the start of the century gave way to the proprietor running his own factory in the middle of the [19\textsuperscript{th}] century, which in turn gave way to the corporation managing a series of factories at the end of the century.”\textsuperscript{48} Commercialization meant the entry-level positions, once offering apprenticeships and advancement to uneducated individuals, rapidly dwindled.\textsuperscript{49} This in turn severely limited opportunities for students who dropped out of school prior to completing their education, a common practice for children as young as eleven years of age prior to the 1890s.\textsuperscript{50}


\textsuperscript{46} Carl F. Kaestle, Pillars of the Republic: Common Schools and American Society 113 (Erica Foner, Hill & Wang 1983).

\textsuperscript{47} William J. Reese, Testing Wars in the Public Schools: A Forgotten History 129 (Harvard University Press 2013).

\textsuperscript{48} David F. Labraree, Someone Has to Fail: The Zero-Sum game of Public Schooling 87 (Harvard University Press 2012).

\textsuperscript{49} David F. Labraree, Someone Has to Fail: The Zero-Sum game of Public Schooling 87 (Harvard University Press 2012).

The hardships industrialization visited upon the uneducated were exacerbated by economic conditions. Rather than enjoying the relatively unfettered growth accompanying the Common Schools Movement, the era between 1870 and the end of the Roaring 1920s was subject to economic panic. The Panic of 1873 began with the failure of the Jay Cooke bank and persisted until the 1880s. After just over a decade of relative prosperity, the overbuilding of railroads led to the Panic of 1893, the effects of which were felt until 1900. This tumult meant public education was now expected to “construct a new social order for the corporate industrial age,” establishing “a precedent for future reformers to call on schools to take on an ever-enlarging array of social problems to solve with new school programs.” Another difference between the Common School Era and the Scientific Efficiency Movement was the latter had its genesis in a different region of the United States. While the heart of the Common School Movement was on the east coast, many of the most prominent education reform publications from the Commercial Industrial era were produced in United States regions farther to the west.

The Corporate Industrial Era shared one trait with the time of the Common School Movement: an increasing population diversity of both the United States as a whole and its public school classrooms. As a result of the market revolution fueling the Common School movement, the first wave of immigrants arrived from England, Germany, and Ireland to provide labor for the emerging market economy. A second larger wave saw millions immigrate from Southern and

---

51 David F. Labraree, Someone Has to Fail: The Zero-Sum game of Public Schooling 88 (Harvard University Press 2012).

52 David F. Labraree, Someone Has to Fail: The Zero-Sum game of Public Schooling 103 (Harvard University Press 2012).

53 David F. Labraree, Someone Has to Fail: The Zero-Sum game of Public Schooling 103 (Harvard University Press 2012).
Eastern Europe from 1890 until federal law limited immigration in 1924. Both groups left their homelands in search of employment available in the U.S. mines, mills, and factories.\textsuperscript{54}

The waves of immigrants, coupled with the increased demand for secondary education, led to exponentially increasing American public school enrollment. According to the United States Census, the public school population increased from 6.8 million in 1870 to 15.5 million in 1900.\textsuperscript{55} This enrollment uptick was particularly marked at the secondary level. At the turn of the 20\textsuperscript{th} Century, only about 10 percent of the 14-17 age group was enrolled in high school. By 1930, more than 50 percent of that age group attended at least one or two years of high school.\textsuperscript{56} Not only did the public school population increase, but it also diversified. School districts wrestled both with integrating students from varying linguistic, cultural, and educational backgrounds into instructional programs and with ascertaining how to measure their progress. While public school enrollments more than doubled, the general population also increased, but at a far slower pace. Taxpayers multiplying at a lower rate than students served to exacerbate the challenges faced by the public education system, and cost control became a high priority for local school boards.\textsuperscript{57}

Ergo, industry and education collided to form the Scientific Efficiency Movement in education reform. The Era of Efficiency first gained momentum in the business sector with

\textsuperscript{54} David F. Labraree, Someone Has to Fail: The Zero-Sum game of Public Schooling 87 (Harvard University Press 2012).


\textsuperscript{56} Daniel P. Resnick, Competency Testing Historically Considered, 8 Review of Research in Education 8 (1980).

Frederick Taylor and his seminal work *Principles of Scientific Management*. Originally written for presentation to the American Society of Mechanical Engineers, an organization of which Taylor served as president, it was eventually published by Harper & Brothers for a wider audience. Taylor’s purpose was to show the losses the United States was suffering due to inefficiency, to propose scientific management as the solution to the inefficiency, and “to prove that the best management is a true science, based upon clearly defined laws, rules, and principles as a foundation.” In the introduction to *Principles*, Taylor presciently posited, “our duty, as well as our opportunity, lies in systematically cooperating to make and to train” competent men rather than “hunting for a man whom someone else has trained” and that “the fundamental principles of scientific management are applicable to all kinds of human activities”

A number of educational reformers, apparently agreeing with Taylor’s assertions, became advocates for the integration of scientific management concepts into education. A year after Taylor published *Principles of Scientific Management*, John Franklin Bobbitt, a professor of education administration at the University of Chicago, published an article in *The Elementary School Teacher* entitled “The Elimination of Waste in Education.” Bobbitt’s article was based

64 Frederick P. Taylor, *Principles of Scientific Management* 7 (Harper & Brothers 1911).
on a case study of education reform in Gary, Indiana. After praising the construction of steel mills on what had been “a region of waste sand-dunes,” Bobbitt described how the rapid growth of the city, including an influx of “immigrant foreign laborers possessing but little taxable property,” caused “a financial problem of peculiar difficulty” for the school district. Gary, Bobbitt wrote, was faced with two choices: either build “inferior buildings” and “employ cheap teachers” to teach classes of large size or create a “thoroughly modern school plant” and operate it “according to recently developed principles of scientific management.”

In parlance reminiscent of the manufacturing sector, Bobbitt described a leader in a position entitled “educational engineer” faced with the task of “operating his plant during school hours at 100 percent efficiency.” This involved, for example, constructing only four regular classrooms for eight primary-level students. To this end, classes were scheduled on a program matrix denoting subject areas, classes of students, and locations. While one group of students was engaged in activities such as nature studies, drawing, music, manual activities, or

play, the other group used the regular classroom.\textsuperscript{77} The Gary school district was also beginning to explore what is currently known as extended day and year-round school programming to avert the counter-productive phenomenon, the superintendent referred to as the “street and alley time”\textsuperscript{78} spent by students no longer bound by household chores as children had been in the past century.\textsuperscript{79}

Scientific efficiencies described in Bobbitt’s Gary case study also extended to staff by using a combination of general educators and specialists assigned to teach students via a scheduling matrix designed to eliminate the co-teaching model prevalent at the time for specialized subjects.\textsuperscript{80} This type of scheduling not only reduced staffing costs, but also allowed teachers to complete all of their work during the school day. This was important to Bobbitt because he believed “the highest working efficiency demands certain qualities of personality that are not to be had without normal association with one’s fellows nor without proper and normal leisure activities.”\textsuperscript{81}

Bobbitt felt the presence of specialized staff and rotating schedules in Gary’s 20\textsuperscript{th} Century schools also benefitted students. To avoid the inefficiency of falling behind, students could be scheduled to receive instruction twice a day in their weakest subjects by forgoing special activities until mastery in the deficient subject area was achieved. Upon achieving

\textsuperscript{77} John Franklin Bobbitt, \textit{The Elimination of Waste in Education}, 12 The Elementary School Teacher 261 (1912).

\textsuperscript{78} John Franklin Bobbitt, \textit{The Elimination of Waste in Education}, 12 The Elementary School Teacher 263 (1912).

\textsuperscript{79} John Franklin Bobbitt, \textit{The Elimination of Waste in Education}, 12 The Elementary School Teacher 263 (1912).

\textsuperscript{80} John Franklin Bobbitt, \textit{The Elimination of Waste in Education}, 12 The Elementary School Teacher 264 (1912).

\textsuperscript{81} John Franklin Bobbitt, \textit{The Elimination of Waste in Education}, 12 The Elementary School Teacher 265 (1912).
mastery, the special activities were scheduled to allow students a seamless re-entry.\textsuperscript{82} The scheduling matrix also included study periods intended for teachers to provide “individual attention to laggards.”\textsuperscript{83} When describing these interventions, Bobbitt cited Leonard Porter Ayers’ work,\textsuperscript{84} who rather than decrying the moral ramifications of student failures instead focused on their inefficiency.\textsuperscript{85}

In a landmark study for the Russell Sage Foundation in New York, Ayers found of the students who were still in school by the eighth grade, more than half lagged behind their age group.\textsuperscript{86} Ayers clarified that while reducing “retardation,”\textsuperscript{87} or retention in the same grade level, would result in some financial savings, the primary motive for addressing this issue was the efficacy of the schools in educating students.\textsuperscript{88} Ayers sought to quantify factors endemic to students who were struggling in school. A bar graph included a summary of “retardation”\textsuperscript{89} by ethnicity, with German-Americans having the lowest rate at 16 percent and Italian-Americans the highest at 36 percent.\textsuperscript{90} In searching for a method for quantifying factors impacting student

\textsuperscript{82} John Franklin Bobbitt, \textit{The Elimination of Waste in Education}, 12 The Elementary School Teacher 267 (1912).

\textsuperscript{83} John Franklin Bobbitt, \textit{The Elimination of Waste in Education}, 12 The Elementary School Teacher 266 (1912).

\textsuperscript{84} John Franklin Bobbitt, \textit{The Elimination of Waste in Education}, 12 The Elementary School Teacher 266 (1912).


\textsuperscript{87} Leonard Porter Ayers, Laggards in Our Schools: A Study of Retardation and Elimination in City School Systems 98 (New York Charities Publication Committee 1909).


learning, the Sage Foundation Study cited the correlation between enlarged glands or adenoids and school failures.\textsuperscript{91} Among the remedies Ayers suggested for high retention rates were more thorough student medical examinations, courses tailored to student abilities, and “a better knowledge of the facts.”\textsuperscript{92}

Later, Ayers added to his body of facts when he led \textit{The Cleveland School Survey} under the auspices of The Cleveland Foundation. The results of \textit{The Cleveland School Survey} were initially published as fifteen monographs and eventually compiled into a 363 page summary volume.\textsuperscript{93} During the study, Cleveland was compared to other major metropolitan areas and found to be “average.”\textsuperscript{94} The study provided an interesting glimpse into schools and districts of the Efficiency Era.

Student retention—which remained a major emphasis in \textit{The Cleveland Study} as it was in Ayers’s earlier work—also served as a key data point in the chapter on measurement.\textsuperscript{95} Rather than merely decrying the failure rates as he had in the Sage Foundation paper, in \textit{The Cleveland Study}, Ayers attempted to identify some of the causes through the use of student achievement data. This “brought to hand abundant evidence on which to base a wholly impersonal view of the


\textsuperscript{93} Leonard Porter Ayers, \textit{The Cleveland School Survey: Summary Volume 19} (The Survey Committee of the Cleveland Foundation 1917).

\textsuperscript{94} Leonard Porter Ayers, \textit{The Cleveland School Survey: Summary Volume 159} (The Survey Committee of the Cleveland Foundation 1917).

\textsuperscript{95} Leonard Porter Ayers, \textit{The Cleveland School Survey: Summary Volume 19} (The Survey Committee of the Cleveland Foundation 1917).
classwork of the schools.”  

Upon analyzing Cleveland’s non-promotion data, Ayers dismissed the teacher-reported reason of “mental incapacity,” observing wryly there was “an astonishing increase of so-called mental incapacity through the intermediate grades showing that this diagnosis of the difficulty as offered by the school officials is not right.” Ayers and his team instead turned to student achievement data, including comparisons between Cleveland and other metropolitan school districts such as Los Angeles, Seattle, Detroit, and Milwaukee.

Not only did he compare Cleveland to other urban districts, but Ayers also compared schools within the system, in keeping with his belief that “a vigorous policy of comparative study should be adopted as a regular part of the routine administration of the system.” A twenty-first century researcher might smile ironically in response to Ayers’ description of the early twentieth century as an “ultra-scientific age.” However, the data included in *The Cleveland Study* was fairly sophisticated. Tests were administered and data analyzed in each grade to assess student mastery of several subjects, including silent and oral reading, arithmetic, and handwriting. There was even a growth model of sorts, with the study claiming “progress from grade to grade can be clearly defined. When the results are put together, they show that

---


there is a law of progress. Once the law is exhibited, it is possible to judge how far individual schools conform.”\textsuperscript{101}

Ayers acknowledged the limitations of school comparisons, due in part to the fact that “children in different schools differ one from another.”\textsuperscript{102} However, the study also analyzed factors impacting student achievement that were controlled by school officials. For example, a hand-drawn bubble chart comprised of a three-by-three matrix depicted the relationship between reading speed and the quality of “interpretation,”\textsuperscript{103} which in modern parlance would be known as comprehension. Each bubble was hand-drawn to scale and labeled with the percent of students demonstrating, for example, rapid speed and poor quality (of comprehension).\textsuperscript{104} While the study did not contain tables or graphs comparing individual teachers as it did school districts and individual schools, it identified the fourth grade as a level where reading growth declined in Cleveland.\textsuperscript{105} Ayers also posited “teachers of different degrees of efficiency are sure to be found in all parts of the system.”\textsuperscript{106}

The study also used the term “promotions for merit” as opposed to service and suggested “Cleveland cannot hope to have the best possible teaching force until adequate salaries are paid

\textsuperscript{101} Leonard Porter Ayers, The Cleveland School Survey: Summary Volume 150 (The Survey Committee of the Cleveland Foundation 1917).

\textsuperscript{102} Leonard Porter Ayers, The Cleveland School Survey: Summary Volume 151 (The Survey Committee of the Cleveland Foundation 1917).

\textsuperscript{103} Leonard Porter Ayers, The Cleveland School Survey: Summary Volume 178 (The Survey Committee of the Cleveland Foundation 1917).


and an adequate system of promotion based on evidence of growth and in professional ability is established. The present system is not calculated to enable the school officials sharply to draw the line between ordinary and extraordinary teachers.

Ayers cited the actions of one extraordinary teacher who worked efficiently to mitigate factors impacting student learning. This was a teacher who introduced special “Steamer Classes,” named for the steamships on which most immigrants of the era arrived in the United States. Steamer Classes—much like the English Learner programs of today—were designed to teach English to non-native speakers. Although applauding the teacher’s effort, Ayers noted Cleveland faced “a grave problem in the number of children from homes where the influence of European life is still strong.” Ayers urged teachers to learn more about the native cultures and languages of immigrant students to help them acclimate. In addition, he noted the need for reading materials to suit student needs. Ayers also advocated for a social studies course to explain the civic structure of Cleveland and the duties of its citizens to the community.

Another student population with differing instructional needs, “exceptional children,” included students who were then assigned labels such as “incorrigible” or “socially incompetent,” the latter including the subcategories of “insane,” “epileptic,” or “feebleminded.” The Binet-Simon achievement test was used to identify “exceptional” children to “select” the “feebleminded” and assign them to “classes for defectives.” Ayers’ colleague David Mitchell compared the administration of Binet-Simon to the sputum test used at the time to diagnose tuberculosis, due to the fact that it could be administered by personnel who were not psychologists so long as they were given a basic level of training.


Psychologist Albert Binet developed the first intelligence tests and scales\textsuperscript{124} in response to a request from the education ministry in France—where education had recently been made compulsory. These tests were designed to identify children who were most likely to require extra assistance during their early years in school.\textsuperscript{125} Along with fellow psychologist Theodore Simon, Binet decided to forgo anthropometrics, a then-popular method using body measurements as an indicator of intelligence. Instead Binet and Simon introduced psychometrics by developing a battery of trials of abstract thinking, vocabulary, problem solving, memory tasks, mental arithmetic, and moral dilemmas. Using test results on small groups of Parisian children, they created the concept of mental age, setting norms at tasks on which 75 percent of children at each age were able to perform.\textsuperscript{126}

In \textit{The Cleveland Study}, David Mitchell acknowledged: “it is unwise in the case of suspected mental subnormality to have only the Binet-Simon tests as the final resource in reaching a decision.”\textsuperscript{127} He posited the Binet tests “should not be used as the only available ultimate criteria for determining whether a child is an idiot, feebleminded, a moron, or normal.”\textsuperscript{128} In spite of these proposed limitations, \textit{The Cleveland Study} recommended use of

\textsuperscript{124} Anya Kamenetz, The Test: Why our Schools are Obsessed with Standardized Testing—but You Don’t Have to Be 44 (Public Affairs 2015).

\textsuperscript{125} Jim Horn & Denise Wilburn, The Mismeasure of Education 12 (Information Age Publishing 2013).

\textsuperscript{126} Anya Kamenetz, The Test: Why our Schools are Obsessed with Standardized Testing—but You Don’t Have to Be 45 (Public Affairs 2015).


Binet-Simon tests be expanded beyond the “selection”\(^\text{129}\) of the “feebleminded”\(^\text{130}\) and “undertake a systematic testing of all children who are making seriously slow progress or encountering unusual difficulties in their school.”\(^\text{131}\)

While Binet and Ayers may have had a relatively benign intent of using intelligence testing to identify students in need of support, others had different motives, ranging from the efficiency of scientific management to outright social Darwinism in the form of eugenics.\(^\text{132}\) Elwood Patterson Cubberley, a professor of education at Stanford University, not only directly trained many school administrators in the early twentieth century, but also he wrote *A Brief History of Education*, a textbook that served as a staple of many administrator training programs well into the 1950s.\(^\text{133}\) Cubberley stopped short of eugenics. If for no other reason than the sake of efficiency, he conceded the “seriously defective class of society,”\(^\text{134}\) he referred to as “unfortunates,”\(^\text{135}\) should be “cared for and educated as suitably as possible, for self-respect, self-support, and some form of social and vocational usefulness.”\(^\text{136}\)


\(^{134}\) Elwood Patterson Cubberley, *A Brief History of Education: Educational Practice and Progress Considered as a Phase of the Development and Spread of Western Civilization* 818 (Houghton Mifflin 1920).

\(^{135}\) Elwood Patterson Cubberley, *A Brief History of Education: Educational Practice and Progress Considered as a Phase of the Development and Spread of Western Civilization* 818 (Houghton Mifflin 1920).

As Cubberley instructed future school administrators in 1920, he worried not only about “what best to do with those of small intellectual capacity,”¹³⁷ but he was also concerned with “the selection and proper training of those with superior intelligence” and “the elimination of barriers to the advancement of children of large intellectual endowment.”¹³⁸ These statements echoed what he had written four years earlier in the introduction to his Stanford colleague Lewis M. Terman’s book *The Measurement of Intelligence*. Terman’s book served as a guide to the administration and interpretation of the updated Binet tests, renamed Stanford-Binet after being adjusted to fit American culture and to extend mental age into the adult years. At that time, Cubberly wrote intelligence tests could be used both for the “proper handling of subnormal on the one hand and gifted children on the other.”¹³⁹

At the time the United States entered World War I, Terman was collaborating with Robert Yerkes to develop intelligence tests for the military. The armed services were faced with the intake of phalanxes of new soldiers, among whom they had to quickly discern who was fit to serve as officers. Terman and Yerkes developed two assessments: the Alpha for English speakers and the Beta for non-speakers.¹⁴⁰ It is important to note that while the federal authorities began using intelligence tests at the same time as local schools, educational use was neither mandated nor monitored by the government. In addition to tracing the history of education and making

---


¹⁴⁰ Anya Kamenetz, *The Test: Why our Schools are Obsessed with Standardized Testing—but You Don’t Have to Be 49* (Public Affairs 2015).
recommendations for the future, Cubberley cited increasing state control over local schools in *A Brief History of Education*. However, the section of the text pertaining to the loss of local control was not linked to specific reforms, standardized testing, or the involvement of the federal government.\textsuperscript{141}

In his conclusion, Cubberley presciently noted education found itself with many problems to solve, including defining “proper educational relationships between the State and its subordinate units.”\textsuperscript{142} He believed “scientific organization of the educational process”\textsuperscript{143} was critical to resolving the issues he had raised. To this end, he predicted with unsettling accuracy that being able to analyze student test data “may be included, very soon, as part of the necessary pedagogical equipment of those who aspire to administrative positions”\textsuperscript{144} in schools. Others proposed more radical solutions to the challenges facing society in the early twentieth century.

Like Ayers and Cubberley, in the early pages of his book *Human Efficiency and Levels of Intelligence*, Henry Herbert Goddard wrote about identifying students with special needs using intelligence tests and providing them with training. Goddard, Director of the Bureau of Juvenile Research of Ohio, took the additional step of suggesting schools should form partnerships with local businesses to provide vocational training. Describing an example of students being taught

\textsuperscript{141} Elwood Patterson Cubberley, *A Brief History of Education: Educational Practice and Progress Considered as a Phase of the Development and Spread of Western Civilization* 687 (Houghton Mifflin 1920).

\textsuperscript{142} Elwood Patterson Cubberley, *A Brief History of Education: Educational Practice and Progress Considered as a Phase of the Development and Spread of Western Civilization*, viii, (Houghton Mifflin 1920).

\textsuperscript{143} Elwood Patterson Cubberley, *A Brief History of Education: Educational Practice and Progress Considered as a Phase of the Development and Spread of Western Civilization* 824 (Houghton Mifflin 1920).

to be assistant cooks, Goddard reported that “many of these morons find occupation as helpers in the bakery.”

Goddard revealed his more extreme beliefs, however, by writing about “social control of the unintelligent and inefficient.” He expected a broad consensus, as he posited that protecting society from those of lower intelligence was “an axiom which no one will attempt to deny.”

Goddard went on to describe a veritable eugenic utopia wherein all persons with low intelligence scores would be identified during childhood and institutionalized for life in colonies with others of similar cognitive abilities. This would entail “a big expense, but what compensation!” including saving the lives of “those who are killed by feeble-minded people” and removing children “who are the bane of the teacher in every class” so “your child and mine would not have to sit in school beside an imbecile.”

Goddard envisioned lifelong institutionalization of children with low intelligence scores would result in those with lower intelligence scores being “kept from propagating their kind” the ultimate goal of eugenics.

Reproduction was addressed by another eugenics proponent, Edward L. Thorndike, a professor of education at Columbia University. However, rather than discussing how to impede

---

146 Henry Herbert Goddard, Human Efficiency and Levels of Intelligence 116 (Princeton University Press 1920).
147 Henry Herbert Goddard, Human Efficiency and Levels of Intelligence 116 (Princeton University Press 1920).
148 Henry Herbert Goddard, Human Efficiency and Levels of Intelligence 120 (Princeton University Press 1920).
149 Henry Herbert Goddard, Human Efficiency and Levels of Intelligence 120 (Princeton University Press 1920).
150 Henry Herbert Goddard, Human Efficiency and Levels of Intelligence 121 (Princeton University Press 1920).
151 Henry Herbert Goddard, Human Efficiency and Levels of Intelligence 121 (Princeton University Press 1920).
152 Henry Herbert Goddard, Human Efficiency and Levels of Intelligence 121 (Princeton University Press 1920).
153 Henry Herbert Goddard, Human Efficiency and Levels of Intelligence 121 (Princeton University Press 1920).
procreation by people with low intelligence, Thorndike applied his prior studies of ethology when he asserted “selective breeding can alter a man’s capacity to learn, to keep sane, to cherish justice, or to be happy.” By 1922, Thorndike had developed his own assessment of intelligence, *Tests for Mental Alertness*. After Columbia University began using this test as part of its admissions criteria, Jewish enrollment declined by half, from 40 percent to 20 percent. However, Thorndike did not appear to be entirely fatalistic in his view of innate intelligence and the ability to learn.

He posited if psychology were properly applied, one could transform education by learning “the effect of every possible stimulus and the cause of every possible response in every possible human being.” Thorndike also challenged traditional pedagogy by asserting the belief that all instruction was transferable to the world of work. As an example, he compared the applicability of mathematics to vocations as varied as those of scientists, grocers, and carpenters to that of the type of writing typical of schools of the era. “To write ‘letters to a friend,’ or ‘stories about a day in the country,’ or ‘essays on the characters in The House of Seven Gables’ is not the same thing as to write an efficient business proposal.”

The ability to write a business proposal as an important goal of education was fitting for the Efficiency Era, given its roots in the private sector. One would be mistaken to assume the involvement of early twentieth century business leaders in the dialogue about public schools

ended with the influence of Frederick Taylor on educational scholars. Three years after the release of *The Cleveland Study*, Edwin Salisbury Carman, one of Taylor’s successors as President of the American Society of Mechanical Engineers, participated in a panel, entitled “The Problems of Education for the Industries,” in Cleveland with the Assistant Superintendent of Schools and the Principals of East and West Technical High Schools entitled. Even Thorndike’s contemporary John Dewey, who decried the compartmentalization of education, including technical schools such as those favored by Taylor remarked that waste in education from a child’s point of view included being “unable to apply in daily life what he is learning at school.”

1930s to Mid-1950s: The Technocratic Meritocracy

Some educational researchers make little mention of the time that elapsed between the Scientific Efficiency Era and the next scientifically-driven wave of reform heralded by the advent of Sputnik in 1957. Those who choose to proffer an opinion are often divided. Educational researcher C.H. Judd posited that after World War I, standardized tests had “quietly gone on their way.” An advisory panel convened by the United States Congress, however, quoted educational historian Lawrence A. Cremin describing the students of the 1940s as “the

---

159 *The Journal of the Cleveland Engineering Society* 267 (1918). In the interest of full disclosure, E.S. Carman was the great-grandfather of the author.


most tested generation of youngsters in history.”\textsuperscript{163} Whatever the prevalence of testing during this era, changes in the 1930s and 1940s laid the foundation for several data and assessment practices which have persisted into the 21\textsuperscript{st} Century.

Once again, educational shifts were driven by economics. Following a boom of more than two decades, the Great Depression dominated the United States landscape in the 1930s.\textsuperscript{164} The Depression generated significant public concern as to whether students were being guided into higher education and ultimately toward the limited employment opportunities available. This concern, coinciding with increased state funding for education, resulted in the formation of the first state agencies involved in educational data analysis standardized testing.\textsuperscript{165}

One such agency was the New York Board of Regents, who commissioned an eponymous study in 1935, “not to gather great masses of statistics”\textsuperscript{166} but rather to identify possible solutions to major issues that resulted in “a great number of youth each year”\textsuperscript{167} who were not adequately prepared for life.\textsuperscript{168} This pervasive ill-preparedness of both graduates and dropouts in 1930s New York included students who were not ready to “go to work and later to


adapt themselves to shifting economic conditions”\textsuperscript{169} or to pursue “advanced or professional education.”\textsuperscript{170} The Regents’ Study identified several key reasons the education system in the State of New York was failing to meet the needs of students in the 1930s. These included a need to redesign education to meet “new and changing work opportunities [students] must face in modern life.”\textsuperscript{171} More specifically, the public school system in New York had not “caught up with the flood of new scientific knowledge”\textsuperscript{172} and failed to impart a scientific view of the world to its students. In addition to changing vocational needs and scientific advances, the Regents found the schools had also neglected to respond to shifting demographics. High schools in particular had once been the domain of the college-bound, but they now found themselves charged with educating “all the children of all the people, with their many new and different needs.”\textsuperscript{173}

To address the aforementioned problems, the Regents concluded their four-year study by providing an “Outline of a New Educational Program.”\textsuperscript{174} Suggested reforms included abandoning the historical grammar school structure in favor of a secondary system that commenced at grade seven, creating junior highs and high schools of 300-1,200 students to offer


electives and extracurricular activities, establishing guidance departments, improving libraries, and providing physical education. Schools were also admonished to “give more attention specifically to gifted youth and to handicapped youth, not only for their sake and the future of society, but also as a means of improving instruction for the average group of students.” One of the Regents’ reforms heralded the future trend of using state assessments for accountability. The Regents’ Exams were to no longer serve as graduation tests but instead be transformed into “examinations designed to discover the weak spots in curriculum and teaching.”

Prior to this time, as evidenced by the Cleveland Study, testing and research had enjoyed growth primarily at the local levels, with nearly 60 larger school systems establishing research bureaus to analyze data and coordinate standardized testing. By the end of the decade, nine of the ten largest cities in the United States had research offices. The New York Regents’ Study and the similar Pennsylvania Study marked the beginning of a new era in the 1930s, during which the role of assessment and research began to shift to the state level. In 1937, fifteen states participated in a conference organized by the American Council on Education, Committee of

---


179 Daniel P. Resnick, Competency Testing Historically Considered, 8 Review of Research in Education 7 (1980).

Measurement and guidance.\textsuperscript{181} At first, however, the role of states in assessment and research was limited, serving in more of a consultant role, offering “a range of back-up and advisory services to the local school districts, occasionally initiating testing programs for the gifted or the handicapped, or supporting the introduction of new forms of guidance testing at the district level.”\textsuperscript{182}

The nascent emergence of state leadership did not mean, however, that the role of industrialists and corporate funding in education reform had ended. Although the Pennsylvania Study involved gathering data on high school seniors statewide in Pennsylvania\textsuperscript{183} from 1928 through 1932,\textsuperscript{184} it was funded not by the state, but by the Carnegie Corporation.\textsuperscript{185} Carnegie also sponsored multi-state research by the Progressive Education Association, which came to be known as the Eight-Year Study,\textsuperscript{186} a reference to the fact that it was conducted from 1932 through 1940 under the leadership of Ralph W. Tyler.\textsuperscript{187}


\textsuperscript{186} Wilford M. Aikin, Adventure in American Education Volume 1: The Story of the Eight-Year Study preface (Harper and Brothers 1942).

The Progressive Education Commission, a group of “gray-haired principals and teachers”\textsuperscript{188} and “young teachers recently out of college”\textsuperscript{189} that had previously focused on reforms at the elementary level, met in Washington, D.C. in April 1930 to discuss how to provide students with “experiences in high school that would develop their powers and equip them to assist in the rebuilding of our already profoundly disturbed national life.”\textsuperscript{190} The group decided that such a task would take far longer than the duration of the two-day conference, so they established the Commission on the Relation of School and College,\textsuperscript{191} a working study group comprised of representatives from thirty public and private high schools.\textsuperscript{192} Citing statistics that indicated only 60 percent of students who entered high school graduated and only 10 percent matriculated to college, the commission identified key problems they felt were the cause.

High schools failed to prepare students for civic life, leaving them “without insight into the great political, social, and economic problems”\textsuperscript{193} of the nation, in no small part because they were “autocratic, rather than democratic.”\textsuperscript{194} The committee felt it was also far too easy for a

\textsuperscript{188} Wilford M. Aikin, Adventure in American Education Volume 1: The Story of the Eight-Year Study 1 (Harper and Brothers 1942).

\textsuperscript{189} Wilford M. Aikin, Adventure in American Education Volume 1: The Story of the Eight-Year Study 1 (Harper and Brothers 1942).

\textsuperscript{190} Wilford M. Aikin, Adventure in American Education Volume 1: The Story of the Eight-Year Study 1 (Harper and Brothers 1942).

\textsuperscript{191} Wilford M. Aikin, Adventure in American Education Volume 1: The Story of the Eight-Year Study 2 (Harper and Brothers 1942).

\textsuperscript{192} Wilford M. Aikin, Adventure in American Education Volume 1: The Story of the Eight-Year Study 1 (Harper and Brothers 1942).

\textsuperscript{193} Wilford M. Aikin, Adventure in American Education Volume 1: The Story of the Eight-Year Study 4 (Harper and Brothers 1942).

\textsuperscript{194} Wilford M. Aikin, Adventure in American Education Volume 1: The Story of the Eight-Year Study 4 (Harper and Brothers 1942).
student to merely “get his lessons.”\textsuperscript{195} “Many a brilliant mind developed habits of laziness, carelessness, superficiality,”\textsuperscript{196} so that “even the conscientious student of superior ability did not often find himself seriously involved in great intellectual enterprise”\textsuperscript{197} that would render him willing to “go through whatever drudgery might be necessary to achieve his purpose.”\textsuperscript{198}

The commission felt these ills were due in no small part to the fact that “the classroom was formal and completely dominated by the teacher”\textsuperscript{199} and students were busy merely “doing assignments.”\textsuperscript{200} The classroom environment was not the only problem identified. The commission also found that “young people wanted to get ready to earn a living, to understand themselves, to learn how to get on with others, to become responsible members of the adult community, and to find meaning in living. The curriculum seldom touched on such genuine problems of living.”\textsuperscript{201} Instead, the course sequence “resembled a picture puzzle”\textsuperscript{202} whereby the

\textsuperscript{195} Wilford M. Aikin, Adventure in American Education Volume 1: The Story of the Eight-Year Study 5 (Harper and Brothers 1942).

\textsuperscript{196} Wilford M. Aikin, Adventure in American Education Volume 1: The Story of the Eight-Year Study 5 (Harper and Brothers 1942).

\textsuperscript{197} Wilford M. Aikin, Adventure in American Education Volume 1: The Story of the Eight-Year Study 5 (Harper and Brothers 1942).

\textsuperscript{198} Wilford M. Aikin, Adventure in American Education Volume 1: The Story of the Eight-Year Study 5 (Harper and Brothers 1942).

\textsuperscript{199} Wilford M. Aikin, Adventure in American Education Volume 1: The Story of the Eight-Year Study 6 (Harper and Brothers 1942).

\textsuperscript{200} Wilford M. Aikin, Adventure in American Education Volume 1: The Story of the Eight-Year Study 6 (Harper and Brothers 1942).

\textsuperscript{201} Wilford M. Aikin, Adventure in American Education Volume 1: The Story of the Eight-Year Study 6 (Harper and Brothers 1942).

\textsuperscript{202} Wilford M. Aikin, Adventure in American Education Volume 1: The Story of the Eight-Year Study 8 (Harper and Brothers 1942).
student’s “chief purpose was to collect enough pieces to graduate.”203 The teachers themselves, meanwhile, worked in isolation. “Seldom did they confer, and when they did, the results were usually unsatisfactory”204 because “common language and unity of purpose were in danger of being lost.”205

All of this had transpired in spite of the fact that emphasis on measurement in education had increased dramatically since the beginning of the 20th Century, with “numberless tests”206 being “devised, published, and used in schools.”207 The committee found, however, that “most of the tests used by schools were designed to measure chiefly accretions of information and proficiency in certain skills”208 in spite of the fact that “every school has other purposes that it believes to be equally, if not more, important.”209 To this end, the 30 schools participating in the Eight-Year Study “took the position that evaluation is important only in relation to purpose. Unless objectives are clearly defined, there can be no significant measurement of results.”210

---


One principal even drew on a Biblical analogy, positing that “the results sought by a school must be constantly before the faculty as a pillar of cloud by day and a pillar of fire by night.”

As a result, early in the study, the schools dedicated themselves to “thinking through and stating plainly the objectives they hoped to achieve,” including helping students to “work satisfactorily with others, to read intelligently and express themselves well in speech and in writing, and to learn how to investigate a topic and follow its leadings.” The commission soon found that “standardized tests were usually based upon the traditional content of conventional subjects” and established an Evaluation Service because “more comprehensive programs of appraisal were needed.” The staff of the Evaluation Service “worked intimately with the thirty schools,” with a mission to “help develop effective ways to find out what changes were produced in students by their school experiences.”

Teachers participated in every step of not only test design, but in revisions of assessments after the initial administration, with the published summary of the Eight-Year Study emphasizing, “This was done co-operatively.”

---


The schools were saying to the Evaluation Staff, ‘We do not know surely whether our work is producing the results we desire. We need to know…..If our teaching is not bringing about these results, we shall change our curriculum and teaching methods in the hope that we can accomplish our purposes. Then we shall examine results again.’

For their part, the Eight-Year Study Evaluation Staff responded to the challenge.

The task is difficult. Many technicians have said that it is impossible to devise reliable measures of progress toward such intangible objectives. We think it can be done. It will take time. The first instruments we construct may not be satisfactory. If you will try them out in your classes, we will discover wherein the tests are faulty and try again. We hope that eventually we shall be able to provide instruments of evaluation that will be useful to you in appraising the results of your work.

In the course of the study, the Evaluation Staff and teachers from the 30 schools created around two hundred tests that were “used experimentally, refined, and tried out again and again.” Sixteen of the instruments saw wide use across secondary schools. The topics of these included interpretation of data, application of the general principles of science, social problems, logic, and familiarity with sources of information. Perhaps the most lasting contribution of the Eight-Year Study Evaluation Staff, however, was that they “rendered another service equally important: they taught hundreds of teachers how to devise their own tests.”

The Eight-Year Study “confirmed the importance of following student progress on a continuous basis, recording data from standardized tests as well as other kinds of

---


achievement” and “demonstrated the potential power of educational research as an agent of change.” The study also affirmed a pattern of belief that “high school performance ought to be especially closely followed in periods of youth unemployment.” When they considered who should monitor high school performance, Eight-Year Study Evaluation Staff felt that states “appeared as the appropriate geographical, political, and educational unit for such an inquiry, challenging the almost exclusively local district basis for earlier testing activity.”

One state, however, did not wait for the findings of the Eight-Year Study. In 1929, the University of Iowa, under the leadership of Dr. E.F. Lindquist, started the first statewide testing program for high schools in 1929. Participation was voluntary, but most high schools in the state of Iowa chose to administer the assessments. The tests were revised on an ongoing basis, with new editions released annually. The directions and scripts used to administer the tests were highly structured, laying the groundwork for future standardized achievement tests.

---


The initial raison d’être for the test was a competition originally known as “The Iowa Academic Meet,”\textsuperscript{230} deliberately named to resemble a track meet to convey the notion that academics should be considered as important as athletics.\textsuperscript{231} Newspapers, however, had another idea for a moniker, quickly re-christening the competition “The Brain Derby.”\textsuperscript{232} Participating schools administered one-hour achievement tests to their students, which assessed each of the core high school subjects.

Top-scoring students from each school advanced to a district-level contest, during which they took additional examinations. Results from these exams were used to select the top 1,000 students, who convened at the University of Iowa in June to take two-hour examinations in various subjects.\textsuperscript{233} While students received awards, the “Brain Derby” pertained not only to individuals, but to schools. High schools with the highest composite\textsuperscript{234} weighted averages on the initial round of “all pupil”\textsuperscript{235} assessments received an award\textsuperscript{236} and were announced at the final competition in Iowa City.\textsuperscript{237}

\textsuperscript{230} E.F. Lindquist, \textit{The Iowa Testing Programs – A Retrospective View}, 91 Education 8 (1970).
\textsuperscript{231} E.F. Lindquist, \textit{The Iowa Testing Programs – A Retrospective View}, 91 Education 8 (1970).
\textsuperscript{232} E.F. Lindquist, \textit{The Iowa Testing Programs – A Retrospective View}, 91 Education 8 (1970).
\textsuperscript{233} E.F. Lindquist, \textit{The Iowa Testing Programs – A Retrospective View}, 91 Education 8 (1970).
\textsuperscript{235} E.F. Lindquist, \textit{The Iowa Testing Programs – A Retrospective View}, 91 Education 8 (1970).
\textsuperscript{237} E.F. Lindquist, \textit{The Iowa Testing Programs – A Retrospective View}, 91 Education 8 (1970).
These early days of the Iowa testing program “served as a golden opportunity to experiment with, and learn how to improve objective tests,” according to Lindquist. The team who created the assessments—primarily faculty from the University of Iowa and department heads from University High School—performed what Lindquist described as “a thorough post mortem” each year on both the test items (questions) and the system as a whole, a process through which they introduced innovations such as the first separate answer documents to expedite scoring. They also “learned a great deal about desirable item characteristics.” It was an observation of external factors, however, that may have precipitated the most substantial change to the types of items which comprised the early Iowa tests. According to Lindquist,

Beneficial and rewarding as the program was in many respects, too much emphasis on the competitive features began to build up. This led to an over-emphasis on the teaching of informational content, and upon the rote learning of facts, since that is what many teachers felt would be most effective in improving test standing. Teachers tended to become more and more teachers of individual subjects, rather than mentors of the pupils most interested in their well-rounded over-all development. In a few instances, teachers were even hired and fired on the basis of the average tests standing of their pupils.

In response to these developments, the Iowa item writers reduced the number of questions about “detailed, specific information,” replacing them with items assessing students’ “reasoned understanding of broad concepts and large units of comprehension, and the ability to do some critical thinking about the material studied.” Almost equally important was

communication to counteract teacher perception that rote memorization was the key to success. The Iowa writers “tried to draw as much attention as possible, through manuals and otherwise, to these characteristics of the tests, and thus to improve instruction.”\textsuperscript{245} In addition to changes in the composition of the tests, the district-level awards were also discontinued, and schools were allowed to administer the exams to their students without entering the school competition.\textsuperscript{246}

In spite of the unintended negative consequences, the program did benefit students by helping schools identify those who may have otherwise been overlooked. In the words of founder Lindquist,

\begin{quote}
We found many instances in which students astounded their teachers by scoring at the top of the test distributions—students who before had been regarded as unpromising or ‘problem’ students and who had regularly been getting grades of C, D and F, perhaps because they refused or failed to follow routine assignments slavishly, and instead did things such as pursuing hobbies or going off to the library to read on their own. Interestingly enough, once those pupils had been identified by the tests as talented, their grades often improved quite suddenly!\textsuperscript{247}
\end{quote}

Several decades later, Lindquist observed that “startling discoveries of this kind at the high school level are seldom reported today,”\textsuperscript{248} attributing this to the use of standardized testing in the lower grades.\textsuperscript{249}

The Iowa testing system was at the forefront of this trend. Between 1935 and 1940, the battery of assessments was expanded down to grades three through eight.\textsuperscript{250}

\textsuperscript{245} E.F. Lindquist, \textit{The Iowa Testing Programs – A Retrospective View}, 91 Education 9 (1970).
\textsuperscript{246} E.F. Lindquist, \textit{The Iowa Testing Programs – A Retrospective View}, 91 Education 9 (1970).
\textsuperscript{247} E.F. Lindquist, \textit{The Iowa Testing Programs – A Retrospective View}, 91 Education 8 (1970).
\textsuperscript{248} E.F. Lindquist, \textit{The Iowa Testing Programs – A Retrospective View}, 91 Education 8 (1970).
\textsuperscript{249} E.F. Lindquist, \textit{The Iowa Testing Programs – A Retrospective View}, 91 Education 8 (1970).
assessments were released first, in 1935. The battery for grades six through eight, which comprised the first version of the *Iowa Test of Basic Skill (ITBS)*, “succeeded from the beginning in placing major emphasis on the development of basic skills and of generalized intellectual abilities, rather than on the rote learning of subject matter.” The statistical reporting of the new system was also innovative, including grade-equivalent scales that were comparable year-to-year and student profile charts. These reflected the primary use of the data: the individualization of instruction for and guidance of students. However, the data also included reports of school averages. In 1940, tests for grades three through five were added. The early elementary tests included not only grade-level content, but also below-level and above-level content. Furthermore, in 1940, Houghton Mifflin assumed responsibility for publication of ITBS, starting a trend of for-profit publishing companies being involved in assessments.

The need for a relationship with Houghton Mifflin may have been linked to the expansion of ITBS beyond the state of Iowa. “In just a few years,” claimed Lindquist, “our test authors were turning out tests of such good quality as to attract attention outside the state,” after which ITBS “built up a considerable volume of sales of the tests to non-Iowa users.” While the Iowa tests may have been of exceptional quality, the popularity of ITBS may have been

---


influenced by the “converging concerns over the employability of students and their achievement levels” triggered by the Great Depression.

By the end of the Depression, one-third of states had instituted testing programs. According to educational historian Daniel P. Resnick, this reflected the expectation—very germane to the Depression era—that “high school performance ought to be especially closely followed in periods of youth unemployment.” Resnick further clarified at which level the monitoring should take place, asserting that “the state, moreover, appeared as the appropriate geographical, political, and educational unit for such an inquiry, challenging the almost exclusively local district basis for earlier testing activity.”

While large-scale studies and the Brain Derby were generating excitement at the high school level in the Depression-era Midwestern United States, changing socio-economic forces were also shaking up the elite colleges of the East Coast. This included Harvard, which welcomed a new president in 1933. The son of an engraver, James Bryant Conant hailed from a decidedly middle-class background and had specialized in the workmanlike study of chemistry, as opposed to the more typical Ivy League disciplines of literature or history. The selection of a president who possessed such an anomalous pedigree and “burned with a fierce disapproval of the old ways of Harvard” was a deliberate move by the Harvard Corporation, to whom

---


Conant was very direct during the interview process about the problems he had observed at the school.\textsuperscript{262} Just as states were increasing their scrutiny of high schools, the United States elites were also impacted by the 1929 stock market crash and the Great Depression that ensued. As a result, “The linked issues of what would succeed this old elite and how American education should be organized were being debated”\textsuperscript{263} throughout the hallowed halls of the Ivy League, including Harvard.

Conant first turned his attention to the student body. His predecessor, Abbott Lawrence Lowell, had “presided over the gradual decay of Harvard as a first class institution devoted to learning and research in favor of preserving Harvard as a bastion for the elite of America.”\textsuperscript{264} This had been exacerbated by the Depression, when “paying students were in short supply.”\textsuperscript{265} As a result, “to be admitted was not much of a feat if you had the money and the right background.”\textsuperscript{266}

Two issues arose at Harvard as a result. First, “as a group, these young men were not notably studious.”\textsuperscript{267} Like the real-life embodiment of Evelyn Waugh’s character Sebastian Flyte in \textit{Brideshead Revisited},\textsuperscript{268} “rich young men conducted a life barely recognizable to today’s

\begin{itemize}
\item \textsuperscript{261} Nicholas Lemann, The Big Test: The Secret of the American Meritocracy 21 (Farrar, Straus, and Giroux 2000).
\item \textsuperscript{263} Nicholas Lemann, The Big Test: The Secret of the American Meritocracy 21 (Farrar, Straus, and Giroux 2000).
\item \textsuperscript{265} Nicholas Lemann, The Big Test: The Secret of the American Meritocracy 27 (Farrar, Straus, and Giroux 2000).
\item \textsuperscript{266} Nicholas Lemann, The Big Test: The Secret of the American Meritocracy 27 (Farrar, Straus, and Giroux 2000).
\item \textsuperscript{267} Nicholas Lemann, The Big Test: The Secret of the American Meritocracy 27 (Farrar, Straus, and Giroux 2000).
\item \textsuperscript{268} Evelyn Waugh, Brideshead Revisited 56 (Little, Brown, & Co. 1944)
\end{itemize}
college students. At a time when a quarter of the American workforce was unemployed and desperate, they lived in private apartments, attended by butlers and maids, in a district called the Gold Coast; went to debutante balls in Boston; did not customarily attend classes; and enrolled briefly in special tutoring schools at the end of each semester so they would be able to pass their exams.”

Conant believed the second problem generated by the Harvard student population of the early 1930s was reflective of one impacting American society in general. It threatened the very Jeffersonian ideals on which the country was founded. As he wrote in The Atlantic (which one might ironically observe is a publication favored by the very elites he was challenging),

A high degree of social mobility is the essence of the American ideal of a classless society. If large numbers of young people can develop their own capacities irrespective of the economic status of their parents, then social mobility is high. If, on the other hand, the future of a young man or woman is determined almost entirely by inherited privilege or the lack of it, social mobility is nonexistent.

The first step Conant took to execute his vision of equity was to create a new four-year full scholarship based solely on academic promise. Conant assigned the task of selecting scholarship students to a pair of assistant deans, Wilbur Bender and an earnest young man named Henry Chauncey. Identifying students for the new scholarship would require an assessment. Chauncey, a “testing enthusiast,” first developed his fondness for psychometrics while a student at Ohio State University. Herbert Toops, a student of Edward Thorndike, administered

---


the first “psychological examination” to Chauncey’s freshman class. When the test correctly identified Chauncey’s grades, he was intrigued. Chauncey continued to voraciously study the relatively young field of testing the following year at Harvard, where he was eventually employed. Chauncey later reflected,

I think I was interested in the full development of each individual. And one could learn about individuals from tests…It seemed to me that we knew more about the horses in the country than we knew about the people of the country. And that it would be useful to know more about all the different people.

However, the College Boards—a weeklong battery of essay exams administered at exclusive boarding schools and favored by Ivy League admissions personnel up to that time—had done nothing to alleviate the lack of social mobility or enhance the intellectual prowess reflected in the Harvard student body. “From Conant’s point of view, the problem with the college boards was that they were so much a test of mastery of the boarding-school curriculum that they couldn’t be used to size up the Midwestern public-school boys he wanted to bring to Harvard.”

Chauncey and Bender traveled to Princeton to meet with Carl Brigham. While promoting the Scholastic Aptitude test (SAT), Brigham expressed a view of the College Board exams similar to that of Conant, calling for innovation and research with “the virility of an

---

experimental attack. Not only was the SAT geared more toward individual aptitude as opposed to a mastery of curriculum, but it could also be scored more objectively and efficiently. In Chauncey’s own words as he later reflected on this era, “The adoption of ‘objective,’ or restricted-answer, examinations made it possible to reduce test scoring to a clerical operation, thus considerably lowering costs and increasing the reliability of scoring.”

Brigham modeled the original version of the SAT after the Army Alpha Intelligence test developed by his former mentor Yerkes. The Alpha had been used during World War I to determine which army recruits were “officer material.”

The early administrations of the SAT, which made its debut in 1926, were given in addition to other placement tests for the sole purpose of validating the new assessment, a practice that at the time involved correlating scores to freshman college grades. During the seven years that had elapsed, however, Brigham had changed. Although he was a former assistant to eugenicist Robert Yerkes, “his educational philosophy was never static from one season to another.” By the time he met Chauncey, Brigham had publicly renounced not only his former views, but even his most noted book, A Study of American Intelligence. His next book, published

in 1932, just a year before he met Chauncey, was entitled *A Study of Error*\textsuperscript{286} The revisions to the SAT reflected this intellectual change. Initially, the SAT had generated a single score, as would an intelligence test. In fact, Brigham had even created “a crude scale for converting it to an IQ score.”\textsuperscript{287} The revised version, however, resulted in two scores, one verbal and one mathematical, and the IQ conversion scale was abandoned.\textsuperscript{288} In January 1934, scores from this version of the SAT were utilized as one of the criteria to identify ten Harvard scholarship students, who were to be the first of many.\textsuperscript{289} Chauncey later recounted one such student during an interview:

Now I'll give you an example. When I was in charge of the scholarships at Harvard, there was a fellow from Kentucky, not far from Nashville, Tennessee. He happened to see a poster on the bulletin board of the school about the Harvard national scholarships, and somebody suggested to him he ought to apply. And so he did apply. He did well. The principal didn't have much excitement about this but I sent a letter to the boy asking if he could meet me in Nashville for an interview. Well, the principal then took him down, got him a haircut, and got him properly dressed and drove him down. And I interviewed him, and he won a scholarship.\textsuperscript{290}

He came from a family where he was the seventh of fifteen children, virtually none of whom had had any significant time in school. He himself had been in school for only about a year's worth at the time he was, he had an accident, which is what caused it all. His brother was chopping down a bushy tree, and by mistake, hit him on his top of his foot and severed some tendons, so he couldn't work on the farm. So he was allowed to go to school. And then because of the test score, he was found out.\textsuperscript{291}

\textsuperscript{286} Nicholas Lemann, The Big Test: The Secret of the American Meritocracy 29 (Farrar, Straus, and Giroux 2000).

\textsuperscript{287} Nicholas Lemann, The Big Test: The Secret of the American Meritocracy 33 (Farrar, Straus, and Giroux 2000).

\textsuperscript{288} Nicholas Lemann, The Big Test: The Secret of the American Meritocracy 33 (Farrar, Straus, and Giroux 2000).

\textsuperscript{289} Nicholas Lemann, The Big Test: The Secret of the American Meritocracy 38 (Farrar, Straus, and Giroux 2000).

\textsuperscript{290} *Interview with Henry Chauncey, founder of the SAT and first Director of ETS*, PBS Frontline (2000), [http://www.pbs.org/wgbh/pages/frontline/shows/sats/interviews/chauncey.html](http://www.pbs.org/wgbh/pages/frontline/shows/sats/interviews/chauncey.html).

\textsuperscript{291} *Interview with Henry Chauncey, founder of the SAT and first Director of ETS*, PBS Frontline (2000), [http://www.pbs.org/wgbh/pages/frontline/shows/sats/interviews/chauncey.html](http://www.pbs.org/wgbh/pages/frontline/shows/sats/interviews/chauncey.html).
Not only did the administration of the SAT identify scholarship students, but it also drew other new students to the school, “undermining what it had meant to be a ‘college man’ in earlier decades.”292 According to Henry Chauncey’s son, also named Henry, but commonly known as Sam,

He believed in meritocracy; he believed that the people who should be admitted to colleges were people who deserved it based on their intelligence and achievement. When he started the national scholarship program at Harvard, it only admitted young people from blueblood schools. He crisscrossed the country, looking for qualified people from little-known high schools…He found James Tobin and Casper Weinberger and John Morton Blum.293

This success soon spread to other Ivy League schools, but it took a war for the SAT to replace the traditional College Board exams. When the Japanese bombed Pearl Harbor on December 7, 1941, the College Board officials were having lunch in Princeton. Given the state of emergency, scoring essay exams was not feasible. They were suspended—never to resume—in favor of the SAT.294 Conant, however, wanted more: an organization that would combine the SAT with other tests.295 Others agreed, referring to the College Board as “narrow and provincial.”296 Brigham, however, was strongly opposed, even issuing a written warning in a scholarly journal, School and Society, about “the possible hazards to education in setting up any

---


organization to perpetuate the present types of tests.” In 1943, Brigham’s death at the age of 52 halted this opposition, paving the way for the founding of Educational Testing Service.

In the meantime, someone needed to lead the College Board. They recruited Henry Chauncey to leave Harvard and serve as their president. Chauncey had been conducting some research at Harvard and spoke about it with Conant in 1945, a conversation he recounted years later for an interviewer:

I had hoped that maybe Conant would be interested in increasing the amount of activity in this area. I went to see him, and as we talked, I indicated this other opportunity of going to the College Board. He said, ‘Well, how much salary do you want?’ Well, that wasn't what I was interested in. So I was a little disappointed he didn't seem to respond to what I had really been interested in.

In spite of his disappointment in Conant, Chauncey did not make the decision to leave lightly, but instead he took a quantitative approach:

I thought about it very carefully and put down a whole list of criteria by which I might make a decision as to whether to stay at Harvard or go to the College Board. And it was a close call, but there were 34 points for going to the College Board, and 32 for staying at Harvard. One doesn't always abide by that kind of thing. It might have been if it came out 34/32 the other way, and I still would have gone.

According to Henry Chauncey’s obituary,

The Educational Testing Service was a post-World War II creation. In 1946, hundreds of thousands of veterans were clamoring to enroll in college. To streamline the admission process, the Carnegie Committee on Testing recommended merging the American

---


Council of Education, the Carnegie Foundation for the Advancement of Testing, and the College Entrance Examination Board.\textsuperscript{301}

The report of the Carnegie committee “emphasized the overlapping, duplication, and consequent waste”\textsuperscript{302} of multiple agencies operating competitively in the field of testing. Their recommendation was to form “a single testing agency which could unify, strengthen, and expand the present testing functions of the nonprofit agencies”\textsuperscript{303} and also “sponsor distinguished research”\textsuperscript{304} on testing.\textsuperscript{305}

As was the case with the SAT, it took a national-level catalyst to enact change. George Zook, former Commissioner of Education under President Franklin D. Roosevelt, was appointed to lead the Presidential Commission on Higher Education by President Harry Truman. Zook submitted his report to Truman on December 11, 1947, calling for higher education enrollment to triple to 4.6 million—or 30 percent of each graduating cohort—matriculating to college by 1960.\textsuperscript{306} Just a week after the Zook Report was issued, the Educational Testing Service, or ETS, was chartered, fulfilling the vision of a combined testing conglomerate. With Conant as chairman and Chauncey as president,\textsuperscript{307} ETS was launched as a test-development and research


\textsuperscript{302} Claude M. Fuess, \textit{The College Board: Its First Fifty Years} 184 (Columbia University Press 1950).

\textsuperscript{303} Claude M. Fuess, \textit{The College Board: Its First Fifty Years} 185 (Columbia University Press 1950).

\textsuperscript{304} Claude M. Fuess, \textit{The College Board: Its First Fifty Years} 185 (Columbia University Press 1950).

\textsuperscript{305} Claude M. Fuess, \textit{The College Board: Its First Fifty Years} 185 (Columbia University Press 1950).

\textsuperscript{306} Nicholas Lemann, \textit{The Big Test: The Secret of the American Meritocracy} 64 (Farrar, Straus, and Giroux 2000).

\textsuperscript{307} Nicholas Lemann, \textit{The Big Test: The Secret of the American Meritocracy} 65 (Farrar, Straus, and Giroux 2000).
agency working on behalf of the College Board,\footnote{About ETS: What We Do, ETS, \url{https://www.ets.org/about/what/} (last visited Jul. 15, 2016).} with the hopes of rendering it less “narrow and provincial”\footnote{Claude M. Fuess, The College Board: Its First Fifty Years 111 (Columbia University Press 1950).} as its critics had previously described it.

Mid to Late 1950s: Era of Excellence

ETS was not the only product of the post-World War II era. In addition to the 1944 G.I. Bill that enabled many veterans to become the first in their families to attend college,\footnote{Education and Training: History and Timeline, Veterans’ Administration As retrieved from \url{http://www.benefits.va.gov/gibill/history.asp} (last visited Aug.11, 2016).} another result of World War II was the Cold War between the West and the Soviet Union. Many historians view the events that took place in the Crimean port city of Yalta as the “shorthand explanation of the origins of the Cold War.”\footnote{Fraser J. Harbutt, Yalta 1945: Europe and America at the Crossroads 3 (Cambridge University Press 2010).} The eponymous Yalta Conference was attended by three nations: the United Kingdom, represented by Prime Minister Winston Churchill; the United States, led by President Franklin Delano Roosevelt; and the Soviet Union, under the control of Communist Party General Secretary Joseph Stalin. The results of the conference “briefly created a widespread sense of euphoria.”\footnote{Fraser J. Harbutt, Yalta 1945: Europe and America at the Crossroads 333 (Cambridge University Press 2010).} The United States framed the results using a “glowing postwar vision organized on Atlantic Charter lines,”\footnote{Fraser J. Harbutt, Yalta 1945: Europe and America at the Crossroads 341 (Cambridge University Press 2010).} while the Soviet propaganda organization Pravda “reaffirmed ‘unity’ of the Big Three”\footnote{Fraser J. Harbutt, Yalta 1945: Europe and America at the Crossroads 333 (Cambridge University Press 2010).} against Adolph Hitler. Yes, the West had made considerable concessions to communism, but they chose the route of “crowning
the artifice for a world audience after Yalta with the false vision of a reformed Soviet Union bent on bringing democracy to Eastern Europe.”

The reality, however, was much more complicated. Stalin had successfully driven a wedge between the United Kingdom and the United States, a task made easier by the fact that Winston Churchill, while concerned about communist expansion, also had trepidations about the potential for “looming American hegemony.” As a result, “Stalin outdid both of them [Churchill and Roosevelt] in the scope and vigor of his self-assertion” and proceeded to take over Poland, decline “diplomatic opportunities to restore trust,” and “made the first Security Council meeting in London into a bear pit.” Thus, by the 1950s the United States was “a fearful place in the grips of a Cold War with the Soviet Union.”

On October 4, 1957, in a maneuver typical of the Cold War era, rather than actually dispatching a bomb, the Soviets instead incited panic by launching the first space satellite, Sputnik I. In spite of the fact that it was the size of a beach ball, weighed only 184 pounds, and was fueled by kerosene, Sputnik created a “sudden crisis of confidence in American

---

315 Fraser J. Harbutt, Yalta 1945: Europe and America at the Crossroads 404 (Cambridge University Press 2010).
316 Fraser J. Harbutt, Yalta 1945: Europe and America at the Crossroads 406 (Cambridge University Press 2010).
317 Fraser J. Harbutt, Yalta 1945: Europe and America at the Crossroads 407 (Cambridge University Press 2010).
318 Fraser J. Harbutt, Yalta 1945: Europe and America at the Crossroads 405 (Cambridge University Press 2010).
319 Fraser J. Harbutt, Yalta 1945: Europe and America at the Crossroads 405 (Cambridge University Press 2010).
320 Fraser J. Harbutt, Yalta 1945: Europe and America at the Crossroads 405 (Cambridge University Press 2010).
technology, values, politics, and the military.”324 This crisis was undoubtedly fueled at least in part by the hyperbolic descriptions given in the media. A newsreel film declared that Russia had “launched a man-made moon into outer space”325 and that “for a little fella, he’s certainly getting a lot of attention.”326 Sputnik continued to transmit its high-pitched signal until October 26, at which point it had weakened to the extent that Radio Moscow declared it dead.327

Before the American public, who were “already cringing in their bomb shelters”328 prior to Sputnik, could calm themselves, the Soviets launched Sputnik II on November 3, 1957. Not only did this satellite weigh over 1,000 pounds, but it also carried a specially-trained dog and the life-support equipment she needed to survive in space.329 News footage from the time showed a terrified little canine being loaded into the satellite, claiming that “Laika, the first space traveler, was ready for the takeoff.”330

Just over a month later, on December 6, 1957, “a shocked America attempted to launch a grapefruit-sized satellite on the Vanguard rocket with disastrous results.”331 Footage of the rocket failing to launch and being enveloped in flames rolled across the newsreel screen as the

326 Movietone News, Sputnik Spotted and Filmed from Two Far Removed Sites, youtube.com (Movietone 1957), https://www.youtube.com/watch?v=FsPKD4tNe-Y.
announcer referred to “Russia’s space supremacy.” The report concluded with alarmist rhetoric typical of the Cold War era. “In the rocket’s fiery wake was America’s sober realization that the battle has just been joined and that the work of self-preservation was at hand.”

The aforementioned sober reflection soon morphed into scrutiny of education, as “Sputnik came to represent in the mass media another example of the U.S. educational system asleep at the wheel while a foreign nation drove away with the prize of first place in the space race.” A 1957 newsreel showed clips of Soviet propaganda films, including one of student groups tracking Sputnik. As footage of studious-looking Soviet teens unspooled in the background, the American announce claimed that the clips served to “underscore the emphasis on science in Russian schools.” The newsreel narrator went on to add that Soviet science instruction “is a challenge that President Eisenhower has said America must meet to survive in the Space Age—an era of danger, of challenge, of opportunity.”

In fact, within a year, Eisenhower did sign a bill into law that changed the locus of control for education reform to the federal level. Many historians portray the era that followed as a direct, knee-jerk response to the first Sputnik launch. In fact, during each of the three previous years, the United States Senate had passed bills that provided federal funding for education, but

---


none of them had survived the House of Representatives. On the day of the Sputnik launch, Senator Stewart McClure of the Education and Labor Committee drafted a memo the committee chairman, Lister Hill. Hill, an Alabama Democrat and a “savvy legislative tactician,” seized upon the idea, and the National Defense Education Act (NDEA) quickly reached the president’s desk.

As he signed it into law on September 2, 1958, President Dwight D. Eisenhower proclaimed that “this Act, which is an emergency undertaking to be terminated after four years, will in that time do much to strengthen our American system of education so that it can meet the broad and increasing demands imposed upon it by considerations of basic national security.” He added, however, that “much remains to be done to bring American education to levels consistent with the needs of our society.”

In the Declaration of Policy included in the preface of the NDEA document, Congress announced that “the security of the nation requires the fullest development of the mental resources and technical skills of its young men and women. The present emergency demands

---


that additional and more adequate educational opportunities be made available.”

To accomplish this, Title I of the NDEA stated that “we must increase our efforts to identify and educate more of the talent of our Nation. This requires programs that will give assurance that no student of ability will be denied an opportunity for higher education.” More specifically, Title I called for the United States to “correct as rapidly as possible the existing imbalances in our educational programs which have led to an insufficient proportion of our population educated in science, mathematics, and modern foreign languages and trained in technology.”

The NDEA provided federal funds to colleges and their students, and as a result, “the education boom redoubled with a sense of almost wartime urgency.” Just as the SAT helped Casper Weinberger on his career path, the NDEA also provided opportunities through college loans and fellowships for many accomplished Americans. For example, Duffy White, a professor of Russian literature at Wesleyan University and past NDEA funding recipient, declared, “Sputnik changed my life.” Also included among NDEA alumni is the infamous Unabomber, Theodore Kaczynski.

---


The law also targeted America’s public high schools to better meet “the strategic and economic needs of cold war competition.” From the end of World War II until Sputnik, the emphasis in American public high schools had centered around the life-adjustment philosophy, which expanded the range of offerings to meet students’ needs. Schools paid scant attention to those in the top 20 percent and bottom 20 percent and instead focused on the 60 percent of students in the middle of the achievement and aptitude bell curve. High school instruction needed to change for this critical mass of students who, proponents believed, “did not benefit from either college preparation or vocational training.” During the life adjustment education era, the curriculum devoted considerable time to courses centered on “hygiene, family living, drivers’ education, and social relations with peers,” or to quote educational historian Joel Spring, life adjustment education taught students “how to deal with their zits and whether or not to kiss on the first date.” In his autobiography, literary agent David Obst recalled how he thought during this era that “school was just fine. I liked my teachers, and the homework wasn’t bad.”


Sputnik, however, “spurred many Americans to question whether the battlefield victories in World War II were sufficient for America to win the peace that followed.” In the months between the Soviet satellite launch and the passage of NDEA, secondary education in the United States was the target of media scrutiny, especially as compared to the Soviet Union. Life Magazine ran an “urgent” series of articles. The first article began with a photo essay depicting the lives of one Soviet student, Alexei Kutzkov, and one American, Stephen Lapekas. The photos of American students depicted them dancing, laughing at Stephen making errors in geometry class, and surreptitiously reading magazines during English class. Alexei and his classmates, on the other hand, answered questions in a technical class, wrote lab reports, and read complex literature in two languages. Perhaps the most marked of the contrasts was between two science photos, both of which centered on groups of students looking at a glass sphere. In the Russian photo, solemn students peered inquisitively at a radio tube, while the photo of amused American students shows that a “biology experiment of dead guinea pigs momentarily diverts Stephen and others.”

Life series author Sloan Wilson neatly summarized the “salient points of the crisis.” In America, she opined, “there is no general agreement on what the schools should teach.”

---


359 Sloan Wilson, Crisis in Education, Life, March 2, 1958, at 25.

360 Sloan Wilson, Crisis in Education, Life, March 2, 1958, at 25.


362 Sloan Wilson, Crisis in Education, Life, March 2, 1958, at 32.

363 Sloan Wilson, Crisis in Education, Life, March 2, 1958, at 25.

364 Sloan Wilson, Crisis in Education, Life, March 2, 1958, at 25.
instead having squandered twenty-five years “squabbling over whether to make a child well-adjusted or to teach him something.”  

In response to this pedagogical vacuum, Wilson felt, secondary education in the United States had “gone wild with elective courses” thanks to their “eagerness to be all things to all children,” and in the end focused on a program that would “build up the bodies with in-school lunches and let the minds shift for themselves.” In addition to curricular chaos, Wilson found “most appalling, the standards of education are shockingly low.” As if to illustrate this, the pages that followed included sketches of students physically hunching over as they receive their diplomas and another of a lone studious boy reading while several other pupils in pointy dunce caps leer at him because “an eager student is branded a queer duck.”

As a result of Sputnik and NDEA, K-12 education changed, including a national emphasis on “higher standards of schooling overall.” After Sputnik, according to the autobiography of literary agent David Obst, his teacher read the class a newspaper article about the Russian education system, describing how Soviet students attended classes six days a week and were assigned four hours of homework each day. He and his classmates were “given to

366 Sloan Wilson, *Crisis in Education*, Life, March 2, 1958, at 32.
367 Sloan Wilson, *Crisis in Education*, Life, March 2, 1958, at 32.
368 Sloan Wilson, *Crisis in Education*, Life, March 2, 1958, at 32.
369 Sloan Wilson, *Crisis in Education*, Life, March 2, 1958, at 32.
370 Sloan Wilson, *Crisis in Education*, Life, March 2, 1958, at 32.
believe that the only way that America was to prevail was by loading on the homework,”373 and as a result, they “became staunchly anti-communist.”374

In addition to higher overall workload and a shift from rote memorization to tasks of higher cognitive levels,375 certain curricular areas received greater emphasis under the NDEA. Traditional classes in classical languages such as Latin and Greek were supplanted by modern languages, fluency in which was developed in language labs.376 Not only were languages taught, but the NDEA also provided funds to study how best to teach “other fields needed to provide a full understanding of the areas, regions, or countries in which such languages are commonly used.”377 Science and math were also subjects of primary interest under the NDEA. For both elementary and high schools, Section 308 provided funds for “acquisition of laboratory and other special equipment, including audio-visual materials and equipment and printed materials (other than textbooks)”378 as well as “minor remodeling of laboratory or other space.”379

The NDEA also responded to criticism of the life-adjustment education philosophy and in particular its neglect of the top 20 percent of the student population, as depicted in the scathing portrait issued by Wilson in *Life*: “The nation’s stupid children get far better care than the bright.

377 National Defense Education Act, 16 U.S.C §602 (1958),
378 National Defense Education Act, 16 U.S.C §308 (1958),
379 National Defense Education Act, 16 U.S.C §308 (1958),
The geniuses of the next decades are even now being allowed to slip back into mediocrity.”

Section 503 required states who requested funding to “set forth a program for testing students in the public secondary schools, and if authorized by law, in other secondary schools,” to “identify students with outstanding aptitudes and ability.” To make use of the results of the increased aptitude testing, the same section of NDEA also funded a program of guidance and counseling in the public secondary schools…to advise students of courses of study best suited to their ability, aptitudes, and skills, and to encourage students with outstanding aptitudes and ability to complete their secondary school education, take the necessary courses for admission to institutions of higher education, and enter such institutions.

The NDEA authorized the release of $15 million per year in Title V funds for increasing guidance and testing. Over the four years in which the funds were provided to states, the number of K-12 students to whom an aptitude test was administered increased from one-third to nearly all pupils, which represented an increase in administered tests from 10 million to 45 million. In addition to testing funds being given to state and local entities, the NDEA also provided the impetus and funding for Project Talent, a limited but precedent-setting national assessment. While it was not universal, Project Talent “administered tests to a large national

---


sample of the nation’s secondary school students.” The Project Talent assessment required students to sit for three days of administration.

During the Title V era, the United States experienced increased graduation rates, and high school diplomas were not the only result. College enrollment doubled. In addition to the immediate impact, NDEA introduced the idea that “substantial reform of education had to occur at the national level and be funded by the federal government.”

Mid 1950s to Early 1960s: Technology, Teaching, & Testing Intersect

In the meantime, back in Iowa, new technology—more successful than the Vanguard rocket—had been launched by E.F. Lindquist. He and his University of Iowa colleagues had invented the first optical scanner for scoring tests, filing for a patent in 1955. “Obviously,” wrote Lindquist in the patent application, “the problem of merely scoring the tests for thousands of pupils wherein each test can involve, say one thousand answers, is a staggering problem.”


Not only that, posited Lindquist, but “the fact is that the problem is more than one of merely scoring the test because, in the field of education, the so-called raw scores have to be converted,”\(^3\) and “converted scores must frequently be combined into weighted linear composites”\(^4\) so as to “obtain meaningful and useful information from an educational statistical standpoint.”\(^5\) “Furthermore,” argued Lindquist, educational testing organizations needed to generate “summary data for groups of answer sheets,”\(^6\) “roster type reports giving the names of examinees and their scores,”\(^7\) and “report cards for individual examinees.”\(^8\) Lindquist bemoaned the fact that at that time, the post-scoring conversions and reporting relied “primarily on manual labor.”\(^9\)

Until that time, the gold standard in automated test scoring had been the 805 Test Scoring Machine\(^0\) invented in 1934 by International Business Machines (IBM) in conjunction with high school physics teacher Reynold Johnson,\(^1\) who designed a device “in his own basement”\(^2\) in Ironwood, Michigan. Johnson was inspired by childhood memories of a prank he would play on

---

\(^3\) U.S. Patent No. 3,050,248 col.1 l.27 (filed September 21, 1955).


his sisters’ dates by drawing on their car engines with a pencil. This caused the sparks from the spark plug to travel toward the pencil marks, preventing the engine from starting. His device used wire brushes to read the amount of electricity running through graphite pencil marks, a process that proved more efficient than hand-grading. Johnson was hired as a senior engineer by IBM, and the 805 was born.

The 805 was more efficient than hand grading and was soon used for tests such as the New York Regent’s Exam and the SAT. Its separate answer cards also allowed early state testing programs such as one in Connecticut to stay within budget by having students re-use test booklets. The IBM device had its limitations, however. Each student answer card had to be hand-fed into the machine by an operator, who also had to press a knob and hand-record the total number of correct answers that appeared on a dial. After those steps, this raw score had to be hand-converted to a standard or percentile score before it could serve as a useful educational data point.

Using this type of technology was no longer feasible for the Iowa program, however. In 1942, the attention redirected to World War II had enabled Lindquist and colleagues to “drop the ‘Brain Derby’ with minimal protest from the schools.” The instrument used for the Brain Derby was replaced by the Iowa Test of Educational Development (ITED) in 1947. As compared

---

403 Anya Kamenetz, The Test: Why our Schools are Obsessed with Standardized Testing—but You Don’t Have to Be 59 (Public Affairs 2015).


to the earlier high school test, the ITED placed greater emphasis on “interpretation of materials,”408 which included the ability to comprehend text, think critically, and solve problems in broad subject areas such as science and social studies.409 As Lindquist wrote decades later, “We had hardly begun the ITED program when a young publishing house in Chicago, Science Research Associates (SRA), became keenly interested in it and asked for an opportunity to publish the ITED outside of Iowa.”410 Rather than hand the test over carte blanche to SRA, however, Lindquist’s desire was to “keep the nationwide use of ITED similar to its use in Iowa—that is, to have the publisher sell a complete package, including the tests, scoring services,”411 and reporting, which was “regarded as a rather radical innovation in test publishing at the time.”412

SRA agreed to Lindquist’s ground-breaking proposal, but this solution generated a new problem for the Iowa program.413 Lindquist was once described by a colleague as “a masterful organizer of human resources,”414 who during the first years of the Iowa Testing Program “developed highly efficient test scoring and processing procedures using only a corps of unskilled, part-time workers.”415 By the time the ITED was poised to be used on a national

scale, however, the University of Iowa “no longer had on campus the large number of returning
war veterans whose wives had earlier constituted a major source of good clerical help”\textsuperscript{416} and
thus found it difficult to secure the number of staff needed to score large-scale assessments\textsuperscript{417}
using the scanners available from IBM at that time. As is often the case, this conundrum led to
innovation. The same colleague who described Lindquist’s acumen in organizing human workers
to score tests noted that, “in the development of the test-processing equipment,”\textsuperscript{418} Lindquist
“demonstrated a remarkably creative talent for electronic and mechanical engineering.”\textsuperscript{419} This
was soon readily apparent, as Lindquist launched a scanner of his own.

In contrast to the IBM models, the Lindquist scanner relied on an optical reader rather
than detecting electrical impulses emitted by the granite in pencil lead. This new technology
allowed the sheets to automatically feed into the machine at a high rate of speed, which in turn
greatly reduced the cost and turnaround time previously involved with scoring large-scale
multiple choice assessments.\textsuperscript{420} Not only did Lindquist’s new scanning machine generate raw
scores far more rapidly, but, per the inventor’s own words in the patent application, “The
capabilities and apparatus according to the present invention are such that many more
converting, analyzing, and reporting operations can be performed on the raw score data without

\textsuperscript{416} E.F. Lindquist, \textit{The Iowa Testing Programs – A Retrospective View}, 91 Education 14 (1970).

\textsuperscript{417} E.F. Lindquist, \textit{The Iowa Testing Programs – A Retrospective View}, 91 Education 14 (1970).

\textsuperscript{418} Leonard S. Feldt, \textit{Evert F. Lindquist 1901 – 1978: A Retrospective Review of His Contributions to Educational Re}

\textsuperscript{419} Leonard S. Feldt, \textit{Evert F. Lindquist 1901 – 1978: A Retrospective Review of His Contributions to Educational Re}

\textsuperscript{420} George Madaus, Michael Russell, and Jennifer Higgins, The Paradoxes of High Stakes Testing: How They
Affect Students, Their Parents, Teachers, Principals, Schools, and Society 129 (Information Age Publishing 2009).
sacrificing, to any appreciable extent, the speed of completing the desired reports.”

These functions of the new device would generate a significant efficiency by eliminating the need for testing companies to recruit and train large groups of new staff during each year’s testing windows. The Iowa program continued to shrink and increase the speed of its scanners, soon making it feasible to administer assessments on a national scale.

Lindquist would soon expand the use of this scoring technology beyond the Iowa Test of Basic skills to just such a national test—also of his own invention. That year, Lindquist, “a recognized leader in the field of testing,” partnered with a colleague, University of Iowa Admissions Director Ted McCarrel, who “knew what kind of information colleges needed to make effective admission and placement decisions,” to create a new national admissions test. The first American College Test (ACT) was administered on November 9, 1959, to 75,000 students, and the SAT had viable competition in college entrance exam market.

This event was foreshadowed by Lindquist almost exactly a year earlier in a paper he authored. Ironically, Lindquist presented the paper during a conference organized by none other


422 U.S. Patent No. 3,050,248 col.2 l.23  (filed September 21, 1955).


than ACT’s competitor ETS and its chairman Henry Chauncey. On November 1, 1958, the nation was still reeling from Sputnik and poring over the aforementioned *Life* magazine series comparing the Russian and American education systems. Lindquist acknowledged this, positing, “If we are to meet the Russian challenge, we must, among other things, find more effective ways of motivating our students, particularly our most talented students, or of inducing them to work harder, both in and out of school, at the task of self-improvement.” Assessments, Lindquist felt, could play a crucial role in this, but only if a new college entrance exam was created.

“To appreciate fully what kind of problem we are here considering,” Lindquist opined, it was critical that the leaders of American testing give careful consideration to the “general purpose or requirement of college entrance and scholarship examinations.” This echoes Lindquist’s words eight years earlier, in a book he edited—and to which Henry Chauncey contributed—that was considered to be the definitive text on educational assessment at the time. “It should be apparent that the decisions which are made preliminary to actual test construction

---


are, from the broadest point of view, far more important or crucial than those which follow.”

This preliminary work, per Lindquist, needed to include a clear vision of not only the academic—but also the human—subjects involved.

The ACT, Lindquist later recalled, “was designed the needs of the large state universities, state and municipal colleges and junior colleges, and the great majority of the smaller private and denominational colleges of the country—generally, the institutions that were not already served by the College Entrance Examination Board.” Prior to this time, some of these schools had unsuccessfully attempted to use the results of state-specific assessments and administer their own tests, but were unsuccessful due to issues including consistency, manpower, and timing. This problem was exacerbated by the influx of new college students generated by the G.I. Bill and Sputnik.

ETS had attempted to fill this gap by experimenting with a new instrument, the Test of Developed Ability (TDA), which was designed to provide more subject-specific data. This, in turn, could have allowed ETS to expand beyond the East Coast and appease “people in higher education, especially the state universities,” who “grumbled about how Conant and his allies

---


had hijacked educational testing and set up a national IQ operation to feed the Ivy League.”

The TDA was declared a failure, however, after only two administrations. Compared to the SAT, the TDA was “longer and therefore more complicated to administer, and because of the essay section, it could not be entirely machine-scored.” The TDA was also double the per-student cost of the SAT.

This left the aforementioned state and smaller private colleges in need “not for a test that would skim the cream off the top of a distribution of applicants, but rather for one that would help screen out the few who might least profit from college opportunities.” To avoid the mistakes of the TDA, Lindquist knew “the cooperation of the colleges of my own state and of college admissions officers everywhere was essential.” To this end, Lindquist recruited Ted McCarrell, who not only served as the registrar and admissions officer at the University of Iowa but “was very influential among college admissions officers, having just finished a term as president of National Association of College Admissions Officers.”

Together, Lindquist and McCarrell followed the blueprint for test development laid out by the former in his chapter of the aforementioned book. In Lindquist’s opinion, any list of steps taken to construct a test “takes for granted that certain preliminary decisions have already been

---

made concerning the general nature and purposes of the test to be constructed.\textsuperscript{447} Lindquist later described how what colleges needed from the ACT was “a test battery that would describe the student and define his educational needs and abilities in meaningful differential terms, terms related to different areas of instruction.”\textsuperscript{448}

In addition to reflecting the students’ diligent work and informing college admission decisions, Lindquist felt that a college entrance examination should also “make the high schools more keenly aware of their own responsibilities and shortcomings”\textsuperscript{449} while simultaneously giving them “positive aid in meeting these responsibilities by drawing attention to broad areas or aspects of achievement most in need of improvement.”\textsuperscript{450} Lest any participants leave the conference ambiguous about whether Lindquist intended to criticize his hosts and their SAT exam, he provided clarity. “I need hardly point out that college entrance examinations of the type generally regarded as intelligence tests or scholastic aptitude tests, or differential aptitude tests, are almost wholly useless for these purposes, as they are for motivating the individual student.”\textsuperscript{451} This caused consternation at the headquarters of the SAT parent company ETS.

\textsuperscript{447} E.F. Lindquist, Preliminary Considerations I Objective Test Construction, E.F. Lindquist, et. al, Educational Measurement 120, (1951), https://archive.org/stream/educationalmeasu00lind#page/120/mode/2up https://archive.org/stream/educationalmeasu00lind#page/120/mode/2up.

\textsuperscript{448} E.F. Lindquist, The Iowa Testing Programs – A Retrospective View, 91 Education 17 (1970).


“Up to the present year,” noted an internal report, “ETS has enjoyed what amounts practically to a monopoly in college admissions testing.”

After collaborating with him on the TDA, the aforementioned book and conference, and other projects, one might reasonably ask at this point why Chauncey did not simply hire Lindquist to work for ETS. Rather than seeing his time on the standing committee of the College Board for the SAT as a logical springboard to employment by ETS, Lindquist instead later described how his position meant he “had special occasion and opportunity to think about alternatives to the improvements upon the Board’s program.” To preserve the aforementioned ETS stranglehold on large-scale college entrance testing, Chauncey had attempted to recruit Lindquist to the position of vice president for programs shortly after ETS was founded. Lindquist declined, perhaps because he “thought himself as being far above Chauncey in professional eminence, and at the same time, he played the Midwestern populist to Chauncey’s Eastern elitist.”

To distinguish his Midwestern exam from the East Coast’s SAT, Lindquist “broke new ground” when designing the ACT by “focusing not on cognitive reasoning, but instead on the information taught in schools.” He also heeded his own earlier advice and carefully considered the potential purpose and participants for his new test. Lindquist and McCarrell


believed that “America’s students and colleges could best be served by an achievement test that measured what students had learned, rather than an aptitude test measuring student potential.”

“If the examination is to have the maximum motivating value for the high school student, it must impress upon him the fact that his chances of being admitted to college, or of being awarded a scholarship, depend not only on his ‘brightness’ or ‘intelligence’ or other innate qualities.”

Instead a college placement test should assess

how hard he [a student] has worked at the task of getting ready for college, both in high school and in the years preceding high school. The examination must make him feel that he has earned the right to go to college by his own efforts, not that he is entitled to college admission because of his innate abilities or aptitudes.

To ensure the correlation to achievement, the ACT was “tied more closely than the SAT to high-school curricula.” Since Lindquist already had in his possession a bank of tests from the ITED about which “a great mass of evidence of validity” had already been compiled, he adapted four of them as a base for the first edition of the ACT. As one would have expected, the ACT covered English, mathematics, and reading. Unlike the SAT, however, in a move
that reflected his understanding of the impact of Sputnik, Lindquist also made the ACT “the first and only college entrance exam to include a science assessment.”

As a colleague once described, Lindquist felt too many students “tended to take a passive attitude toward the learning process,” and he felt assessments could play a role in raising the cognitive rigor of instruction. Thus, “as against those who would caricature such tests as measuring only rote recall of facts,” Lindquist insisted on assessments that measured students’ higher-level thinking skills, with the caveat that they be measured in the context of the curriculum. To this end, the ITED, on which parts of the first ACT were based, was “concerned with interpretive and problem-solving abilities in broad areas,” including social studies and science.

The ACT was not merely a renaming of the ITED, however. To distinguish the ACT from not only the ITED but the SAT, Lindquist also specified that “the real need was not for an efficient predictor of a single doubtful grade-point-average criterion.” He felt “each institution needed to know everything about the student that is significant and relevant to his college

---


Ergo, Lindquist and Carrell designed the ACT to include a student survey component that included “his past achievements, his interests, his educational and vocational aspirations and so forth.”

Mid 1960s: Era of Equity

Other aspirations—those of a third grade girl and her family that came to symbolize civil rights in education—had also been reshaping the educational landscape. *Brown v. Board of Education of Topeka* was indirectly initiated by the National Association for the Advancement of Colored People (NAACP). Future Supreme Court Justice Thurgood Marshall followed his mentor, Charles Hamilton Huston, as special counsel to the NAACP. The organization soon adopted Marshall’s strategy to “devote its efforts solely to an all-out attack on segregation in education, rather than pressing for the equalization of segregated facilities.”

The school district in Topeka, Kansas housed just such segregated educational facilities. For example, in spite of the fact that her house was in a diverse neighborhood and stood only a four-block walk from Sumner School, Linda Brown was compelled by segregation practices to traverse a set of railroad tracks and take a bus to an all-black school. In 1950, the NAACP

---


asked her father, Oliver Brown, to join others in attempting to enroll their children in all-white schools, expecting them to be rebuffed, which is exactly what transpired.476

Linda Brown and her complaint against the Topeka, Kansas Board of Education also came to represent four other cases from various states: Briggs v. Elliot, Davis v. Board of Education of Prince Edward County, Boiling v. Sharpe, and Gebhart v. Ethel. All were granted certiorari as a single case by the Supreme Court.477 While the facts differed among the cases, they addressed the same issue: Is state-sponsored segregation, as allowed by the “separate but equal”478 decision in Plessy v. Ferguson, constitutional?479 On May 17, 1954, the court held in the field of public education, the doctrine of ‘separate but equal’ has no place. Separate educational facilities are inherently unequal. Therefore, we hold that the plaintiffs and others similarly situated for whom the actions have been brought are, by reason of the segregation complained of, deprived of the equal protection of the laws guaranteed by the Fourteenth Amendment. This disposition makes unnecessary any discussion whether such segregation also violates the Due Process Clause of the Fourteenth Amendment.480

“Such segregation,”481 however, persisted. Due to “the complexities arising from the transition to a system of public education freed of racial discrimination,”482 the defendant school districts from Brown sued for relief, and the case, which is often described as Brown II, was heard by the Supreme Court. The Court recognized that desegregation required the “solution of


478 Plessy v. Ferguson, 163 U.S. 537 (1896).

479 Plessy v. Ferguson, 163 U.S. 537 (1896).


varied local school problems”\footnote{Brown \textit{v. Bd. of Educ.}, 349 U.S. 294, 301 (1955).} and “school authorities have the primary responsibility for elucidating, assessing, and solving these problems.”\footnote{Brown \textit{v. Bd. of Educ.}, 349 U.S. 294, 301 (1955).} The justices remanded the cases to the original courts, charging them with determining “whether the action of school authorities constitutes good faith implementation of the governing constitutional principles.”\footnote{Brown \textit{v. Bd. of Educ.}, 349 U.S. 294, 301 (1955).} While it released responsibility back to the original courts “because of their proximity to local conditions and the possible need for further hearings,”\footnote{Brown \textit{v. Bd. of Educ.}, 349 U.S. 294, 301 (1955).} in 1955, the court ordered the District Courts to “take such proceedings and enter such orders and decrees consistent with this opinion as are necessary and proper to admit to public schools on a racially nondiscriminatory basis with \textit{all deliberate speed} the parties to these cases.”\footnote{Brown \textit{v. Bd. of Educ.}, 349 U.S. 294, 301 (1955).}

In spite of the precedent of federal involvement in education reform set by the NDEA,\footnote{David A. Gamson, Kathryn A. McDermott, and Douglas S. Reed, \textit{The Elementary and Secondary Education Act at Fifty: Aspirations, Effects, and Limitations}, 1 The Russell Sage Foundation Journal 8 (2015), \url{http://www.rsjournal.org/doi/pdf/10.7758/RSF.2015.1.3.01}.} progress toward desegregation and equity was slow during the years following Brown. A decade after the landmark case, only 2.3 percent of black students in the South attended a majority white school.\footnote{Erika Frankenberg and Kendra Taylor, \textit{ESEA and the Civil Rights Act: An Interbranch Approach to Furthering Desegregation}, 1 The Russell Sage Foundation Journal 33 (2015), \url{http://www.rsjournal.org/doi/full/10.7758/RSF.2015.1.3.01}.} This led some to believe that “segregation would not be eradicated
solely through the judiciary.” 490 However, as of the beginning of 1963, little traction on further education reform was gained in the executive or legislative branches of the federal government. 491

Two events not directly related to education dramatically changed the federal landscape and status of education reform at the national level. 492 In May 1963, Dr. Martin Luther King and the Southern Christian Leadership Conference dispatched over 1,000 black youth, aged six to eighteen, to march through Birmingham, Alabama, in a call for an end to the segregation of local businesses. The march triggered a response from the local police, led by chief Eugene “Bull” Connor. Images of under-age protestors being sprayed with high-pressure fire hoses and attacked by police dogs 493 were shared with a broad audience via television and news outlets across the nation. 494 Following Birmingham, public opinion polls showed that civil rights was now considered “a pressing item on the national agenda.” 495

---


491 Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 31 (Cornell University Press 2012).


A few months later, on November 22, 1963, President John F. Kennedy was assassinated by gunshot while riding in an open motorcade as part of a visit to Dallas, Texas. Although the only live motion-picture account of the shooting, 26 seconds of amateur eight-millimeter film shot by Dallas dressmaker Abraham Zapruder, was entrusted to Life Magazine “to be protected out of respect for the Kennedys” and was not broadcast until 1975, television news outlets offered unprecedented coverage. “Once people heard this had happened, they were glued to their televisions” and “within an hour of the shooting, 68 percent of Americans heard the news; within two hours, 92 percent had heard.” Coverage continued, comprising a “four-day long drama that played out on national television,” the end result of which included increased public support for Democratic agendas.

---


One item on the Democratic agenda that had been introduced in a June 1963 speech by John F. Kennedy but was enacted only after the wave of support which followed his assassination was the Civil Rights Act. “The events in Birmingham and elsewhere have so increased the desires for equality that no city or state or legislative body can prudently choose to ignore them. The fires of frustration and discord are burning in every city, North and South, where legal remedies are not at hand,” Kennedy had declared. Yet it was not until the wave of popular support following his death that the Civil Rights Act was enacted on July 2, 1964.

Titles IV and VI of the Act pertained to education. While it is rightfully remembered by many for ordering desegregation or “assignment of students to public schools and within such schools without regard to their race, or national origin,” the Act also quietly laid the groundwork for additional federal involvement in education reform and standardized testing, which has persisted into the 21st century in an often-forgotten section of the law.

Title IV of the Civil Rights Act charged the newly-created Civil Rights Commission with conducting a survey of “the lack of availability of equal educational opportunities for individuals by reason of race, color, religion, or national origin in public educational institutions at all levels

---


in the United States” and reporting its results to the president and Congress. Like the Civil Rights Act itself, this idea had its genesis during the John F. Kennedy administration. Francis Keppel was appointed to lead the Office of Education by President Kennedy in 1962. While researching his new role, Keppel uncovered an 1867 law that required the Office of Education to “report annually on the progress of students in the United States” and “marveled at the fact that, in nearly a century, the Office had never done so.”

In 1965, the Office of Education awarded the contract for the “Survey and Report of Educational Opportunities” to ETS, the testing arm of the College Board, which remained under the leadership of Henry Chauncey. The involvement of ETS in the Equality of Educational Opportunity Survey (EEOS) was not widely known. Instead, the resulting report was named for Harvard professor James S. Coleman, who was on sabbatical in

---


Germany during the time the survey was conducted and results were compiled.\textsuperscript{517} For the
designed Robert Solomon as the lead, along with technical expert
also advocated for “a test of skills that you couldn’t connect
in any way, shape, or form to IQ.”\textsuperscript{519} The EEOS called for such tests, along with questionnaires
for principals, teachers and students to afford the opportunity for “investigation of correlates of
students achievement”\textsuperscript{520} to be administered to one million students across the United States in
grades 1, 3, 6, 9, and 12.\textsuperscript{521}

The ability of ETS to use pre-existing achievement tests, thereby eliminating the need for
field testing of items, was a strength of the study. However, participation was voluntary, leading
to issues with sample size. The EEOS was administered to only 65 percent of the original one
million projected sample due to the inability or unwillingness of some\textsuperscript{522}—including all of the
Chicago Public Schools—to participate,\textsuperscript{523} a result that would not be considered acceptable in a
current government study. Technology in the mid-1960s made generating measurement error

\begin{itemize}
\item \textsuperscript{517} Nicholas Lemann, The Big Test: The Secret of the American Meritocracy 160 (Farrar, Straus, and Giroux 2000).
\item \textsuperscript{518} Nicholas Lemann, The Big Test: The Secret of the American Meritocracy 159 (Farrar, Straus, and Giroux 2000).
\item \textsuperscript{519} Nicholas Lemann, The Big Test: The Secret of the American Meritocracy 161 (Farrar, Straus, and Giroux 2000).
\item \textsuperscript{523} Nicholas Lemann, The Big Test: The Secret of the American Meritocracy 160 (Farrar, Straus, and Giroux 2000).
\end{itemize}
difficult as well.\textsuperscript{524} The Civil Rights Act required that a report of the study be issued “within two years of the enactment of this title,”\textsuperscript{525} granting Solomon and his team just over one year to complete this ambitious task by the July 1, 1966, deadline.

After the passage of the Civil Rights Act, there continued to be disagreement over the role of the federal government in education.\textsuperscript{526} Some felt the federal government should completely avoid any involvement in “the sanctity of the state-local-private preserve of education”\textsuperscript{527} because it would pit “Democrat against Republican, liberal against conservative, Catholic against Protestant and Jew, federal power against states’ rights, white against black, and rich constituency against poor.”\textsuperscript{528} Others agreed on a federal role but posed a question about the nature of education reform: “should reformers push for academic excellence, as NDEA tended to do, or should they insist on equal educational opportunity, a goal that had been a tenet of American schooling since the common school reform era?”\textsuperscript{529}

\textsuperscript{524} Nicholas Lemann, The Big Test: The Secret of the American Meritocracy 160 (Farrar, Straus, and Giroux 2000).


\textsuperscript{527} Patrick J. McGuinn, No Child Left Behind and the Transformation of Federal Education Policy 28 (University Press of Kansas 2006).

\textsuperscript{528} Patrick J. McGuinn, No Child Left Behind and the Transformation of Federal Education Policy 28 (University Press of Kansas 2006).

In 1964, John Gardner, president of the Carnegie Corporation, a close friend and sometime critic of Henry Chauncey and an adviser to Presidents Eisenhower and Kennedy, argued that the aforementioned goals of education reform were not mutually exclusive. Gardner published his claims in his book *Excellence: Can We Be Equal and Excellent Too?* Not only did Gardner argue that excellence and equity were not mutually exclusive, but rather that, at least in the United States, the latter could not exist without the former.

It must never be forgotten that ours is one of the few societies in the history of the world in which performance is a primary determinant of status. What the individual can ‘deliver’ in the way of performance is a major factor in how far he can rise in the world.

Gardner also preemptively refuted the claim that schools could not help mitigate for factors in students’ homes. Gardner struck a balance, purporting that schools could help mitigate for the effects of an impoverished home environment, while still acknowledging their significance:

We now know beyond any doubt that the social and cultural influences of home have a good deal to do with both motivation and performance in school. The child growing up in

---


534 John W. Gardner, *Can We Be Equal and Excellent, Too?*, 68 (Harper & Brothers 1961), [https://archive.org/stream/excellencecanweb00ingard#page/76/mode/2up](https://archive.org/stream/excellencecanweb00ingard#page/76/mode/2up).
a home barren of educational or cultural influences may require a longer exposure to
school before he wakes up.  

Lyndon B. Johnson continued to benefit from both John F. and Robert F. Kennedy’s
posthumous popularity, winning the 1964 presidential election with 61.1 percent of the
popular vote against the “spectacularly unsuccessful” campaign run by Republican Barry
Goldwater. Johnson soon selected Gardner to serve as his Secretary of Health, Education, and
Welfare. Per Gardner’s obituary, the president “was so impressed by the way Mr. Gardner
handled the 1965 White House Conference on Education,” that Johnson asked Gardner join
his cabinet. United by their “almost obsessive interest in education,” the Republican
Gardner and Democrat Johnson collaborated on an unprecedented volume of federal education
legislation.

535 John W. Gardner, Can We Be Equal and Excellent, Too?, 70 (Harper & Brothers 1961),
https://archive.org/stream/excellencecanweb00ingard#page/76/mode/2up.

536 Patrick J. McGuinn, No Child Left Behind and the Transformation of Federal Education Policy 29 (University
Press of Kansas 2006).

537 Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 31 (Cornell
University Press 2012).


540 Robert D. McFadden, John E. Gardner, 89, Founder of Common Cause and Adviser to Presidents, Dies, New
cause-and-adviser-to-presidents-dies.html.

541 Robert D. McFadden, John E. Gardner, 89, Founder of Common Cause and Adviser to Presidents, Dies, New
cause-and-adviser-to-presidents-dies.html.

542 Robert D. McFadden, John E. Gardner, 89, Founder of Common Cause and Adviser to Presidents, Dies, New
cause-and-adviser-to-presidents-dies.html.

543 Lyndon B. Johnson National Historic Park Texas: Junction School, National Park Service,
In Johnson’s case, the strong interest in education likely had deep roots in his childhood and early career. In 1912, he enrolled early—at age four—at Junction School. Junction was a rural one-room schoolhouse heated by a wooden stove, lit by kerosene lamps, and ruled by Miss Katie Deadrich, the sole teacher for grades one through eight. Johnson’s family relocated to Johnson City, Texas during his school years, and he went on to attend Southwest Teachers’ College in San Marcos, Texas. After graduating from San Marcos at the age of 20, Johnson taught at a high-poverty majority-Hispanic school in Cotulla, Texas. He used part of his salary to purchase supplies and equipment for students, who he had witnessed “going through a garbage pile, shaking the coffee grounds from the grapefruit rinds and sucking the rinds for the juice that was left.”

In the presence of dignitaries such as his “old friend” Senator Eugene McCarthy, Johnson sat next to his first teacher Miss Deadrich at a picnic table in front of Junction School.

---


on April 11, 1965,\textsuperscript{550} and delivered a speech prior to signing the Elementary and Secondary Education Act (ESEA). Johnson hoped ESEA would “bridge the gap for more than five million educationally deprived children,”\textsuperscript{551} quoting Thomas Jefferson when he said, “Preach, my dear sir, a crusade against ignorance establish the law for educating the common people.”\textsuperscript{552} In addition to the carefully chosen setting for the signing, Johnson also invoked his own past during his speech.

As the son of a tenant farmer, I know that education is the only valid passport from poverty. As a former teacher—and, I hope, a future one—I have great expectations of what this law will mean for all of our young people. As President of the United States, I believe deeply no law I have signed or will ever sign will mean more for all of our young people.\textsuperscript{553}

Johnson perceived the ESEA as necessary to fulfilling the mandate of the Civil Rights Act he had signed the previous year.\textsuperscript{554} As the Coleman Report would later demonstrate,\textsuperscript{555} “President Johnson believed that issues of educational inequality were intimately linked to race

\begin{flushleft}
\textsuperscript{554}Patrick J. McGuinn, No Child Left Behind and the Transformation of Federal Education Policy 29 (University Press of Kansas, 2006).
\end{flushleft}
and poverty.”\textsuperscript{556} Under the ESEA, “not only would racial walls come down on campuses, no longer would it be right to have major gaps between the academic achievement of white and minority students.”\textsuperscript{557} Johnson also saw the ESEA as a logical next step in his “Great Society”\textsuperscript{558} antipoverty program, through which he had “rejected an income-transfer strategy in favor of an emphasis on job training and education,”\textsuperscript{559} embodying his belief that “very often a lack of jobs and money are not the cause of poverty, but the symptom.”\textsuperscript{560}

The ESEA was by no means the first piece of federal education legislation proposed since the NDEA. Rather, it was the first to succeed. The landmark law not only followed on the heels of Johnson’s landslide presidential election victory,\textsuperscript{561} but it also managed to avoid the pitfalls described by historian Gareth Davies as the “three Rs of federal education politics: race, religion, and reds.”\textsuperscript{562} The first R, race, had been addressed by Title VI of the Civil Rights Act,\textsuperscript{563} which stated that “No person in the United States shall, on the ground of race, color, or national origin,

\begin{flushright}


\textsuperscript{558} Patrick J. McGuinn, \textit{No Child Left Behind and the Transformation of Federal Education Policy} 29 (University Press of Kansas, 2006).

\textsuperscript{559} Patrick J. McGuinn, \textit{No Child Left Behind and the Transformation of Federal Education Policy} 29 (University Press of Kansas, 2006).

\textsuperscript{560} Patrick J. McGuinn, \textit{No Child Left Behind and the Transformation of Federal Education Policy} 29 (University Press of Kansas, 2006).


\textsuperscript{562} Jesse H. Rhodes, \textit{An Education in Politics: The Origins and Evolution of No Child Left Behind} 29 (Cornell University Press 2012).

\textsuperscript{563} Jesse H. Rhodes, \textit{An Education in Politics: The Origins and Evolution of No Child Left Behind} 31 (Cornell University Press 2012).
be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving federal financial assistance.”

The second R, religion, was a less substantial issue following the recent upholding by the Supreme Court of “child benefit theory.” Child benefit theory, first introduced to education law by *Borden v. Louisiana State Board of Education* posits that public funds can be allocated directly to students in parochial schools, so long as the institutions themselves do not benefit. Francis Keppel ensured the ESEA passed the child benefit test and avoided opposition from Democrat voters who were Catholic or belonged to the National Education Association by allocating funds based on a per-pupil basis. The law itself also contained provisions to protect separation of church and state. Section 605, entitled “Limitation on Payments Under This Act,” forbade any payment via the act for religious worship or instruction.

In the context of the ESEA, the third R, reds, referred not to the fear of communist aggression evidenced by the drafting of NDEA in response to *Sputnik*, but rather to an opposition to anything that hinted at a Soviet-style centralized bureaucracy at the federal level in

---


567 Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 31 (Cornell University Press 2012).


the United States. As was the case with religious concerns, the ESEA also contained direct verbiage to address fears of federal control usurping state and local rights in the area of education. Section 604 reassured citizens that

nothing contained in this Act shall be construed to authorize any department, agency, officer, or employee of the United States to exercise any direction, supervision, or control over the curriculum, program of instruction, administration, or personnel of any educational institution or school system, or over the selection of library resources, textbooks, or other printed or published instructional materials.

The final R was also averted by use of an innovative funding strategy. ESEA monies were distributed as categorical assistance to students suffering from the effects of poverty rather than general education aid. The Johnson administration applied to ESEA the principles of impact aid, first used under the Lanham Act of 1941, that granted federal funds to districts harmed by the presence of un-taxable federal installations. Also like the Lanham Act, ESEA funds were disbursed to nearly every Congressional district in the United States.

In addition to avoiding the aforementioned pitfalls in their education bill, Democrats had also added seats in both houses of Congress during the 1964 election. These seats, added to a majority already attained in 1958, left the party with two-thirds of the vote in both the House and

\[572\] Patrick J. McGuinn, No Child Left Behind and the Transformation of Federal Education Policy, 30 (University Press of Kansas 2006).


\[574\] Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 31 (Cornell University Press 2012).

\[575\] Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 30 (Cornell University Press 2012).

\[576\] Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 31 (Cornell University Press 2012).
As a result, ESEA passed with 263 votes in the House and 73 in the Senate, and a precedent was set for a broader federal reach into education policy.

Title I of ESEA provided funds to states to disburse to local education agencies for programs or projects “designed to meet the special educational needs of educationally deprived children in school attendance areas having high concentrations of children from low-income families.” Any initiative receiving a Title I allocation was required to be “sufficient in size, scope, and quality to give reasonable promise of substantial progress toward meeting” the aforementioned special educational needs of low-income students. Among acceptable appropriations of Title I funds were “supplementary educational services,” including remedial instruction and programs outside of regular school days and hours. Title II provided monies for books, including library materials, teacher resources, and textbooks. The ESEA also included funds to train educators on what Johnson described as “new teaching techniques.”

---

577 Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 31 (Cornell University Press 2012).


The ESEA mandate of “reasonable promise of substantial progress” toward meeting students’ needs was fairly vague, a fact not lost on Senator Robert F. Kennedy, whose distrust of Johnson and the political establishment in the South was well known. Kennedy expressed his concerns on the Senate floor during debate about the ESEA:

I think it is very difficult for a person who lives in a community to know whether, in fact, his educational system is what it should be, whether if you compare his community to a neighboring community they are doing everything they should be, whether the people that are operating the educational system in a state or local community are as good as they should be.

Kennedy proposed a solution to the problem he had identified: “I wonder if we couldn’t have some system of reporting” via some testing system that would be established by which the people at the local community would know periodically what progress had been made.

The advocacy of Robert F. Kennedy was only one of the factors impacting the implementation and evaluation of ESEA. Secretary of Defense Robert McNamara and other federal government officials of the 1960s had revived the concept of accountability shared by the Progressives of the 1890s and proponents of the Scientific Efficiency movement of the early

---


592 Neil J. Salkind, Directions in Educational Psychology 286 (Sage Publications, 2011).
20th century.\textsuperscript{593} In the 1960s, McNamara and his colleagues tied accountability directly to funding and christened it “program budgeting”\textsuperscript{594} for a new era.

This new approach soon expanded from the Department of Defense to other branches of the federal government and the public sector in general.\textsuperscript{595} Management under program budgeting emphasized objectives, goals, and cost-benefit analysis.\textsuperscript{596} Demand for accountability was also felt at the state level. States were paying “the largest share of school bills in history”\textsuperscript{597} as voters rejected bond issues and other increases to local property taxes proposed by local school districts.\textsuperscript{598} These factors manifested themselves in “a call for schools to provide evidence that they were successful,”\textsuperscript{599} and that evidence was provided by scores on standardized tests.\textsuperscript{600}

To this end, Section 205 of the ESEA mandated that a “provision for appropriate objective measurements of educational achievement will be adopted for evaluating at least annually the effectiveness of the programs in meeting the special educational needs of

\textsuperscript{593} Frederick P. Taylor, Principles of Scientific Management 7 (Harper & Brothers 1911).
\textsuperscript{594} Neil J. Salkind, Directions in Educational Psychology 286 (Sage Publications 2011).
educationally deprived children.”601 To avoid the “third R” and the attendant fears of centralization,602 the provision stopped short of Kennedy’s suggestion of a national system, instead remanding control to lower levels of government. The ESEA required local school districts to submit annual reports to their states, “including information relating to the educational achievement of students participating in programs.”603 In other words, “performance data were required on students receiving assistance from these programs.”604 To accomplish this, districts “had to revitalize their old school research bureaus or create new ones.”605 To continue receiving ESEA funds, not only were districts required to collect test scores, but they also were mandated to report them to the state in a standardized format,606 along with the demographic data required to prove eligibility.607 Titles III and V of ESEA provided funds to states to coordinate data collection at the state level.608

The end result was that the ESEA set the precedent of evaluating programs using standardized assessments.609 This new requirement revealed some instances in which the

605 Daniel P. Resnick, Competency Testing Historically Considered, 8 Review of Research in Education 7 (1980).
achievement of poor and minority students had been kept from the public view. For example, the school district in Dallas had not previously been required to administer the Iowa Test of Basic Skills to minority students. Still more factors eventually triggered an additional layer of standardized testing linked to ESEA. After peaking in 1964, average SAT scores began to decline in 1965. The mid-1960s also saw the release of the results from the Sputnik-inspired Project Talent survey, after which the Coleman Report—chartered by the Civil Rights Act and implemented through data gathering via the EEOS—was also released.

Two key members of the EEOS team reviewed the assessment scores and their correlation to questionnaire responses. They drew opposite conclusions. Coleman, who reviewed the data upon his return from sabbatical, concluded that differences in students’ families and the socioeconomic composition of their schools that often reflected families’ ability to move to an area with better schools produced the most significant difference in test results for minority students. Beaton, on the other hand, found that differences among schools accounted


for more variance in achievement than differences among students’ families. Beaton’s study, however, was internal to ETS and not widely publicized.617

In spite of the aforementioned limitations and wide variance in interpretation even among those closely involved with the EEOS, raw data from the Coleman Study were made available to any analyst with technology. The Public-Use Data Tape (PUDT) offered open access to the results of the study with identifying information redacted.618 The data tapes from the Coleman Report were “widely distributed and used.”619

Fiscal expert Alice Rivlin found common ground between Coleman’s analysis of the EEOS results and those reported from Project Talent. “The most general result of these statistical studies has been the finding that variables reflecting the socioeconomic characteristics of students and their families explain most of the variation in test scores.”620 “Resource inputs,”621 Rivlin concluded, “explain very little.”622 That interpretation was not emphasized by the Johnson administration, which was depending on ESEA “to buy Southern support to end apartheid schooling in the South.”623 Democrats feared that “Republicans, already resistant to


more federal spending,” would seize onto and exploit the assertion that spending levels were clearly not the prime factor in performance discrepancies.”

Thus, the stage was set for the National Assessment of Educational Progress (NAEP). It is worth noting that the Accountability Handbook published “in the face of the burgeoning demand for formal evaluations” described the demand as “a product of the period’s dramatic expansion of social services and proposals for services” that “did not come primarily from tightfisted conservatives.” In other words, “advocates in the War on Poverty wanted to know whether all students were moving ahead.”

Francis Keppel and John Gardner were tasked with developing this new federal-level assessment. The nascent negative trends in SAT data have been cited as having “greatly influenced the design of NAEP.” Keppel and Gardner’s task was to create “a good indicator

---

627 Neil J. Salkind, Directions in Educational Psychology 286 (Sage Publications, 2011).
628 Neil J. Salkind, Directions in Educational Psychology 286 (Sage Publications, 2011).
629 Neil J. Salkind, Directions in Educational Psychology 286 (Sage Publications, 2011).
of the performance of all students, not just the college-bound."\textsuperscript{633} As a result, NAEP, as compared to other large-scale standardized tests, was designed to "reflect a degree of consensus about what students should know"\textsuperscript{634} with "unusually broad"\textsuperscript{635} coverage of the domains it measures.\textsuperscript{636} The initial administration of NAEP included citizenship, science, and writing\textsuperscript{637} and soon grew to assess ten domains,\textsuperscript{638} including reading and math.\textsuperscript{639}

The creators of NAEP were given a lofty charge to "examine achievement in ten learning areas, to spot changes in level of achievement over the years, and to apply the implication of those changes to national education policy."\textsuperscript{640} They also faced the unenviable challenge of reconciling their assignment with the need to avoid awakening the "Third R"\textsuperscript{641} by appearing to encroach on state and local rights and mandating a sweeping national assessment. To accomplish


\textsuperscript{638}Daniel P. Resnick, Competency Testing Historically Considered, 8 Review of Research in Education 8 (1980).

\textsuperscript{639}Jim Horn & Denise Wilburn, The Mismeasure of Education 123 (Information Age Publishing 2013).


\textsuperscript{641}Patrick J. McGuinn, No Child Left Behind and the Transformation of Federal Education Policy, 30 (University Press of Kansas, 2006).
all of the aforementioned, Keppel, Gardner, and their team employed a variety of innovative methods for the first NAEP administration in 1969.\textsuperscript{642}

First, a large pool of items was developed, and a sampling of students in both public and non-public schools was selected by the Research Triangle Institute (RTI), the vendor contracted to administer NAEP. To avert triggering the “third R”\textsuperscript{643} by requiring each state to enroll students in school for the first time at the same age, students were grouped for assessment by age rather than grade for the first administration of NAEP. Students aged nine, thirteen, and seventeen—including out-of-school seventeen-year-olds—were tested in 1969. (The out-of-school group was later dropped due to the cost of administration.) For the same reason, participation in the early administrations of the NAEP was voluntary.\textsuperscript{644} Rather than assessing all students across the country, a nationally representative sample of students was tested.\textsuperscript{645} In addition to the aforementioned precautions, assessment sessions for the first administration of the NAEP were limited to one hour per student to avoid triggering fears of federal co-opting of instructional time.\textsuperscript{646}

The efficiency of measuring 10 domains in 60 minutes was made possible by employing matrix sampling. Matrix sampling assigned a group of items from each domain to each student.


\textsuperscript{643}Patrick J. McGuinn, No Child Left Behind and the Transformation of Federal Education Policy, 30 (University Press of Kansas, 2006).


In the end, each item was answered by a group of students representing the demographic composition of the sample size as a whole. Tape recordings of directions were used to ensure standardization. This in turn required that students in the same session be assigned the same booklet (and therefore, the same set of items). Because matrix sampling tested random clusters (as defined by schools) and not random students, Quenouille and Tukey’s jackknife method was applied for the first time to a large-scale assessment. The jackknife is a method of replicating data in complex sampling designs to estimate the variability of the overall sample. Due to the application of Tukey’s method, the NAEP was able to report standard errors, contributing to the perceived validity of results which was lacking in the EEOS.

Results of the first NAEP were also reported in ways that varied from standardized testing practices until that time. For the 1969 administration, results were reported by denoting the percentage of students answering each individual item correctly. In addition to all students, item data were also reported by gender, race, parental education levels, and other demographic factors. However, since this “left the reader to generalize to an entire subject area” (e.g.,

---


science) based on scores for a large set of items, an average percentage of items answered correctly for an entire domain was soon added.\textsuperscript{651}

More importantly, data reporting for NAEP broke new ground for large-scale standardized tests by breaking away from the practice of reporting norm-referenced results that assign students’ final scores based on how they compare to others taking the same assessment. To make possible its initial charge of tracking growth over time, NAEP scores were criterion (content)-referenced, making NAEP one of the first minimum-competency tests in the world of large-scale standardized assessment.\textsuperscript{652} In other words, reports centered “not on the percentile ranking of the student who had taken the test, but on the kinds of knowledge and skills that he had demonstrated.”\textsuperscript{653} In spite of the references to students, in the early days of the NAEP, results were not reported for individual students, teachers, schools, districts, or states.\textsuperscript{654} Instead score reports focused on representing the performance of students across the United States over time.\textsuperscript{655} Over time, NAEP score reports came to be popularly known as “The Nation’s Report Card.”\textsuperscript{656}

\begin{itemize}
  \item \textsuperscript{652} Daniel P. Resnick, \textit{Competency Testing Historically Considered}, 8 Review of Research in Education 8 (1980).
\end{itemize}
Ironically, as the first NAEP score trends were compiled, another set of trend data was analyzed at a deeper level. The decline of SAT scores had gained significant attention in the 1970s due to wide reporting in the media that “viewed this decline as an indicator of the deterioration of the U.S. educational system.” The College Board commissioned a blue ribbon panel to investigate. After introducing the results of Project Talent and the National Educational Longitudinal Study of 1972 to the data set, the commission found that “the decline was largely due to the steady increase in the number of students taking the SAT examination,” including many with lower verbal ability (i.e., IQ) than previous college-bound test takers.

As the late 1960s and 1970s unfolded, the ESEA continued to grow in scope and cost as the definition of what constituted an “educationally deprived student” expanded. Funds were added to create programs for disabled (special education) students, bilingual students, and dropouts, among other groups. Some liberals considered Goldwater’s defeat a signal of the end of conservatism in the United States. However, presidential politics proved difficult to

---


predict. In November 1967, President Lyndon Johnson’s “old friend,” Minnesota Senator Eugene McCarthy, took the unusual step of challenging a sitting president for the Democratic nomination during the primary cycle, running on an anti-war platform in response to the conflict in Viet Nam. McCarthy finished a close second to Johnson in the March 12 New Hampshire primary election, after which Robert F. Kennedy also entered the race for the Democratic nomination.

Within weeks, Johnson declared he did not intend to run for another term. On June 5, 1968, less than five years after the death of his brother in a similar manner, Robert F. Kennedy was killed by a single gunman. This left McCarthy and Johnson’s vice president Hubert Humphrey as the two most viable Democratic candidates. Due in part to McCarthy’s unpopularity with Kennedy’s now-bereft supporters, he lost the race for the Democratic nomination to Humphrey. McCarthy failed until just days before the general election to endorse Humphrey, who lost to Republican Richard Nixon.

Late 1960s through 1970s: Education Takes a Back Seat

In a 1970 speech about education, Nixon cited the increasingly large amount of federal monies allocated to education in recent years and argued this was based on a faulty premise the

---


Carter, however, served only one term in the White House. In 1980, former California Governor Ronald Reagan defeated Carter in the presidential election.\footnote{Nicholas Lemann, The Big Test: The Secret of the American Meritocracy 244 (Farrar, Straus, and Giroux 2000).} During the 1980 race, Reagan focused primarily on issues other than education. In fact, in a campaign speech, Reagan had referred to the newly-formed United States Department of Education as “President Carter’s
new bureaucratic boondoggle” and posited that education was among the governmental functions that should “primarily be carried out at the state and local levels.” Ergo, at the beginning of his first term, it appeared Reagan’s primary interest in the field of education was to eliminate the Department of Education.

1980s: Excellence Returns

In 1981, Reagan appointed Terrell Bell, a former teacher and school superintendent as Secretary of Education. Reagan reiterated the tenets of his “New Federalism” program as he assumed office, stating that it was “time to bury the myth that bigger government brings more opportunity and compassion.” In keeping with this philosophy, Bell was tasked with eliminating the Department of Education and soon delivered “a detailed plan” on how to fulfill the charge. The Education Consolidation and Improvement Act (ECIA) was passed in 1981 and


677 Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 54 (Cornell University Press 2012).


altered the ESEA. Included in the changes was a transfer of decision-making to the state level and a 20 percent cut to the federal education budget.\footnote{Patrick J. McGuinn, No Child Left Behind and the Transformation of Federal Education Policy 42 (University Press of Kansas 2006).}

Bell, however, came to believe that a cabinet-level education agency was vital.\footnote{Robert C. Johnston, Terrel Bell Known for Defending Federal Role in Education, Dies, Education Week (July 10, 1996), \url{http://www.edweek.org/ew/articles/1996/07/10/40bell.h15.html}.} According to Thomas Anderson, Bell’s chief of staff, Bell “really was Ronald Reagan’s master teacher as far as education.”\footnote{Terrell H. Bell, 74, Education Chief in Reagan Years, New York Times (June 24, 1996), \url{http://www.nytimes.com/1996/06/24/us/terrel-h-bell-74-education-chief-in-reagan-years.html}.} In the face of “very strong political pressures to dismantle the department,” Bell proceeded to “weave and bob”\footnote{Robert C. Johnston, Terrel Bell Known for Defending Federal Role in Education, Dies, Education Week (July 10, 1996), \url{http://www.edweek.org/ew/articles/1996/07/10/40bell.h15.html}.} as he “elongated the process until the momentum died.”\footnote{Robert C. Johnston, Terrel Bell Known for Defending Federal Role in Education, Dies, Education Week (July 10, 1996), \url{http://www.edweek.org/ew/articles/1996/07/10/40bell.h15.html}.} In the words of Bell’s undersecretary Gary Jones, “We’d been in Siberia for two years. They just wanted to reduce our budgets. Suddenly, Ted Bell and the department became important.”\footnote{Robert C. Johnston, Terrel Bell Known for Defending Federal Role in Education, Dies, Education Week (July 10, 1996), \url{http://www.edweek.org/ew/articles/1996/07/10/40bell.h15.html}.}

Bell’s efforts came to fruition on April 26, 1983, when Reagan delivered a speech to the National Commission on Excellence in Education\footnote{Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 54 (Cornell University Press 2012).} that signaled a shift in his administration’s
focus toward education. Regan opened the speech with his trademark humor, remarking that he could have been introduced as the “late president,” reflecting not only a delay in his schedule but also to his having been shot during an attempted assassination. Levity was quickly cast aside as Reagan addressed the committee he had “asked to assess the quality of teaching and learning in America compared with our own educational tradition and the rising competition from other industrial nations.” Like Lyndon Johnson, Reagan also included a reference to Thomas Jefferson, but in Reagan’s case, it was a warning. “If a nation expects us to be ignorant and free, it expects what never was and never will be.”

There were “few areas of American life as important to our society” as education, said Reagan, in part because of “a parallel between a decline in education and a decline in our economy.” The commission had found public education in the United States was in crisis,

---

690 Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 54 (Cornell University Press 2012).


“not because today’s students are any less capable than their predecessors,” but rather due to “low standards, lack of purpose, ineffective use of resources, and failure to challenge students to push performance to the boundaries of individual ability.” This “uninterrupted decline in achievement” since 1963 coincided, noted Reagan, with the “decades in which the federal presence in education grew” and “squandered the gains of the Sputnik era.” Citing $215 billion in total annual spending on education, Reagan claimed the United States had spent “more on education at all levels than any other country in the world.” He then posed a question: “What have we bought with all that spending?”

Reagan’s speech made it clear that the focus of education reform was shifting from equity back to excellence. He did not, however, cite a specific threat as was the case with Sputnik and NDEA. Instead the Nation at Risk Report (NAR), which was the focus of Reagan’s speech,

---


701 Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 44 (Cornell University Press 2012).
referred to “unilateral educational disarmament” in the form of declining public schools. Released by the National Commission on Excellence, NAR differed geographically from earlier excellence efforts as well. Horace Mann had compared United States schools unfavorably to those in Western Europe. The NDEA focused on the perceived superiority of Eastern European education in response to the Russian launch of Sputnik. Rhetoric surrounding NAR, however, made unfavorable comparisons of United States schools to those in Japan. While Japan no longer posed a military threat to the United States following World War II, its economic growth during the 1980s raised fears in business leaders as it arrived on the heels of economic instability in the United States.

The Nation at Risk report “relied in the main upon five sources of information.” These included “papers commissioned from experts on a variety of educational issues” and testimony


705 Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 44 (Cornell University Press 2012).


711 Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 43 (Cornell University Press 2012).

at hearings and other meetings by school administrators, teachers, students, parents, business leaders and other stakeholders. The National Commission for Excellence in Education also reviewed “existing analyses of problems in education,”\textsuperscript{714} letters from concerned citizens, and reports about “notable programs and promising approaches.”\textsuperscript{715}

The report identified several key “indicators of the risk”\textsuperscript{716} in American public education. These risk indicators included the decline in SAT scores first noted in the EEOS chartered during the Johnson administration. The negative trend for SAT scores had continued, as noted the Nation at Risk Report, “virtually unbroken” through 1980.\textsuperscript{717} The NAR also referred to “international comparisons of student achievement completed a decade ago.”\textsuperscript{718} On nineteen of these assessments, “American students were never first or second and, in comparison with other industrialized nations, were last seven times.”\textsuperscript{719} In summary, the report stated, “average

\footnotesize


achievement of high school students on most standardized tests is now lower than twenty-six years ago when Sputnik was launched.” Echoing other themes from the 1950s education reform, NAR cited “a steady decline in science achievement scores of seventeen-year-olds” and the fact that “over half the population of gifted students do not match their tested ability with comparable achievement in school.”

These negative trends in education, as stated the Nation at Risk report, came amid a time when “the demand for highly skilled workers in new fields was accelerating rapidly.” Technology was transforming homes and the workplace. In fact, the NAR predicted that “by the turn of the century millions of jobs will involve laser technology and robotics.” Paul Copperman, an expert quoted in the report, came to the “sobering conclusion” that “for the first time in the history of our country, the educational skills of one generation will not surpass,

---


will not equal, will not even approach, those of their parents.” The report identified a list of “findings.” These reflected what the committee felt were the proximal causes for the risks and reflected “disturbing inadequacies in the way the educational process itself is often conducted.” The findings section of the report centered on four aspects identified by the committee: content, expectations, time, and teaching.

The Nation at Risk report defined content as the “stuff” of education, or curricula. At the secondary level, the committee felt curricula had been “homogenized, diluted, and diffused,” describing a purposeless “cafeteria style” curriculum and bemoaning the fact that students had “migrated from vocational and college preparatory programs to ‘general track’

---


courses in large numbers.” The issue was exacerbated, posited the report, by the fatal combination of “this curricular smorgasboard” and “extensive student choice.” Students, as noted the report, were opting not to take French, geography, science courses, or more than one year of math. Instead 25 percent of general track students’ credits were derived from physical education, work experience, “training for adulthood and marriage,” or remedial English and math. The report also criticized current state tests, which at the time were administered in 37 out of 50 states. These assessments were of the type often described by psychometrists as “minimum competency tests,” and the NAR claimed that the “‘minimum’ tends to become the ‘maximum,’ thus lowering educational standards for all.”


To rectify the issues it identified, the *Nation at Risk* report made several recommendations. These centered around the “New Basics,” which included graduation requirements of four years of English; three years of math, science, and social studies; two years of foreign language; and a semester of computer science. In addition to requiring more courses, the report recommended that within the courses, schools should “adopt more rigorous and measurable standards” and that colleges raise their requirements for admission.

For students to take the “New Basics” courses and meet higher standards, the report also made recommendations about teachers. In response to its findings that teachers tended to be drawn from the bottom 25 percent of college graduates, the NAR called for higher educational standards for teachers and a system for rating university education programs. To attract those who met the higher standards, the report recognized that average teacher salaries, which at the time were around $17,000, needed to be raised. The report recommended an eleven-month contract for teachers that would not only result in “a more adequate level of teacher compensation” but also afford opportunities for curriculum work, professional development,

---


and extra programs for students with special needs.\textsuperscript{745} The report called for the primary responsibility for funding and managing schools as they implemented these recommendations to fall on state and local units of government.\textsuperscript{746}

In contrast to Reagan’s campaign platform, however, the Nation at Risk report outlined a federal role in education oversight. The federal government, the report posited, should work with state and local governments to ensure that the needs of certain “key groups of students such as the gifted and talented, the socioeconomiclly disadvantaged, minority and language minority students, and the handicapped”\textsuperscript{747} were met. The report also outlined “several functions of national consequence that states and localities alone are unlikely to be able to meet.”\textsuperscript{748} These functions included protecting the rights of students and school staff, collecting data about education, supporting research, and providing financial assistance to students. This assistance, the report recommended, “should be provided with a minimum of administrative burden and intrusiveness.”\textsuperscript{749}


The aforementioned goodwill toward Reagan continue through the 1984 election. Reagan won in a landslide against Walter Mondale, who had served as Vice President under Jimmy Carter. Bell resigned following the election and was replaced by William Bennett as Secretary of Education. Under Bennett, the Department of Education not only continued to survive but became “a megaphone for state-led excellence-in-education reform.” The Department published reports entitled “What Works” and began ranking states’ education programs. Bennett also supported the idea of diverting education funds to vouchers for parents who wished for their children to attend private schools, but that idea did not progress beyond the concept phase.

Two key trends emerged because of the “excellence movement.” States provided more funding for and became more involved in oversight of education, enacting approximately 3,000 state-level measures of education reform. Business leaders, such as those belonging to the Business Roundtable, also became more involved in education policy. One of the earliest examples this type of collaboration was found in Boston. The Boston Compact was a pledge by

---


business leaders to provide entry-level jobs for Boston high school graduates in exchange for higher standards and attendance rates in the public schools.\textsuperscript{757} Throughout Reagan’s second term, Secretary Bennett had a United States map on the wall on which he tracked states’ progress.\textsuperscript{758}

1990s: Lofty Goals, Unrequited

Reagan was succeeded by his Vice President, George H.W. Bush, in 1989.\textsuperscript{759} Bush, however, did not favor shuttering the Department of Education or diverting public school funds to private school vouchers. This was reflected in his reference to teachers “giving their heart and soul to their jobs”\textsuperscript{760} and in the content of his first presidential address to Congress, which came to be known as the “Building a Better America”\textsuperscript{761} speech. A centerpiece of the speech was Bush’s “competitiveness program,”\textsuperscript{762} and he referred to improving education in the United States as the most important of the program’s initiatives. “The longer our graduation lines are today,” Bush intoned, “the shorter our unemployment lines will be tomorrow.”\textsuperscript{763} Referring to

\begin{flushleft}
\textsuperscript{757} Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 46 (Cornell University Press 2012).
\textsuperscript{759} Jim Horn & Denise Wilburn, The Mismeasure of Education 48 (Information Age Publishing 2013).
\textsuperscript{760} Patrick J. McGuinn, No Child Left Behind and the Transformation of Federal Education Policy 58 (University Press of Kansas 2006).
\textsuperscript{761} Patrick J. McGuinn, No Child Left Behind and the Transformation of Federal Education Policy 58 (University Press of Kansas 2006).
\textsuperscript{762} George Bush Address Before a Joint Session of Congress, February 9,1989, University of Virginia Miller Center, \url{http://millercenter.org/president/bush/speeches/speech-3420} (last visited Jan. 12, 2017).
\textsuperscript{763} George Bush Address Before a Joint Session of Congress, February 9,1989, University of Virginia Miller Center, \url{http://millercenter.org/president/bush/speeches/speech-3420} (last visited Jan. 12, 2017).
\end{flushleft}
accomplishing this by cutting dropout rates, the president affirmed that the excellence movement was alive and well. “We cannot tolerate mediocrity in education.”764

“When some of our students actually have trouble locating America on a map of the world,”765 said Bush, “it is time for us to map a new approach to education.”766 This new approach, to which Bush hoped to allocate $500 million, included rewarding excellent schools and teachers, increasing math and science achievement, expanding magnet schools, encouraging the use of alternative teacher certification, providing assistance to the neediest schools, and “cut[ting] through bureaucracy.”767 “I’ve said I’d like to be the ‘Education President,”’768 concluded Bush, “And tonight, I’d ask you to join me by becoming the ‘Education Congress.”’769

Congress, however, appeared uninterested, failing a few months later to pass the Bush administration’s Education Excellence Act. Noted among the reasons for the lack of supporting votes were the recession impacting the economy, the relationship between Democratic Congressmen and teachers’ unions, and the preference of many Republican members of Congress for Ronald Reagan’s efforts to eliminate the Department of Education and issue


vouchers. While those Reagan-era philosophies on vouchers and elimination of the Department of Education were not shared by Bush, the emphasis on corporate and state-level influence on education reform continued. In fact, Bush derived some of the strongest interest in his education policies from the National Governors Association (NGA), who assembled a summit in North Carolina in the fall of 1989. A group of Southern governors who had been active in raising standards for education in their states—including Richard Riley of South Carolina, Lamar Alexander of Tennessee, and William Jefferson Clinton of Arkansas—played pivotal roles in the conference. The NGA released a statement at the conclusion of the summit, stating, “We believe that the time has come, for the first time in U.S. history, to establish clear national performance goals, goals that will make us internationally competitive.”

President Bush announced the six goals developed at the summit during his 1990 State of the Union address: All children should arrive at school ready to learn. The high school graduation rate will increase to at least 90 percent. Students in grades four, eight and twelve will be “competent” in English, mathematics, science, foreign language, civics and government, economics, arts, history, and geography. Every school will be free of drugs, violence, firearms, and alcohol and will offer a disciplined learning environment. U.S. students will be the first in

---


the world in math and science achievement. Every adult will be literate and will possess the knowledge and skills necessary to compete in a global economy.\textsuperscript{774}

George H.W. Bush, however, did not see this vision realized during his time as president. His administration’s second attempt at education reform, America 2000, was passed by the House of Representatives in 1992, but was drastically altered. Spending was increased, while school choice and assessment requirements were minimized. The Bush administration threatened a veto.\textsuperscript{775} Senate Republicans, however, cognizant that a president vetoing an education bill during an election year would be damaging to their party filibustered the bill until the 1992 Congress adjourned.\textsuperscript{776} In the end, it was not only Congressional resistance and the economy that derailed George H.W. Bush’s attempts to impact federal education policy.\textsuperscript{777} Perhaps the most important factor was that Bush served only one term, after which he was defeated in the 1992 election by one of the aforementioned governors, Bill Clinton.\textsuperscript{778}

Clinton selected his fellow education reformer, South Carolina Governor Richard Riley, to serve as Secretary of Education.\textsuperscript{779} Within a year of taking office, the Clinton administration passed two pieces of major education reform legislation. The first was the Goals 2000: Educate


\textsuperscript{775} Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 93 (Cornell University Press 2012).

\textsuperscript{776} Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 94 (Cornell University Press 2012).

\textsuperscript{777} Patrick J. McGuinn, No Child Left Behind and the Transformation of Federal Education Policy 58 (University Press of Kansas 2006).

\textsuperscript{778} Bill Clinton: Campaigns and Elections, University of Virginia Miller Center, (February 9, 1989), http://millercenter.org/president/biography/clinton-campaigns-and-elections.

\textsuperscript{779} Patrick J. McGuinn, No Child Left Behind and the Transformation of Federal Education Policy 86 (University Press of Kansas 2006).
President Clinton delivered a speech at Zamorano Fine Arts Academy in San Diego as he signed Goals 2000 into law on March 31, 1994. Addressing his remarks directly to the high school students, Clinton referred to his and Riley’s presence at the 1989 summit as he reviewed the six goals established by the Bush-led gathering in Charlottesville, including early childhood readiness, increased graduation rates, drug-free schools, and “world-class” standards in academic subjects.

In fact, during the 1992 presidential race, an aide remarked that much of Clinton’s education campaign platform was “indistinguishable from policies of the Bush administration.” This may have been in part because, like Reagan and Bush before him, Clinton maintained close ties with the business community, including the National Alliance of Business, a group that was frustrated with the lack of progress toward national standards and assessments. Goals 2000 added two objectives to the original six tenets from the 1989 summit. One new goal pertained to partnerships with businesses. The other involved professional

---


784 Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 104 (Cornell University Press 2012).
development for teachers, which Clinton described in his speech as “taking care of our teachers better.”

To pass both houses of Congress, however, the verbiage on standards and assessments was modified to the point at which the development and use of national standards was voluntary. Not only could states develop their own curriculum and tests, but the law explicitly stated that a state was “not required to have its standards and assessments certified”

to receive federal funds through Goals 2000. In keeping with this compromise, Clinton referred in the same section of his Sand Diego speech to “world class standards” and “measuring whether the children learn them or not” but also to “grassroots reforms.”

Paraphrasing his charge to the states, Clinton said, “Now, you figure out how to do it. Use your mind, use your energy, and we will support you.” The support came primarily in the form of $400,000,000 of federal funding allocated for fiscal year 1994 that was spread across

---


the federal and state levels for wide-ranging list of initiatives. Programs receiving a portion of
Goals 2000 funds included several national research organizations,\textsuperscript{793} the aforementioned pre-
service training and professional development for teachers,\textsuperscript{794} and grants to states to develop
standards and assessments.\textsuperscript{795}

Within months, the Clinton administration launched its second wave of education
legislation, the Improving America’s Schools Act (IASA). While Goals 2000 was a relatively
small new categorical grant to states that opted to take it,\textsuperscript{796} the IASA was a reauthorization of
the ESEA, the grants from which had long been accepted by all fifty states.\textsuperscript{797} Not only did the
IASA have history and participation rate on its side, but it also provided much more significant
funding—appropriating nearly $10 billion\textsuperscript{798} in comparison to the $400 million proffered by
Goals 2000.\textsuperscript{799} As a result, according to the recollections of Clinton staffer Michael Cohen, the


\textsuperscript{795} Patrick J. McGuinn, No Child Left Behind and the Transformation of Federal Education Policy 90 (University Press of Kansas 2006).

\textsuperscript{796} Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 112 (Cornell University Press 2012).

\textsuperscript{797} Patrick J. McGuinn, No Child Left Behind and the Transformation of Federal Education Policy 92 (University Press of Kansas 2006).

\textsuperscript{798} Patrick J. McGuinn, No Child Left Behind and the Transformation of Federal Education Policy 92 (University Press of Kansas 2006).

IASA had the potential to be “more coercive and more specific”\(^{800}\) in achieving the desired reforms.\(^{801}\)

To accomplish broad reforms, however, the ESEA would need to be reframed, a task initiated by a brochure authored by Secretary of Education Richard Riley. The IASA, per the brochure, would replace “the piecemeal structure of the old ESEA.”\(^{802}\) Instead ESEA programs would be “integrated into a state’s overall school improvement efforts, focused around a core of challenging state standards.”\(^{803}\) Accountability for implementing the standards would come in the form of “state assessments that measure students’ progress toward new state standards.”\(^{804}\) Also included were provisions mandating school improvement planning\(^{805}\) and District and School Report Cards.\(^{806}\)

The reach of the IASA extended further. It addressed staff training through a new Title II, the Eisenhower Professional Development Program. President Clinton’s reauthorization of

\(^{800}\) Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 113 (Cornell University Press 2012).

\(^{801}\) Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 113 (Cornell University Press 2012).


\(^{806}\) Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 96 (Cornell University Press 2012).
ESEA also required corrective action for schools or districts not meeting achievement targets, provided training via Title VII to promote instruction for Limited English Proficient students, and granted school choice to parents in some instances. It also included provisions for innovation, public charter schools, and coordinated services among districts. Lastly, IASA set a precedent with the addition of Title IV, The Safe and Drug-Free Schools and Communities Act of 1994.\textsuperscript{807} By the time IASA passed both houses of Congress and was signed by President Clinton, the original 32-page ESEA legislation\textsuperscript{808} from 1965 had grown to 545 pages.\textsuperscript{809} Ambitious as they were, however, the aims of Goals 2000 and IASA were never fully realized. Resources were scarce, as the Clinton administration prioritized departments such as Health and Human Services and the reduction of a projected deficit of $300 billion.\textsuperscript{810} The Clinton administration also faced growing opposition to the standards movement as the 1996 election approached. In 1994, the Republicans gained 52 seats in the House and eight seats in the Senate, giving them a majority in both houses of Congress.\textsuperscript{811} Led by newly-elected Speaker

\begin{thebibliography}{9}


\bibitem{808} Elementary and Secondary Education Act of 1965, 20 U.S.C 6301.


\bibitem{810} Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 113 (Cornell University Press 2012).

\bibitem{811} Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 128 (Cornell University Press 2012).

\end{thebibliography}
Newt Gingrich, the Republicans organized their policies under the umbrella of a “Contract with America.”

While education was not a major component of the Contract with America, both houses passed HR 1158 that proposed cutting more than $870 million from the budget of the United States Department of Education, a department some legislation proposed eliminating. The cuts included reducing Goals 2000 by $84 million.

In June 1995, Clinton used his presidential veto power for the first time to defeat HR 1158. However, to gain Republican support for an appropriations bill, the president agreed the next month to cut more than $570 million from education—including $31.5 million from Goals 2000. In 1995, Secretary Riley issued a “Dear Senator” letter, promising not to attach regulations to 38 of the IASA programs and to utilize only very limited regulations for the remaining seven programs. Congress, however, continued to introduce legislation to counter

---


814 Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 130 (Cornell University Press 2012).


Goals 2000 and the IASA. The “Goals 2000 Elimination Act” introduced in 1996 failed to pass, but several amendments to Goals 2000 were enacted the same year. The amendments abolished the National Education Standards and Improvement Council and also eliminated opportunity-to-learn standards, outcome-based education, and requirements regarding social services, all of which had been controversial provisions of Clinton education policy. The amendments also granted waivers to six states and to the others the ability to simply state that they had formed an improvement plan rather than submitting one in detail. Allocation of funds, including the ability to purchase technology in lieu of developing standards or assessments, was also left largely up to the states.

Even left to their own devices, some states—including those with both Democrat and Republican governors—joined the standards and accountability movement. Massachusetts, North Carolina, and Texas were among the most prevalent. Beginning with his election in 1995, the governor of Texas was George W. Bush, son of former President George H.W. Bush. Under Governor Bush, Texas required the annual assessment of students in reading and math beginning in grade three. Results from the Texas Assessment of Academic Skills (TAAS) were reported not only as a percentage of all students meeting or exceeding standards but were also disaggregated into subgroups, including race, income, disability, and gender.

---


In 2000, George W. Bush entered the presidential race against Clinton’s vice president, Al Gore, and made education an important part of his platform. According to Margaret Spellings, a key Bush campaign adviser, “The standard shtick [among Republicans prior to 2000] had been ‘abolish the Department of Education’” and minimize federal involvement in education, a philosophy that changed dramatically after Bush won the presidential election. This shift, and the reauthorization that followed, had not only the strong support, but considerable input, from the Business Roundtable, the same private-sector group that had influenced federal education reform during the Clinton administration. In the public sector, Bush built a coalition that included Republican Congressman John Boehner of Ohio, Democratic Representative George Miller of California, Senator Judd Gregg, a Republican senator from New Hampshire, and Democratic Senator Ted Kennedy of Massachusetts, brother of the late John F. Kennedy and Robert Kennedy. Together, they crafted legislation reauthorizing the ESEA that came to be known as No Child Left Behind (NCLB). The bipartisan approach led to a dual emphasis on both excellence and equity that was captured by the phrase often used by Bush as he

824 Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 139 (Cornell University Press 2012).

825 Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 140 (Cornell University Press 2012).

826 Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 140 (Cornell University Press 2012).


persuaded additional members of both parties to support the NCLB Act, which he posited would combat “the soft bigotry of low expectations.”

As NCLB, officially known as HR 1, wove its way through the 107th Congress, President Bush toured the country to promote it. One such promotional visit was to Emma E. Booker Elementary School in Sarasota, Florida. On September 11, 2001, Bush was reading to a class of second graders who were “sitting in their chairs, looking forward, and were kind of giggling and glad to see the president” when he was interrupted by a whisper in the ear from Andrew Card, his chief of staff. From Card, Bush learned that two planes hijacked by terrorists had hit the twin World Trade Center towers. The attack “strengthened the resolve of both Congress and the administration to proceed with the business of the people,” and NCLB was ratified by both houses of Congress within just a few months.

---


On January 8, 2002, President George W. Bush signed NCLB into law. With Democratic Senator Ted Kennedy on the dais with him, Bush spoke to students, staff, and parents at Hamilton High School in Hamilton, Ohio, home of Republican Senate Education Committee Chairman John Boehner. “I know you all are anxious to get back to class. So please be seated,” George W. Bush opened with his characteristic humor. Bush thanked Boehner, Kennedy, and a number of political figures from both parties, including his Secretary of Education, Rod Paige. Paige, who had begun his career as a teacher, had come to Bush’s attention during his tenure as Superintendent of the Houston Independent School District. “I asked Rod to join my administration,” stated Bush, “because I wanted somebody who understood what it meant to run a school district in Washington, D.C. I didn’t need somebody that based his knowledge on theory. I wanted somebody who based his knowledge on experience.”

“Today begins a new era, a new time in public education in our country. As of this hour, America’s schools will be on a new path of reform, and a new path of results,” intoned Bush.

---

“It’s not exactly light reading,” said President Bush as he pointed to a box containing a hard copy of NCLB. Under its 2002 reauthorization, the ESEA had grown to 670 pages, but Bush asserted that it “contains some very important principles that will help guide our public school system for the next decades.” The president outlined four of the principles, the first being accountability.

“Every school has a job to do, and that’s to teach the basics and teach them well.” To monitor this, Bush explained, “in return for federal dollars, we are asking states to design accountability systems to show parents and teachers whether or not children can read and write and add and subtract in grades three through eight.” “I understand taking tests aren’t [sic] fun,” conceded Bush. “Too bad. We need to know in America. We need to know whether or not children have got the basic education.”

Bush hinted that one reason NCLB required annual testing in grades three through eight might be to end social promotion, defined by social policy historian Timothy Hasci as “the practice of promoting children from one grade to the next whether they have learned much or


not."848 “I read a quote from a young lady in New York,” shared Bush, “She said, ‘I don’t ever remember taking an exam. They just kept passing me along. I ended up dropping out in the 7th grade. I basically felt nobody cared.’”849 The focus of NCLB, however, clearly went beyond individual students to schools as a whole. “No longer is it acceptable to hide poor performance,” declared Bush. “We do not want children trapped in schools that will not change and will not teach,”850 he summarized.

In keeping with his comment to the educators in the audience that “yours is indeed a noble profession,”851 Bush reassured them that “we’re never going to give up on a school that’s performing poorly.”852 Instead “when we find poor performance, a school will be given time and incentives and resources to correct their problems. A school will be given time to try other methodologies, perhaps other leadership, to make sure that people can succeed.”853 Bush also made it clear, however, that after this period of support, if a school was “unable to solve the problem of not educating their children,”854 there would be “real consequences.”855

These consequences would be public: “No longer is it acceptable to keep results away
from parents,” Bush declared, transitioning to the second principle of NCLB. “We trust
parents to make the right decisions for their children.” In the case of “any school that doesn’t
perform, any school that cannot catch up and do its job,” there would be “a moment” Bush
promised, “in which parents can say, ‘I’ve had enough of this school.’” “Parents,” said the
president “must be given real options in the face of failure in order to make sure reform is
meaningful.” Under NCLB, these choices would include “a better public school, a tutor, or a
charter school.”

The third principle of NCLB outlined by President Bush was that “we have got to trust
the local folks on how to achieve standards, to meet the standards. In Washington, there’s some
smart people there,” he conceded, but “the people who care most about the children” were

to be found at the local level. “The federal government will not micromanage how schools are run,”865 promised Bush. “I can’t think of a better way to say to teachers, ‘We trust you…We want you to have as much flexibility as possible to see to it that every child that walks in your classroom can succeed,”866 he assured the educators in the audience.

As Bush introduced the fourth principle, however, it was evident that flexibility had its limits under NCLB. “We’re going to spend more money, more resources,”867 stated Bush, “but they’ll be directed at methods that work. Not feel-good methods, not sound-good methods, but methods that actually work.”868 Bush continued, this principle is applied “particularly when it comes to reading,”869 a subject area for which NCLB “tripled the amount of federal funding for scientifically-based early reading programs.”870 “We’re going to spend more on our schools,” 871 summarized Bush, “and we’re going to spend it more wisely. If we’ve learned anything over the last generations, money alone doesn’t make a good school.”872


Having learned from the lessons of Goals 2000 and the IASA, however, Bush knew that Washington alone could not maintain an accountability system. To this end, Secretary of Education Rod Paige wasted no time, meeting the next night with 30 state chief education officers. Paige “warned them that they would be held to the letter of the law and that he would not grant waivers or tolerate noncompliance as his predecessors had done.”

Unlike the Clinton administration, Paige informed the state-level leaders that the Department of Education under Bush had created “tough, detailed regulations in support of NCLB” and “threatened to withhold federal funds from states that do not comply with its mandates.”

These mandates included many of the initiatives Goals 2000 had merely suggested. NCLB required states to adopt “challenging academic standards.” Student mastery of the reading and math standards was to be measured annually in grades three through eight and during high school using common state assessments that tested “higher order thinking skills and understanding,” something only 13 states had undertaken as a result of Goals 2000.

---


Suggesting that the specter of Sputnik still hovered over education, states were also directed by NCLB to administer a science assessment to all students once during each grade span (elementary, middle school, and high school). In addition to its own tests, each state was required to administer the National Assessment of Educational Progress (NAEP) annually to a sample of students. However, NCLB also contained verbiage expressly prohibiting other federally sponsored testing. “No funds provided under this Act to the Secretary or to the recipient of any award may be used to develop, pilot test, field test, implement, administer, or distribute any federally sponsored national test in reading, mathematics, or any other subject.”

Results of the reading and math assessments were not only to be reported for each state, local education agency, and school, but also by each gender, racial, ethnic, disability, English Proficiency, and economic subgroup, as they had been in the state of Texas during George W. Bush’s tenure as governor. Subgroup data would be reported down to the school and grade level, so long as the size of the group was large enough to prevent breaching student

---


confidentiality by revealing personally identifiable information.\footnote{The No Child Left Behind Act of 2001, 20 U.S.C § 6311(2)(C)(II) (2002). As retrieved from https://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf (February 16, 2017, 7:44 PM)} These data were to be provided to the public in the form of a “report card” for each state, district, and school.\footnote{Patrick J. McGuinn, No Child Left Behind and the Transformation of Federal Education Policy 180 (University Press of Kansas 2006).} Not only would report cards display demographic and achievement data, but also the percentage of each district and school’s teachers who were “highly qualified”\footnote{Patrick J. McGuinn, No Child Left Behind and the Transformation of Federal Education Policy 180 (University Press of Kansas 2006).} as defined by NCLB. This entailed obtaining certification and demonstrating “a high level of competence”\footnote{Patrick J. McGuinn, No Child Left Behind and the Transformation of Federal Education Policy 180 (University Press of Kansas 2006).} in the subjects taught.

“statistically valid and reliable,”\textsuperscript{895} generate “continuous and substantial academic improvement for all students,”\textsuperscript{896} and measure the progress of all public K-12 schools and districts “based primarily on the academic assessments”\textsuperscript{897} described earlier.

If any school failed to make AYP for two consecutive school years, NCLB required that it be identified as being in “School Improvement” status.\textsuperscript{898} In such instances, the district was required to provide students who wished to transfer with transportation to another school—district or public charter—that was not in status.\textsuperscript{899} If space was limited, first priority was to be given to “the lowest-achieving children from low-income families.”\textsuperscript{900} In the meantime, the district was also required to provide “technical assistance,”\textsuperscript{901} including professional development for staff and parent involvement facilitation.\textsuperscript{902}

If the same school failed to make AYP for a third consecutive year, NCLB required that funds previously used by districts to assist students of poverty instead be paid to a provider for


“supplemental educational services,” including tutoring, as chosen by parents. A fourth consecutive year without AYP moved a school further along the status continuum. School choice and supplemental educational services would continue, and the district would be required to take one of several “corrective actions.” Options for corrective action under NCLB included “replacing the school staff who are relevant to the failure to make adequate yearly progress,” providing “scientifically based” professional development and curriculum for staff who showed “substantial promise” of helping the school make AYP, “significantly decreasing designating management authority at the school level,” designating “an outside expert to advise the school on its progress,” and extending the hours of school per day or days of school per year.


If a school reached a fifth year without making AYP, it was placed in the Restructuring phase, which offered five options. A district could reopen the school as a public charter school;\textsuperscript{913} contract “with an entity, such as a private management company with a demonstrated record of effectiveness, to operate the school;”\textsuperscript{914} or turn control of the school over to the state.\textsuperscript{915} If the district chose to retain control of the day-to-day management of the school, it could replace “all or most of the staff,”\textsuperscript{916} which the legislation explicitly stated might include the principal.\textsuperscript{917}

The same AYP calculations, timelines, and five phases of sanctions applied to districts as a whole under NCLB, with states required to impose the sanctions.\textsuperscript{918}

Through its graduated AYP model, each state, local education agency, and school was required by NCLB to move 100 percent of its students to meeting or exceeding standards by 2014 to avoid the aforementioned sanctions.\textsuperscript{919} When the four principles of NCLB were applied, 


\textsuperscript{918} Patrick J. McGuinn, No Child Left Behind and the Transformation of Federal Education Policy 180 (University Press of Kansas 2006).

\textsuperscript{919} Patrick J. McGuinn, No Child Left Behind and the Transformation of Federal Education Policy 180 (University Press of Kansas 2006).
Bush felt this “grand goal for the country”\textsuperscript{920} was “not an impossible goal.”\textsuperscript{921} Instead it was “a great symbol of what is possible in Washington.”\textsuperscript{922} Responsibility for implementing NCLB, however, depended on the states. By Spring 2002, only 16 states had fully implemented the recommendations of Goals 2000 and the IASA, such as developing rigorous new state standards and standardized state tests for grades three through eight.\textsuperscript{923} Under No Child Left Behind, however, all 50 states had obtained federal approval of the process for developing their reading, math, and science standards by 2003 and completed them by the 2006-07 school year.\textsuperscript{924} Within the same time frame, each of the 50 states had also developed and administered annual statewide assessments in grades three through eight.\textsuperscript{925} States also complied with the NCLB requirement to test students “not less than once” during grades ten through twelve in reading, math,\textsuperscript{926} and science.\textsuperscript{927} In a precedent-setting move, many states chose to adopt the ACT or SAT as their high school accountability test. By


\textsuperscript{923} Patrick J. McGuinn, No Child Left Behind and the Transformation of Federal Education Policy 181 (University Press of Kansas 2006).


2007, five years into NCLB, 17 states required the ACT, SAT, or Workkeys, a career placement test administered by ACT.\footnote{James Taylor, Jennifer O’Day, & Kerstin Carlson Le Floch, State and Local Implementation of the No Child Left Behind Act 16 (U.S. Department of Education, 2010), \url{http://files.eric.ed.gov/fulltext/ED508912.pdf}.} By 2014, that number had increased, with 23 states either funding, mandating, or contracting for the administration of the ACT or SAT to all high school students.\footnote{Caralee J. Adams, State Initiatives Widen Reach of ACT, SAT Tests, Education Week, \url{http://www.edweek.org/ew/articles/2014/10/29/10satact_h34.html} (last visited Oct. 28, 2014).}

While No Child Left Behind generated tangible compliance in the form of standards adoption and state testing, its implementation did not advance without significant opposition, much of which began to emerge just as President George W. Bush campaigned for a second term against his opponent, Democratic Senator John Kerry. In February 2004, the National Governor’s Association, the group that had been instrumental in starting the standards movement in 1989, called for changes to NCLB during their winter conference.\footnote{Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 164 (Cornell University Press 2012).} The same month, Congressional Democrats, many of whom had voted for the legislation and continued to support the concept of accountability, met with Secretary of Education Rod Paige to express their concerns about implementation issues.\footnote{Alan Richard & Erik W. Robelen, Federal Law is Questioned by Governors, Education Week, (March 17, 2004), \url{https://eric.ed.gov/?q=percent22no+child+left+behind+percent22&ff1=autRobelen+percent2C+Erik+W.&ff2=autRichard+percent2C+Alan&id=EJ755595}.} Chief State School Officers from 16 states drafted a letter requesting flexibility as they worked to meet NCLB accountability requirements.\footnote{Jim Horn & Denise Wilburn, The Mismeasure of Education 113 (Information Age Publishing 2013).} Some state legislatures also became involved. The Virginia House of Delegates—led by a Republican majority— passed a resolution requesting relief from NCLB, citing a history of successful state-
initiated reforms and the burdensome cost of implementation. Virginia was not alone in its concern about costs. Five states passed legislation that prohibited the expenditure of any state funds to meet the mandates of NCLB, and a total of 31 state legislatures introduced some form of legislation seeking flexibility or exemption.

In July 2004, the Education Commission of the States (ECS)—of which Virginia Governor Mark Warner was Chairman—issued a report on the implementation of NCLB across states. The report acknowledged the complex reaction to No Child Left Behind, which had generated “both strong support and deep concern.” “To many,” the report stated, “NCLB embodies—and even elevates—America’s longstanding commitment to public education and the central role it plays in maintaining the nation’s economic competitiveness, the strength of its institutions, the vitality of its communities, and the well-being of its citizens.” The report celebrated the “considerable progress” made since NCLB was signed into law. The ECS found that all states had made at least some growth toward meeting the 40 NCLB requirements included in its analysis and that the “overwhelming majority” had met the expectations to test


new teachers, raise participation rates on state assessments to 95 percent, and set criteria for determining “safe schools.”

The report also, however, cited a number of obstacles during initial implementation that contributed to the opposition to NCLB. While some lauded it, the ECS asserted, “others view NCLB as well-intended but far beyond the capacity of states, districts, and schools to carry out. Still others see the law as burdensome and unwarranted intrusion on state and local prerogatives and responsibilities.” Specifically, the report noted that fewer than half of states were on track to meet requirements to staff all classrooms with teachers defined as “highly qualified” by NCLB, offering professional development for all teachers who would meet NCLB criteria for being “high-quality,” or providing technical assistance to underperforming schools defined as “scientifically based” by the federal law. The final aspect of NCLB the ECS found problematic was the development of data systems. Creating such a system was not an explicit requirement of NCLB, but rather a resource without which “states will have difficulty meeting a number of the law’s requirements.” The ECS reported that “many states do not have in place


the technology infrastructure needed to collect, disaggregate, and report data at the school, district, and state levels.”

Another component of NCLB that generated controversy was the mandate for schools in their third year of status to offer Supplemental Educational Services. The 2002-03 school year was the first in which schools and districts were required to implement SES, a process the U.S. Department of Education acknowledged in a report issued in 2004 “thrust states, districts, and providers into uncharted territory.” Born of bipartisan compromise, SES created a “hybrid market” in which government, while not the end consumer, provided the parameters and funding for services to students. The monies were not new allocations but rather were derived from a requirement that districts set aside 20 percent of their Title I funds to pay for any supplemental educational services selected by low-income parents for their children. Because this was a set-aside that could be returned to their general Title I accounts if not expended, districts “knew that the better they implemented SES, the less of their own funding would be

---


Also, although a school was assigned this sanction due to low test scores, SES had to be offered to all low-income students in a school regardless of their performance.952

Perhaps most critically, NCLB regulations provided “no incentive for states to embrace a rigorous provider approval process.”953 Paired with the use of the words “without delay,”954 “timely,”955 and “promptly”956 throughout NCLB, this led to states approving an “ever-increasing number of questionable providers.”957 Also while No Child Left Behind was “very prescriptive”958 as to expectations for schools and districts, “what no one has learned much about,”959 reported the Department of Education and the end of the first year of Supplemental

---


Educational Services, “is how to monitor and evaluate provider performance.” Ergo, once an SES company was placed on the list of approved providers, it possessed “a hunting license to bag as many enrollments as possible.”

A particularly egregious example of a “questionable provider” was that of two Illinois-based companies (i.e., Brilliance Academy, Inc. and Babbage Net Schools Inc.) the owners of which were indicted for defrauding 200 public school districts in 19 states out of more than $33 million. Father-and-son co-owners Jowhar Soutanali and Kabir Kassam were charged with “misrepresenting the nature and quality of the tutoring services the companies provided,” “falsely inflating invoices,” and “creating and distributing false student progress and improvement reports.” Charged along with the Brilliance and Babbage owners were four school officials, indicted for federal program bribery for “an unspecified amount of money from

---


Soultanali, Kassam, Brilliance and Babbage in exchange for recruiting students. Two of the officials also were charged with receiving bribes in the form of gifts. Cedric Petersen, an SES coordinator and assistant principal from San Antonio, Texas, was charged with accepting a Caribbean Cruise. Arturo Martinez, a state education official from New Mexico responsible for the state’s Migrant Education Program and for approving and auditing SES providers, was alleged to have received “services at a gentleman’s club,”—not the type of “services” for which SES funds were intended. As of August 23, 2016, Soultanali, Kassam, and Martinez had pled guilty. Martinez was sentenced on December 8, 2016, to one year in prison. As of August 30, 2017, Kassam and Soultanali were asking the court for a two-year sentence. Prosecutors, however, argued in their sentencing memorandum that the sentence should be longer due to Kassam and Soultanali’s “failure to fully accept responsibility for their actions,” and attempting to deflect personal responsibility by making reference to “the corruption endemic to the [federal grant] program.”

---


The issues surrounding Supplemental Educational Services have been cited as “a key reason”972 the subsequent presidential administration “provided broad-based waivers for many of NCLB’s requirements.”973 In the meantime, however, while Democratic candidate John Kerry pledged on the presidential campaign trail to “put new resources into our schools and make reform work by fully funding No Child Left Behind, creating a new bargain with America’s teachers, and beginning a national campaign to raise high school graduation rates,”974 George W. Bush held firm in his rhetoric. “We’re not backing down,”975 Bush promised. “I don’t care how much pressure they try to put on the process. I’m not changing my mind about high standards and the need for accountability.”976 This stance was supported by the Business Roundtable.977 However, the Bush administration granted some flexibility to states as the election approached. The Department of Education announced changes in accountability policies for students with disabilities or Limited English Proficiency and also allowed states more leeway in how they calculated participation rates.978


974 Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 166 (Cornell University Press 2012).

975 Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 164 (Cornell University Press 2012).


Bush won the election, and shortly thereafter he replaced Secretary of Education Rod Paige with former campaign aide Margaret Spellings. Spellings instituted additional flexibility for NCLB after reiterating the non-negotiables: annual reading and math assessments, with scores disaggregated by subgroups, highly qualified teachers, school choice, and 100 percent of students meeting standards by 2014. It was the latter requirement that motivated Spellings to make the most significant alteration to NCLB. As states faced “increasingly impossible proficiency targets,” the United States Department of Education established a committee to provide peer review of growth models proposed by states. By November 2005, the committee had evolved into the Growth Model Pilot Project (GMPP), for which 10 states applied and eight were approved.

The first two states, North Carolina and Tennessee, commenced the pilot during the 2005-06 school year. Both used a “value-added” model developed in the 1980s by Dr. Ted Sanders that purported to measure teacher effects on student academic growth. At various points over the next three school years, each of the eight states were given flexibility to

---


deviate from the NCLB mandate of using proficiency benchmarks, to which the pilot report referred as the “status model.”

GMPP states instead calculated AYP for schools and districts using “growth models,” also known as “projection models,” that used longitudinal data to “determine whether each student was ‘on track’ to reach or exceed the state’s proficiency cut points” either within three to four years or by a set grade level. For the final AYP determination, students who were not yet proficient but demonstrated sufficient growth as determined by the state’s model would be “counted the same as proficient students.” All states included in the pilot saw more schools making AYP when they applied a growth model than they had when using the original NCLB status model.

Using growth models was one of several recommendations for the reauthorization of NCLB in a joint statement drafted by the National Governors Association, the Council of Chief


State School Officers, and the National Association of State Boards of Education. The statement also recommended maintaining states’ authority to select appropriate assessments, allowing districts to intervene with SES before offering school choice, committing additional resources—including rewards and incentives—to education reform, providing flexibility in the assessment of special education and English Language Learner (ELL) students, expanding access to college-credit-bearing high school courses such as Advanced Placement (AP) and International Baccalaureate (IB) classes, and using multiple measures to assess student progress.

The joint statement was released in 2007 after a failed attempt to reauthorize NCLB. In 2006, the Democratic Party regained control of both houses of Congress for the first time.

---


1000 Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 168 (Cornell University Press 2012).
since the Republican Revolution of 1994. Civil rights activist Representative George Miller assumed chairmanship of the House Education Committee and warned that a reauthorization of NCLB would not pass without “serious changes.” Representative Miller soon learned that the various interest groups involved in shaping NCLB had vastly differing ideas about which changes needed to be made. Miller compiled a four-hundred page “discussion draft” intended to provide a roadmap for compromise but was later heard to say that stakeholders on all sides “treated it like a piñata.” By the time the discussion draft and reauthorization efforts imploded, there were only months remaining until the next presidential election, after which it was anticipated that NCLB would finally be reauthorized two years behind schedule.

After defeating fellow Democratic Senator Hillary Clinton, wife of former president Bill Clinton, in a hard-fought 2008 primary Illinois Senator Barack Obama went on to campaign for president against Republican Senator John McCain of Arizona. During the campaign, Obama sought to distinguish himself from the Bush administration and McCain as the

---


1004 Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 172 (Cornell University Press 2012).


“change” candidate, committed to “fixing the broken promises of No Child Left Behind.” During a campaign stop at a school in Denver, Colorado, he detailed “what is wrong with No Child Left Behind,” including a lack of funding, “forcing teachers and students to spend most of the year preparing for a single, high-stakes test” and “labeling a school and its students as failures one day and then throwing your hands up and walking away from them” the next. Obama promised that he would provide more funding and collaborate with governors and educators to develop “assessments that can improve achievement all across America by including the kinds of research, scientific investigation, and problem-solving that our children will need to compete in a 21st century knowledge economy.”

Reflecting the bipartisan consensus surrounding 21st century education reform, however, Senator Obama also clarified that he believed that the goals of NCLB—including closing the achievement gap, setting higher standards, and increasing accountability—“were the right

---

1009 Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 159 (Cornell University Press 2012).

1010 Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 159 (Cornell University Press 2012).


ones.” Obama cited his belief that “from the moment our children step into a classroom, the single most important factor in determining their achievement is not the color of their skin or where they come from. It’s not who their parents are or how much money they have. It’s who their teacher is.” Candidate Obama vowed to “find new ways to increase teacher pay…If they consistently excel in the classroom, that work can be valued and rewarded.” Like many of his predecessors from both parties, he listed dismal statistics, referred to the economic future of the nation, and quoted Thomas Jefferson: “Talent and virtue, needed in a free society, should be educated regardless of wealth or birth.”

Obama won the presidential election in 2008. That year was also a momentous one for a private-sector entrepreneur with an interest in shaping the future of public education. Bill Gates, founder and CEO of software giant Microsoft, resigned his post to co-chair his family’s sizable foundation. The Gates Foundation, started by Gates and his wife Melinda in 2000, had initially focused on projects such as connecting public libraries to the Internet that “sprang from our founders’ Microsoft experience.” The work of the Gates Foundation eventually broadened


into a wider array of issues, including global health and public education.\textsuperscript{1022} For the latter field, the foundation set “an ambitious national goal”\textsuperscript{1023} to “graduate all students in the United States ready for success in college.”\textsuperscript{1024}

President Obama selected Arne Duncan as his Secretary of Education. Duncan, who first met fellow Chicagoan Obama on the basketball court in the 1980s, was at the time serving as Superintendent of the Chicago Public Schools. He had been a frequent Obama advisor on education issues during the latter’s tenure as an Illinois State Senator and United States Senator.\textsuperscript{1025} Duncan recruited many of his senior Department of Education staff from both the Gates Foundation and the Broad Foundation,\textsuperscript{1026} an organization headed by philanthropists Eli and Edythe Broad\textsuperscript{1027} known for funding public charter schools and other education reform initiatives.\textsuperscript{1028}

The first fruit of their labor was an initiative known as Race to the Top (RTTT).\textsuperscript{1029} RTTT was funded by the American Recovery and Reinvestment Act (ARRA), a broad economic

\begin{footnotesize}
\textsuperscript{1026} Jim Horn & Denise Wilburn, \textit{The Mismeasure of Education} 2 (Information Age Publishing 2013).
\textsuperscript{1029} Arne Duncan, \textit{The Race to the Top Begins—Remarks by Secretary Arne Duncan} (July 24, 2009), http://www.ed.gov/news/speeches.
\end{footnotesize}
stimulus package\textsuperscript{1030} signed into law on February 17, 2009. Title VIII of ARRA provided supplemental appropriations to several federal departments—including the Department of Education—through FY 2010.\textsuperscript{1031} The appropriations, albeit not guaranteed renewal past 2010,\textsuperscript{1032} were substantial.

In his speech introducing the competitive grant,\textsuperscript{1033} Secretary Duncan emphasized that the available funds amounted to more than the total amount of discretionary funds available to his predecessors over the last 29 years. “All of them,”\textsuperscript{1034} Duncan acknowledged, “fought to improve our schools,”\textsuperscript{1035} but none of them “had the resources to encourage innovation”\textsuperscript{1036} at the level being offered through RTTT. Race to the Top was “education’s moon shot,”\textsuperscript{1037} proclaimed Duncan, and it offered “a once-in-a-lifetime opportunity for the federal government to create incentives for far-reaching improvement in our nation’s schools.”\textsuperscript{1038} To ensure this,

\textsuperscript{1030} Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 176 (Cornell University Press 2012).

\textsuperscript{1031} H.R. 1, 11\textsuperscript{th} Cong. §800(4), (2009).

\textsuperscript{1032} Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 162 (Cornell University Press 2012).

\textsuperscript{1033} Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 176 (Cornell University Press 2012).

\textsuperscript{1034} Arne Duncan, The Race to the Top Begins—Remarks by Secretary Arne Duncan (July 24, 2009), \texttt{http://www.ed.gov/news/speeches}.

\textsuperscript{1035} Arne Duncan, The Race to the Top Begins—Remarks by Secretary Arne Duncan (July 24, 2009), \texttt{http://www.ed.gov/news/speeches}.

\textsuperscript{1036} Arne Duncan, The Race to the Top Begins—Remarks by Secretary Arne Duncan (July 24, 2009), \texttt{http://www.ed.gov/news/speeches}.

\textsuperscript{1037} Arne Duncan, The Race to the Top Begins—Remarks by Secretary Arne Duncan (July 24, 2009), \texttt{http://www.ed.gov/news/speeches}.

\textsuperscript{1038} Arne Duncan, The Race to the Top Begins—Remarks by Secretary Arne Duncan (July 24, 2009), \texttt{http://www.ed.gov/news/speeches}. 
Duncan specified, “we will be scrutinizing state applications for a coordinated and deep-seated commitment to reform.”\textsuperscript{1039} RTTT funds would be awarded based on states’ willingness to make “assurances”\textsuperscript{1040} centered around four “core reforms.”\textsuperscript{1041}

The first assurance a state would be required to provide was a commitment to adopt “standards and assessments that prepare students to succeed in college and the workplace and to compete in the global economy.”\textsuperscript{1042} Secretary Duncan would later deliver another speech—entitled “Beyond the Bubble Tests”\textsuperscript{1043}—when the $350 million of RTTT funding earmarked for the new assessments was awarded to two multi-state consortia. “Today is a new day,”\textsuperscript{1044} announced Duncan, “For the first time, many teachers will have the state assessments they have longed for—tests of critical thinking skills and complex student learning.”\textsuperscript{1045}

States receiving RTTT funding would be required to store and analyze data from these new standards-based assessments by “building data systems that measure student growth and

\begin{itemize}
  \item \textsuperscript{1039} Arne Duncan, \textit{The Race to the Top Begins—Remarks by Secretary Arne Duncan} (July 24, 2009), http://www.ed.gov/news/speeches.
  \item \textsuperscript{1040} Arne Duncan, \textit{The Race to the Top Begins—Remarks by Secretary Arne Duncan} (July 24, 2009), http://www.ed.gov/news/speeches.
  \item \textsuperscript{1041} Arne Duncan, \textit{The Race to the Top Begins—Remarks by Secretary Arne Duncan} (July 24, 2009), http://www.ed.gov/news/speeches.
  \item \textsuperscript{1042} Race to the Top Resources, United States Department of Education https://www2.ed.gov/about/initiatives/implementation-support-unit/tech-assist/index.html?exp=4, (last visited Feb. 21, 2017).
success, and inform teachers and principals about how they can improve instruction.”

States and districts were charged “recruiting, developing, rewarding, and retaining those “effective teachers and principals,” especially in areas of high poverty. RTTT states were also mandated to commit to “turning around our lowest-achieving schools.”

Duncan specified in his speech that the four core assurances were “interconnected.” One of the most significant initiatives required by states who wished to receive RTTT funds combined the first and second assurances. RTTT states were forbidden to have “any legal, statutory, or regulatory barriers at the state level to linking data on student achievement or student growth to teachers and principals for the purpose of teacher and principal evaluation.”

Not only were aspiring RTTT states directed to remove obstacles to the use of test scores for teacher evaluation, they were also mandated to develop a system through which they would “conduct annual evaluations of teachers and principals that include timely and constructive

---


feedback.” Evaluations in RTTT states were also required to “provide teachers and principals with data on student growth for their students, classes, and schools.”

The changes it demanded were significant, but the total Race to the Top allocation of $10 billion was “not chump change,” as Duncan had reminded his audience. “To every governor who ever aspired to be his state’s ‘education governor,’ I say, ‘do not let this unprecedented opportunity slip away.’” The race was on, and 41 states scrambled to comply with the assurances and qualify for the “first round” of funding to be awarded. Duncan and his staff kept their promise to “scrutinize” applications, and only two states—Delaware and Tennessee—secured funds when the first awards were announced in April 2010. Ten additional states, including New York and Florida, were granted awards via the second round in August 2010, and the Race to the Top commenced.

---


Before New York State’s application for RTTT funding, Education Law 3012-b forbade the granting or denial of tenure to a teacher based on student performance data.\footnote{N.Y. State United Teachers, 929 N.Y.S. at 702 (citing Former Education Law §3012-b (repealed July 1, 2010)).} In keeping with the expectations of Race to the Top, Education Law 3012-c repealed Education Law 3012-b effective May 28, 2010. Education Law 2013-c not only reversed the ban on the use of student achievement data for teacher evaluation and employment purposes but expressly required it.\footnote{N.Y. State United Teachers, 929 N.Y.S. at 702 (citing Educ. Law §3012-c(2)(a)).}

The revised legislation provided four rating categories for performance reviews: highly effective, effective, developing, and ineffective.\footnote{N.Y. State United Teachers, 929 N.Y.S. at 702 (citing Educ. Law §3012-c(2)(a)).} Each teacher or principal received a single composite score based on multiple measures of effectiveness to be established in the implementing regulations. The Commissioner of Education was given authority to establish score ranges for each performance level.\footnote{N.Y. State United Teachers, 929 N.Y.S. at 702 (citing Educ. Law §3012-c(2)(a)).}

Education Law 3012-c required that student achievement measures serve as a part of the evaluation scoring, comprising a total of 40 percent of a teacher or principal’s overall rating.\footnote{N.Y. Educ. Law §3012-c(2)(a).} Twenty to twenty-five percent was to be based on student growth data as measured by state-mandated standardized tests of reading and mathematics or comparable measures as needed for other subject areas.\footnote{N.Y. Educ. Law §3012-c(2)(a)(1)(i)).} The remainder of the student achievement measure, 15 to 20 percent, would be derived from data generated by locally selected
assessments. "Locally selected" assessments, while not defined specifically in Education Law 3012-c, are typically defined in assessment literature as those chosen by a local school district due to their alignment to critical skills, topics, and concepts. They can be created locally, but they can also include pre-existing commercial assessments selected at the district level. The defining aspect of “locally selected” assessments is that at a minimum the decision to use an instrument—if not the logistics of administration and scoring—is controlled at the local level. Education Law 3012-c added further specificity to the definition of “locally selected” assessments. The law mandated that the final decision to use an assessment for teacher and principal evaluation was to be made through the collective bargaining process.

Shortly after the passage of Education Law 3012-c, New York submitted an application for federal funding through Phase II of Race to the Top (RTTT). The United States Department of Education announced on August 24, 2010, that New York was to receive an RTTT award of $696,646,000.

---

1064 N.Y. Educ. Law §3012-c(2)(a)(1)(ii)).
1065 N.Y. Educ. Law §3012-c(2)(a)(1)(ii)).
1067 N.Y. Educ. Law §3012-c(2)(a)(1)(ii)).
1068 N.Y. Educ. Law §3012-c(2)(a)(1)(ii)).
1069 N.Y. State United Teachers, 929 N.Y.S. at 701.
Emergency Regulations in Response to Education Law 3012-c

Through Education Law 3012-c, the legislature directed the Board of Regents to develop by July 1, 2011, implementation regulations in collaboration with an advisory committee comprised of teachers and other school district staff. All performance reviews conducted after July 1, 2011, would be based on a system aligned with the new regulations. Although proof was submitted to the court showing that an advisory committee participated in drafting the regulations, the committee was unable to reach consensus within the time frame required to preserve the RTTT funding. As a result, emergency regulations written by the Board of Regents were adopted on May 16, 2011. The Board of Regents’ emergency regulations called for 40 percent of educators’ overall evaluation rating to be derived from a single state assessment system, with 20 percent from student growth on state assessments or comparable measures for subjects for which there is not a state assessment and 20 percent from student achievement as demonstrated by the same state assessments. The New York state testing system in 2011 consisted of the Regents’ Exams. At the elementary and middle school level, they assessed English language arts and math at grades three through eight and science at grades four and eight. At the high school level, they assessed 16 core English, science, social studies, and math courses along with six world languages.

---

1070 *N.Y. State United Teachers*, 929 N.Y.S. at 701 (citing Educ. Law §3012-c[7]).


1072 *N.Y. State United Teachers*, 929 N.Y.S. at 704.

Local school districts were required to respond to the regulations and adopt implementation plans by September 1, 2011. The statute required the use of collective bargaining to determine local implementation plans. However, ironically, staff would be on summer recess during the majority of this time.

New York Litigation in Response to Emergency Regulations

The New York State United Teachers (NYSUT), a public school teachers’ union, filed for a preliminary injunction in the New York State Supreme Court to prevent the Board of Regents from implementing components of the emergency regulations, including some pertaining to the use of student achievement data as part of teacher and principal evaluations. The New York State Supreme Court ruled NYSUT had standing to pursue the challenge against the Board of Regents due to the impact of the regulations on legally negotiated employment contracts in which the union had a vested interest. The court also clarified that local school districts were not parties to the litigation because the challenge was directed at the Board of Regents, a state-level governing body with authority over local districts.

The complaint did not center on Education Law 3012-c itself, but rather on the regulations written in response to the statute. Both parties acknowledged they participated in drafting Education Law 3012-c. The dispute specifically concerned the emergency regulations guiding implementation of staff evaluations during the 2011-12 school year. The

---

1074 N.Y. State United Teachers, 929 N.Y.S. at 701.
1075 N.Y. State United Teachers, 929 N.Y.S. at 702.
1076 N.Y. State United Teachers, 929 N.Y.S. at 701.
most salient issues pertained to whether the emergency regulations were in keeping with the legislative intent of the statute and whether the Board of Regents exceeded the scope of its authority. Logistically, the issue before the court was whether 40 percent of a teacher or principal’s overall evaluation rating could be based on a single state-mandated standardized testing system administered once per year to measure student achievement in reading and mathematics. More specifically, the dispute focused on whether the Board of Regents could dictate the use of data from state-mandated standardized tests as the second 20 percent of the overall rating that was reserved per the statute for locally selected and developed measures.

The key questions to be considered by the court pertaining to the intent of the legislature when it wrote the verbiage in the portion of the statute related to the use of student achievement data as part of teacher evaluation were 1) whether student achievement on state assessments could be considered a “locally selected measure” and 2) if it was acceptable to rate personnel as “ineffective” based solely on data from a single assessment when the statute referred to “multiple measures.” The latter could occur under the pending regulations because a failure to meet achievement and growth targets would virtually ensure an overall performance evaluation rating of “ineffective.” The Board of Regents’ emergency regulations would have used student growth on state-mandated standardized tests as 20 percent of a teacher or principal’s performance evaluation rating and student achievement on the same assessments as an additional 20 percent. Thus, if students did not meet targets for achievement and growth on the state-mandated assessments, their teacher and principal would receive a final evaluation score no higher than 60.

1077 N.Y. State United Teachers, 929 N.Y.S. at 703.
1078 N.Y. State United Teachers, 929 N.Y.S. at 704.
1079 N.Y. State United Teachers, 929 N.Y.S. at 704.
This would virtually ensure that a teacher or principal would be rated “ineffective” because the range for a rating of “ineffective” was set at a static score of 0-64 out of 100.\textsuperscript{1080}

The court granted NYSUT partial relief, ruling that certain aspects of the emergency regulations were invalid, including two provisions pertaining to the use of student achievement data in teacher evaluation. The two main points of legal consideration were the limitations placed on the power of the Board of Regents by state statutes and the incompatibility of the emergency regulations with the intent of the statute.\textsuperscript{1081} The court concluded the Board of Regents could not dictate the second 20 percent of teacher evaluation because that data was intended by the legislature to be a locally selected measure of student growth.\textsuperscript{1082} The court further stated student achievement could not constitute 40 percent of either a teacher or principal’s rating under the existing cut scores for the four proficiency levels. With the score range of 0-64 out of 100 denoted as “ineffective,” student achievement alone could result in the termination of a teacher or principal,\textsuperscript{1083} even if he or she received perfect scores on all other components. The court held this would contravene the multiple measures mandate of the statute.\textsuperscript{1084}

The court acknowledged the Regents possessed broad rule-making authority but were still subject to state law, citing \textit{Moore v. Board of Regents of the State of New York}, a case

\begin{footnotesize}
\begin{enumerate}
\item \textit{N.Y. State United Teachers}, 929 N.Y.S. at 706 (citing Educ. Law §3012-c[2] [a]).
\item \textit{N.Y. State United Teachers}, 929 N.Y.S. at 703.
\item \textit{N.Y. State United Teachers}, 929 N.Y.S. at 706 (citing Educ. Law §3012-c[2] [a]).
\item \textit{N.Y. State United Teachers}, 929 N.Y.S. at 706 (citing Educ. Law §3012-c[2] [a]).
\item \textit{N.Y. State United Teachers}, 929 N.Y.S. at 704.
\end{enumerate}
\end{footnotesize}
pertaining to the Board’s authority to review and evaluate doctoral programs.\footnote{1085} Moore provides a legal precedent for the limitations of the power of the Board of Regents relative to state law; in this case, Section 3012-\textit{c} of New York Education Law. The court also found the statute’s reference to “other locally selected measures of student achievement,” “developed locally,” and “multiple measures” were not a mere technicality but presented “a question of pure statutory reading and analysis dependent only on accurate apprehension of legislative intent.”\footnote{1086} In establishing this, the court cited Kurcsics v. Merchants Mutual Insurance, a case pertaining to the intent of the wording of insurance regulations,\footnote{1087} and Sbriglio v. Novello, addressing the distinction between “incorporation” and “construction” of new entities and the laws that govern them.\footnote{1088} Sbriglio is germane to the case because it supports the need for any implementing regulations to follow the intent of the statute as pertains to “locally selected” and “multiple measures.”\footnote{1089}

In outlining the merits of the case, the court highlighted specific tenets of Education Law 3012-\textit{c}. The court noted the salience of the fact that former Education Law 3012-b specifically forbade the use of student achievement data in tenure decisions.\footnote{1090} The court also underscored “other locally selected,” “developed locally,” and “other” for emphasis, with the caveat that the use of state assessments was not precluded entirely. However, it would be necessary to


\footnote{1086} \textit{N.Y. State United Teachers}, 929 N.Y.S. at 703.


\footnote{1088} Sbriglio v Novello, 845 NYS2d 147 (N.Y. Sup. Ct. 1978).

\footnote{1089} Sbriglio v Novello, 845 NYS2d 147 (N.Y. Sup. Ct. 1978).

\footnote{1090} \textit{N.Y. State United Teachers}, 929 N.Y.S. at 702 (citing Educ. Law §3012-b[2] [b]).
demonstrate how two data points from the same assessment system can serve as distinct measures, and most importantly, that their use be ratified as the second 20 percent through collective bargaining at the local level.\footnote{N.Y. State United Teachers, 929 N.Y.S. at 704 (citing Educ. Law §3012-c[2] [h]).} The court also emphasized the directive in Education Law 3012-c to use “multiple measures” to formulate a final evaluation score.\footnote{N.Y. State United Teachers, 929 N.Y.S. at 704 (citing Educ. Law §3012-c[2] [h]).} “Multiple measures” referred to the process defined in the statute that follows the widely-accepted practice of including in the final evaluation rating diverse elements such as classroom observations, documentation of professional activities, peer review, parent feedback, or student feedback to decrease subjectivity and increase validity and reliability.\footnote{Best Practices for Including Multiple Measures in Teacher Evaluations, hannoverresearch.com, http://www.hanoverresearch.com/wp-content/uploads/2012/05/Best-Practices-for-Including-Multiple-Measures-in-Teacher-Evaluations-Membership.pdf (last visited May 13, 2014).} The court indicated assigning 40 percent of the weight to student achievement was disproportionate, particularly given the “ineffective” range of 0-64 out of 100 that could result in an educator being terminated solely due to student achievement, regardless of performance in the other areas. The court held such a scenario would not be compatible with the “multiple measures” required by the statute because it does not allow a “meaningful impact” to be made by the 60 percent that includes all of the other measures except test scores.\footnote{N.Y. State United Teachers, 929 N.Y.S. at 704 (citing Educ. Law §3012-c[2] [a]).} Perhaps most significantly, the court rejected the State’s argument that an evaluation system must ensure educators actually improve student achievement.\footnote{N.Y. State United Teachers, 929 N.Y.S. at 706.}
“Ground-Breaking Agreement”

Within six months of the decision, on February 6, 2012, New York Governor Andrew M. Cuomo announced via press release that a compromise had been reached between New York State Education Commissioner John B. King and the New York State United Teachers, represented by President Richard C. Iannuzzi. The compromise met the criteria to preserve the $700 million in federal Race to the Top funding that had been imperiled by the court’s decision. The agreement maintained the practice of basing 40 percent of a teacher or principal’s overall effectiveness rating to student achievement data. The agreement included a change, however, which prevented a teacher or principal from being terminated based solely on test scores. The ceiling of the overall score range for “Ineffective” was lowered from 64 to 60. Under the revised parameters, a teacher or principal who received the full 60 points on other measures could not be terminated even if he or she received a score of zero for student achievement on standardized tests.

Another shift contained in the compromise addressed the intent of the statute for 20 percent of the teacher evaluation to be comprised of locally selected assessments. Local districts


were given three options: state tests, third-party assessments approved by the State Department of Education (SED), or locally developed tests, which are subject to approval by the SED.1099

New York Legislation to Address Unresolved Issues

One might reasonably assume an agreement that addressed two of the key issues cited in the original complaint and described in such glowing prose might portend the resolution of the issues surrounding the use of student achievement data for principal and teacher evaluation in New York. However, Aaron Lawson, staff attorney for the United States Court of Appeals for the Seventh Circuit, made a prescient assertion. Lawson posited that *NYSUT V. NYBOR*—the first state court opinion pertaining to the use of student achievement data to comply with Race to the Top—may be the start of a third wave of education litigation. The first two waves of lawsuits pertained to the adequacy and funding of education in general. This third wave, according to Lawson, will specifically address the conditions being applied to the acceptance of federal funding.1100 Lawson also raised the possibility that state legislatures may attempt to circumvent the state courts and federal regulations by contending it is the charge of state legislative branches to interpret state constitutions.1101

In hopes of supporting just such a states’ rights scenario, New York State United Teachers promoted a slew of state legislation under the umbrella of The Truth in Testing Act.

---

http://www.governor.ny.gov/press/02162012teacherevaluations


Four key tenets common to the suite of proposed legislation were cited on a Truth in Testing webpage linked to the NYSUT website: the obligation of the state to better prepare teachers and administrators for the rigor of the Common Core Standards, the urgent need to examine the amount of time and other resources being expended on assessment, the prohibition of non-diagnostic assessment in grades pre-kindergarten through second grade, and the protections that should be afforded students when student data are shared beyond the personnel who work directly with them.1102

This original Truth in Testing Act consisted of a suite of bills that addressed the four tenets. While the initial Truth in Testing legislation did not pass, it is instructive to review the changes made to the proposed legislation during its various iterations. Section 1 of proposed legislation A. 06593 would have amended Section 305 of New York Education Law by adding a new subdivision 43. This subdivision stated that regardless of any other laws or regulations, the commissioner would be prohibited from requiring the administration of state-developed or other similar assessments to students in kindergarten through second grade except for diagnostic purposes.1103 This limitation would have precluded the use of data from any state-selected or state-developed assessments in these grades as a measure for teacher evaluation because the purpose of a diagnostic assessment is solely to determine whether a student meets the criteria for a specific disability, condition, or intervention. Diagnostic assessments are not intended to measure student achievement or demonstrate academic growth.1104


1104 Steven Little & Angeleque Akin-Little, Academic Assessment and Intervention 166 (Taylor & Francis 2014).
Section 2 of the original Truth in Testing Act would have amended Education Law 3012 by adding a new Section 3012-b that would restrict the use of what is known in legal terms as “personally identifiable information.”

Personally identifiable information is defined in best practice recommendations as “any information which can be used to trace or distinguish an individual’s identity,” including students’ names, identification numbers, or dates of birth.

The aforementioned data points are among the elements that make it possible for schools and districts to combine multiple data sources. The restrictions in the proposed legislation would have rendered impossible tasks, such as merging the two years of trend data on a single assessment, that are needed to generate the growth scores that comprised part of the student achievement portion of the state’s teacher evaluation plan. In other words, proposed prohibitions would have effectively circumvented the directives outlined in a letter from the United States Department of Education that clarified the final regulations resulting from the Federal Education Rights and Privacy Act (FERPA). The aim of the letter was to improve “access to education data for research and accountability” by allowing “state educational authorities, such as SEAs and higher education commissions, [to] disclose education records in personally identifiable form,

---


without [parental] consent, to contractors, consultants, and other parties to whom they have outsourced organizational services or functions.”

The original Truth in Testing Act would have expanded the parental consent requirement to information that is currently routinely uploaded into the New York State Department of Education Database without parental permission. To obtain consent, school districts would have been directed to post the request on their website, send home via email, and make otherwise publicly available a detailed disclosure letter. The letter would have needed to include the following: a list of the data fields to be disclosed, the entity to which the disclosure will be made, the time frame during which it will be utilized and when it will be destroyed, the purpose of the project that requires the disclosure, and an explanation as to why the disclosure is necessary.

A mandate to obtain signed parental consent for the electronic data upload conducted by school districts would have rendered it nearly impossible for the State of New York to maintain its current data collection practices related to student assessment and achievement.

Outside of legal circles, “personally identifiable information,” where technology, education, and assessment intersect, is typically referred to by educators as student demographic data. Personally identifiable information often includes data elements such as date of birth and state ID number. Those two fields are typically used along with student names as the unique

---


Identifiers, or primary keys, that allow a student’s achievement to be tracked longitudinally. This facilitates merging different information from multiple sources in a database such as the longitudinal data warehouses described in Race to the Top. Without unique identifiers, it is impossible to merge one large data set pertaining to a group of students (e.g., demographic and enrollment records) with another (e.g., test scores). Maintaining trend data consisting of multiple years of test scores needed to generate growth data would also be impossible without a primary key to link records from different sources. When two sets of test scores from consecutive years are merged, for example, there might be 100 third graders named Matthew Johnson in a given state. The files could be combined to generate a growth score and attribute it to school, principal, and teacher only if the school and district are permitted to share dates of birth and student identification numbers so the database could use them as primary keys to differentiate among the different Matthew Johnsons.

Section 1 of Truth in Testing bill S. 04284 would have added a requirement for parental consent to sharing place of birth, a data point that is often used to track the progress of immigrant English Language Learners. Student passwords, that would be used to log onto computer-based online assessments, were included on the list of protected data requiring parent permission as well. Paired with the aforementioned state ID numbers, passwords need to be

---


housed in a state student information system and shared with third party vendors. Without this data sharing, not only would trend data not be available, but even the scoring of tests might be jeopardized. It would also preclude participation in the aforementioned computer-based online assessments, a format utilized for many 21st Century standardized tests.

In Section 2, of Truth in Testing bill S. 04284 also would have added a student’s “biometric record,” including handwriting samples, to the protected category of “personally identifiable information.” Since nearly all paper-based assessment systems entail scanning hand-written student work into a computer and linking it to a student’s ID number for scoring by personnel working in front of a screen, the inclusion of student handwriting in the category of protected personally identifiable information could preclude the use of paper-based standardized tests. In other words, by requiring written parent permission for each upload of the data elements currently routinely shared by districts with state student information systems and third-party vendors, the Truth in Testing Act would have rendered universal student participation in any standardized testing—paper or online format—next to impossible.

Section 3 of S. 04284 from Truth in Testing legislation would have also directed districts to provide assurance of compliance with state and federal safeguards, district and school rules regarding the use of data, the district or state’s indemnification for any violations that occur, an analysis of the risk of data breaches, and safeguards in place to prevent them. The resulting regulatory burden would have at least slowed—if not paralyzed—the use of state and third-party student information systems. It is important to note that these systems are used not only for

---


assessment, but also for the allocation of funding.\textsuperscript{1119} This is especially true in diverse districts that typically qualify for targeted assistance grants based on student demographics. For enforcement purposes, the Act would have called for the state comptroller to perform regular audits. Any entity found to be in violation would have been banned from using any student data for at least five years.\textsuperscript{1120}

Section 1 of yet another piece of Truth in Testing legislation, A. 06594, proposed an additional difficult, if not impracticable, set of requirements involving the gathering of data. Under this proposed law, the Commissioner of Education would have been required to undertake a massive and unprecedented study of assessments. The first data element—the total cost of developing, administering, and scoring assessments—would most likely have been relatively easy to compile if it involved only state assessments. However, it entailed not only state tests but also the cost of all locally and commercially developed assessments administered to each student cohort in every school and district in the state of New York. These data would have had to be disaggregated to include the amount of public funds paid to outside vendors the state, districts, and schools contracted for services. Costs for licensing, test development, materials, and scoring would also have needed to be denoted in detail.\textsuperscript{1121}

Not only would the state and each school and district have been required to report these costs, but they also would have had to report with specificity their personnel expenses for assessment coordination, data coaches, training for test proctors, and professional development


\textsuperscript{1120} S. 04284, S. 237th Sess. (N.Y. 2013) at §3.

\textsuperscript{1121} A. 06594, A. 237th Sess. (N.Y. 2013) at §1.
with a primary focus on data.\footnote{1122} This might generate an unbalanced picture of the cost of value-added teacher and principal evaluation in districts practicing Response to Intervention (RtI), Professional Learning Communities (PLCs), and other data-driven initiatives that intertwine with instruction but not the evaluation of personnel. For example, if a team of third grade teachers wanted to use data to improve math instruction, research has shown they would need professional development in high-level analysis, possibly using a data coach.\footnote{1123} Additional reporting required for the state, districts, and schools would have included denoting the cost of testing materials, test preparation materials, data warehouses or dashboards, and storage costs for materials. Disaggregation would have also applied to student subgroups, including the per-pupil cost of any additional assessments administered to English Language Learners or students with Individualized Education Plans.\footnote{1124}

Lastly, had Truth in Testing bill A. 08556 been passed, the state, districts, and individual schools would have been directed to calculate the total amount of time spent distributing materials, testing students, analyzing data, and what the Act defines as “preparing students for testing.”\footnote{1125} Given that any standards-aligned instruction might be construed as “preparing students for testing,” determining how to measure this item would be challenging at best. Under S. 05442, student time spent on testing would have been further disaggregated by districts into

\footnote{1122} A. 06594, A. 237th Sess. (N.Y. 2013) at §2.
\footnote{1124} A. 06594, A. 237th Sess. (N.Y. 2013) at §2.
\footnote{1125} S. 05540, A. 237th Sess. (N.Y. 2013) at §3.
student demographic groups, including general education, special education, English Language Learners, and low-income students.\textsuperscript{1126}

In addition to the aforementioned quantitative data, the commissioner would also have been tasked with adding a layer of qualitative data by conducting a state-wide anonymous survey of administrators and teachers that affords them anonymity. Teachers and administrators would have been asked to answer questions about time spent on testing. They would also have been asked to describe nebulous concepts such as the “impact of testing on curriculum and instruction” as well as provide suggestions for making assessment more efficient.\textsuperscript{1127} A wide-scale anonymous survey would inherently need to be delivered without a secured log-in, which would call into question the validity of the data due to the possibility of distortion by users who respond more than once.

It is perhaps due to some of the more extreme of the aforementioned measures that none of the bills included in the original Truth in Testing Act made it past the committee level.\textsuperscript{1128} The last action on any of the legislation took place in late April of 2014, when A. 07442 was referred to committee,\textsuperscript{1129} meaning that none of the Truth in Testing bills ever reached the State House or Senate floor for debate, let alone a vote.

\textsuperscript{1126} S. 05540, S. 237th Sess. (N.Y. 2013) at §3.

\textsuperscript{1127} S. 05540, S. 237th Sess. (N.Y. 2013) at §3.

\textsuperscript{1128} A. 07442, A. 237th Sess. (N.Y. 2013).

Legislators Ask for Delay

The lack of action on the initial Truth in Testing legislation did not indicate, however, that the conflict over the implementation of value-added teacher evaluation in New York was finished. In spite of the fact that the plaintiff in the initial litigation was a teachers’ union, both political parties united to advocate for further relief. Even before the original Truth in Testing Act stalled at the committee level, New York State Assembly and Senate leaders used the media to move the issue forward. On February 4, 2014, Senate Co-Leaders Dean Skelos and Dean Klein, along with Senate Education Chair John Flanagan, issued a joint press release calling for a two-year delay of the use of Common Core tests for evaluating teachers, principals, and students due to their “grave concerns about this flawed roll-out” of Race to the Top, Common Core Standards, and new assessments linked to the Common Core. They also expressed “serious concerns” about the release of student, teacher, and principal information collected on the state Education Data Portal (EDP). Assembly Speaker Sheldon Silver and Assembly Education Chair Cathy Nolan released a similar statement. Governor Andrew Cuomo acknowledged that the implementation was flawed but called for circumspection, promising to assemble a panel


of legislators and educators tasked with identifying problems and taking corrective action by the end of the legislative session.\textsuperscript{1133}

Second Wave of Legislation Emerges

Just over a month after the aforementioned press release, a dark horse emerged and quickly bypassed earlier legislation lying dormant in the Education Committees of both state houses. Within a week of being introduced, the first post-Truth in Testing bill, A. 08929, was passed by the New York State Assembly on March 5, 2014, by a vote of 121 to 10. While this legislation was not specifically cited on the NYSUT Truth in Testing webpage at the time, it not only addressed the aforementioned four tenets but also managed to combine all of them into a single piece of legislation.

Section 1 of proposed A. 08929 would have amended Section 3012-c of the education law by adding a new subdivision, 2-a, prohibiting scores from any standardized assessments aligned to Common Core English language arts or mathematics from being used as the state or locally selected assessments applying toward a teacher or principal’s composite evaluation effectiveness score.\textsuperscript{1134} A. 08929 differed from the original Truth in Testing legislation that inspired it. The original Truth in Testing legislation flouted the authority of the federal government by mandating actions “notwithstanding any provision of law or regulation to the contrary.”\textsuperscript{1135} A. 08929 took a more realistic, if still not financially pragmatic, approach. To

\textsuperscript{1133} Jamie Franchi, Pols Call for Common Core Delay, Cuomo Urges Patience, The Long Island Press (February 5, 2014, 3:52 PM), http://www.longislandpress.com/2014/02/05/pols-call-for-common-core-delay-cuomo-urges-patience/.

\textsuperscript{1134} A. 08929, A. 237th Sess. (N.Y. 2013) at §1.

\textsuperscript{1135} A. 06593, A. 237th Sess. (N.Y. 2013) at §1.
render the provisions of Section 1 feasible, Section 2 of A. 08929 would have directed the New York State Commissioner of Education to apply for and obtain any federal waivers deemed necessary.\footnote{1136}{A. 08929, A. 237th Sess. (N.Y. 2013) at §2.}

The federal regulations from which the commissioner would have been asked to seek relief included the Elementary and Secondary Education Act of 1965, its reauthorization through No Child Left Behind in 2001, and the most recent iteration via Executive branch regulations, Race to the Top, the latter of which provided the State of New York with an award of $696,646,000.\footnote{1137}{N.Y. State United Teachers, 929 N.Y.S. at 701.} In light of this directive, it is not surprising that the fiscal impact of A. 08929 was denoted in the memo section on the legislative tracking function on the Assembly website as “to be determined.”\footnote{1138}{A. 08929, A. 237th Sess. (N.Y. 2013).}

Section 3 of A. 08929 stretched beyond the provisions of the original union-sponsored legislation in some respects. It would have amended Section 305 of the Education Law by adding four new subdivisions: 44, 45, 46, and 47. Subdivision 44 would have prohibited school districts from making student placement or promotion decisions based solely or primarily on state assessment scores, directing them instead to use multiple measures if test scores are a factor in the decision-making process. The additional Subdivision 45 would have forbidden districts and schools from including any state test scores on student transcripts.\footnote{1139}{A. 08929, A. 237th Sess. (N.Y. 2013) at §3.} While the former simply reflects what many educators may already consider to be best practice, the latter could have proven to be a disincentive for students, especially at the high school level. According to
research conducted by Steven D. Levitt, the William B. Ogden Distinguished Service Professor of Economics at the University of Chicago, “incentives matter” to high school students and are correlated with higher achievement on standardized tests.\textsuperscript{1140} Due to reduced student motivation, removing the incentive of a passing test score as a graduation requirement could in turn could make it more difficult for schools to obtain accurate achievement data, leading to difficulties determining accurate student proficiency levels either for the design of instruction or the evaluation of teachers and principals.

Subdivision 46 of the first post-Truth in Testing bill would have directed the commissioner to reduce field tests,\textsuperscript{1141} un-scored test items used to gather data for future tests,\textsuperscript{1142} for students taking new Common Core assessments in English language arts and math.\textsuperscript{1143} This was likely in response to the 2014 field tests for the new Partnership for Readiness for College and Careers (PARCC) assessments. PARCC field tests required up to 15.5 hours of testing time per student at some grade levels,\textsuperscript{1144} a duration significantly longer than traditional standardized tests of reading and mathematics. The commissioner would also have been directed to make “significantly more sample test questions” available to teachers and develop a minimum of 20 test forms for each assessment.\textsuperscript{1145} The additional sample questions and test forms would have


\textsuperscript{1141} A. 08929, A. 237th Sess. (N.Y. 2013) at §3.

\textsuperscript{1142} A. 07442, A. 237th Sess. (N.Y. 2013) at §2.

\textsuperscript{1143} A. 08929, A. 237th Sess. (N.Y. 2013) at §3.


\textsuperscript{1145} A. 08929, A. 237th Sess. (N.Y. 2013) at §3.
significantly increased the cost of testing and the length of testing sessions. In addition, the process of creating them would be incompatible with the directive to reduce field testing time, since the development of additional forms would require the field testing of more items. The Commissioner also would have been directed by proposed legislation A. 08929 to create training materials and design professional development in collaboration with teachers, principals, and other stakeholders for every Common Core grade level and subject.\footnote{A. 08929, A. 237th Sess. (N.Y. 2013) at §3.}

The authors of the legislation appear to have recognized there would be significant costs associated with these activities. To fund the mandates denoted in Section 3, the Commissioner would have been directed in Section 4 to use a portion of the $348,323,000 federal grant funds received under the state fiscal stabilization fund of the American Recovery and Reinvestment Act of 2009 (ARRA).\footnote{A. 08929, A. 237th Sess. (N.Y. 2013) at §4.} Given the condition outlined in Section 4, not only would Section 3 have possibly delayed standardized testing aligned to the Common Core, but it also could have diverted resources needed to use student achievement data as a component of teacher and principal evaluation. A portion of the second disbursement of ARRA funds to education in New York was earmarked for the development of growth metrics, design of a longitudinal data system, and revision of teacher and principal evaluation systems.\footnote{New York to Receive More than 814 Million in Additional Recovery Funds, United States Department of Education, http://www.ed.gov/news/press-releases/new-york-receive-more-814-million-additional-recovery-funds (last visited March 26, 2014).}

Like the original Truth in Testing legislation, the first post-Truth in Testing bill, proposed legislation A. 08929, also would have addressed standardized testing in kindergarten, first grade, and second grade, but in a less restrictive manner. Section 5 would have required the
Commissioner to reject any Annual Professional Performance Review Plan (APPRP) that included the administration of state-developed assessments or non-diagnostic standardized assessments unless the administration is required by federal law. Unlike S. 05442, however, A. 08929 stated districts not only may administer locally developed assessments at grades K-2 for non-diagnostic purposes but may also use them for the evaluation of teachers and principals. “Locally developed” test was defined by A. 08929 as rigorous assessments consistent across classrooms that are developed by a district, Board of Cooperative Education Services (BOCES), or a regional group of districts. A. 08929 would have restored some of the accountability for student performance in grades K-2 that would have been impossible under the earlier legislation. Another less restrictive component of A. 08929 was that it explicitly clarified that “nothing in this subdivision shall preclude the approved assessments that are administered to students in higher grades if otherwise allowed in the section and allowed under state regulations.” Addressing one of the key legal issues from NYSUT v. NYBOR, this legislation would have delegated the authority to revise rejected Annual Professional Performance Reviews (APPRs) to the collective bargaining process at the local district level.

Of course, much of the above would have required local school districts to revise their 3012-c-aligned Annual Professional Performance Review Plans, which in New York must be approved by the Commissioner of Education. A. 08929 sought to guarantee the expediency of

the approval process for any district pursuing revisions categorized as the “elimination of unnecessary student assessments.” 1154 Section 10 not only would have required the Commissioner to expedite the review of changes by focusing the review solely on the portion pertaining to the reduction of student assessments; it also would have directed him or her to provide assistance to districts during the process. 1155

As was the case with assessment in the primary grades, A. 08929 would have upheld some of the restrictions pertaining to the data infrastructure needed to produce the trend data used for value-added measures. However, it was less restrictive in some aspects than the original Truth in Testing Act posted in the NYSUT website. Like the previous bills, the definition of protected student data would have been broadened to encompass data points that have historically been routinely shared in the process of administering student assessments and tracking student achievement. 1156 Unlike the earlier legislation, however, A. 08929 simply would have rendered it more difficult, but still feasible, for districts to share data with the state and third-party providers. Rather than requiring signed permission for each student each time the data are shared, this legislation would have directed that parents be provided with the freedom to “opt out,” placing the burden on the parents to sign and return the form. 1157 Assuming at least some parents would elect not to share their child’s information, it would have prevented most districts from obtaining a full data set.

As it pertains to one issue, A. 08929 would have gone a step further than the original Truth in Testing Act. Beyond having to study the assessments administered to English Language Learners (ELLs) and students with Individualized Education Plans (IEPs), the commissioner would also have been directed to study the implementation of assessment accommodations for these students at the district level.\textsuperscript{1158}

The list of sponsors for A. 08929, the first post-Truth in Testing bill, was impressively bipartisan. The Assembly vote of 121 to 10 gave the appearance of immunity to a potential veto by New York Governor Andrew Cuomo, who “denounced changes to the teacher and principal evaluation system he championed” and formed his own panel to study the issues.\textsuperscript{1159} Bipartisan support in the Assembly, however, did not guarantee A. 08929 would become law. New York’s legislative history suggests “uni bills,” those with a sponsor in both the Assembly and the Senate, are given the most serious consideration.\textsuperscript{1160} As of March 5, 2014, A. 08929 did not have a Senate sponsor or equivalent version in the Senate, and the chair of the Senate Education Committee, Republican Senator John Flanagan, stated his conference “would not support the bill as is.”\textsuperscript{1161} A bicameral compromise remained possible, but whether it would have retained the

\textsuperscript{1158} A. 08929, A. 237th Sess. (N.Y. 2013) at §14.

\textsuperscript{1159} Jessica Bakeman, \textit{Assembly Approves a Common Core Delay}, The Capital New York, \url{http://www.capitalnewyork.com/article/albany/2014/03/8541376/assembley-approves-common-core-delay} (last visited March 5, 2014).

\textsuperscript{1160} \textit{New York Research In-Depth}, Georgetown Law Library, (December 2010), \url{https://www.law.georgetown.edu/library/research/guides/newyork-in-depth.cfm#a-legislation}.

\textsuperscript{1161} \textit{New York Research In-Depth}, Georgetown Law Library, (December 2010), \url{https://www.law.georgetown.edu/library/research/guides/newyork-in-depth.cfm#a-legislation}.

\textsuperscript{1162} \textit{New York Research In-Depth}, Georgetown Law Library, (December 2010), \url{https://www.law.georgetown.edu/library/research/guides/newyork-in-depth.cfm#a-legislation}.
same tenets and have garnered a veto-proof vote similar to that attained by A. 0 8929 will never be known.

Governor Andrew Cuomo decided not to take the risk of waiting to find out if a Senate counterpart to A. 08929, the first post-Truth in Testing legislation, would emerge. In late June 2014, facing an election for which Common Core testing had become an issue followed closely by both teachers’ unions and parents, Cuomo announced that he and state legislators had reached a tentative agreement to delay for two years the use of student scores on Common Core assessments to evaluate teachers who were rated “ineffective” or “developing.” While this most likely pleased NYSUT and the parents who supported them due to perceived over-testing of their children, other parents, represented by the advocacy group Students First New York, indicated that they were not happy with the agreement that they felt would provide a “safety net” for underperforming teachers.

As the gubernatorial election drew near its close in late October, Andrew Cuomo’s campaign released a video advertisement that stated that he wanted to stop “over-testing” students and delay the use of Common Core test scores for at least five years or “until the children are ready.” The Cuomo campaign did not respond to media requests for clarification.


Florida’s Statutory Response to Race to the Top

Like New York, the State of Florida was also awarded a Race to the Top grant in August 2010. Florida’s award was slightly higher, totaling $700,000,000.1167 As a result, Governor Rick Scott signed Florida Senate Bill (SB) 736, popularly known as the Student Success Act, into law on March 24, 2011.1168 SB 736 made significant changes to teacher and administrator evaluation process codified in Chapter 1012 of the Florida Statutes.1169

Under the modified statute, public school districts were required to annually evaluate both teachers and administrators1170 using one of four performance ratings: highly effective, effective, needs improvement (“developing” for teachers in the first three years of employment), and unsatisfactory.1171 The new law required at least 50 percent of each performance evaluation to be based on state assessment data. For teachers of subjects or grades not measured by state assessments, school officials were provided the option of using local assessment data.1172 The aforementioned test scores had to reflect growth as measured by at least three years of trend data for students assigned to a teacher.1173 If three years of growth data were not available, school officials were mandated to use “any available data,”1174 with the percentage of the evaluation

reflecting student growth comprising no less than 40 percent of the teacher’s final performance rating.\footnote{Fla. Stat. §1012.34(3)(a)1-b (2011).} For instructional personnel not serving as either classroom teachers or administrators, student growth data could be reduced to no less than 30 percent, provided “other measurable student outcomes that are specific to the assigned position”\footnote{Fla. Stat. §1012.34(3)(a)1-b (2011).} comprised the remaining 20 percent, for a total of 50 percent. If less than three years of trend data were available for non-classroom instructional personnel, the percentage attributed to student growth could be set at no less than 20 percent.\footnote{Fla. Stat. §1012.34(3)(a)1-b (2011).} Another salient change was the fact that beginning on July 1, 2014, school boards were required to use a teacher’s overall performance evaluation rating as the partial basis of determining teacher compensation.\footnote{Fla. Stat. §1012.22(1)(c)4-5(2011).}

In 2011, Florida’s state test was the Florida Comprehensive Assessment Test (FCAT) 2.0, the latest iteration of a system in use since 1998.\footnote{History of Florida Statewide Assessment, Florida Department of Education, http://www.fldoe.org/accountability/assessments/k-12-student-assessment/history-of-fls-statewide-assessment/fcat/ (last visited July 9, 2015).} The 2011 FCAT was designed to measure student performance on the Next Generation Sunshine State Standards in English language arts in grades 3 through 10, math in grades 3 through 8, and science in grades 5 and 8.\footnote{Cook v. Stewart, No.1:13-cv-000072-MW-GRJ, (N.D.Fla. May 17, 2013) at 11.} Rather than focusing solely on the percentage of students meeting or exceeding standards, as was often the case under No Child Left Behind, Florida’s Race to the Top statute prescribed the use of a value-added model to configure student growth data from the FCAT for
accountability. The model was to be developed by the commissioner no later than June 1, 2011.\footnote{Fla. Stat. §1012.34(7)(a)(2011).}

While the statute did not describe the specific formula to be used, it did define some parameters. The formula was required to utilize prior student performance as baseline data.\footnote{Fla. Stat. §1012.34(7)(a)(2011).} The commissioner was also directed to consider factors that might impact student achievement, including attendance, disability diagnosis, and English Learner status.\footnote{Fla. Stat. §1012.34(7)(a)(2011).} However, not all factors were permitted to be considered as co-variants for the future formula. The statute specifically forbade the commissioner from setting different expectations based on student gender, race, ethnicity, or socioeconomic status.\footnote{Fla. Stat. §1012.34(7)(a)(2011).}

The statute required each school district to use the commissioner’s growth formula beginning in the 2011-12 school year for any courses associated with FCAT.\footnote{Fla. Stat. §1012.34(7)(b)(2011).} For grades and subjects not assessed by statewide tests, each school district was required to measure student growth beginning in the 2014-15 school year using “an equally appropriate formula.”\footnote{Fla. Stat. §1012.34(7)(b)(2011).} To this end, the Florida Department of Education (FDOE) was directed by the statute to “provide models for measuring student learning growth which school districts may adopt”\footnote{Fla. Stat. §1012.34(7)(b)(2011).} by 2014.
Each school district was required to report the results of its new evaluation system to the Florida Department of Education (FLDOE), which in turn was charged with approving each school district’s evaluation system and monitoring the systems for compliance. A temporary measure was included in the statute to bridge the gap between the 2011 FCAT-only model and the anticipated release of additional state growth models. Until July 1, 2015, the statute permitted school districts to use the aforementioned evaluation system approval process to request modifications to the growth model requirement for teachers of courses not measured by statewide assessments. One permissible modification was to request the use of student achievement data as a substitute for growth data or in combination with growth data “if achievement is demonstrated to be a more appropriate measure of classroom teacher performance.”

The statute allowed school districts to request another modification for courses that were measured by district assessments but not by state tests. School districts were permitted to include student growth on FCAT Reading or FCAT Mathematics but give greater weight to student learning growth on the school district assessment, so long as school officials were able to “clearly explain the rationale supporting the request.” A third exception was available for the rating the performance of teachers of courses not aligned to FCAT through the evaluation system.

---

approval process. A school district was allowed to ask to use student growth on “measurable learning targets”\textsuperscript{1194} that were “established based upon the goals of the school improvement plan and approved by the school principal.”\textsuperscript{1195} A district superintendent was able to request a fourth exception permitting him or her to use the aggregate scores of an instructional team for the student growth rating of each teacher on the team.\textsuperscript{1196}

The statute directed the State Board of Education to adopt rules and uniform procedures for the submission, review, and approval of school district evaluation systems and reporting requirements.\textsuperscript{1197} This review included ensuring if student learning growth standards were not met, an employee would receive an unsatisfactory performance evaluation rating.\textsuperscript{1198} School officials were also charged with developing a procedure to allow school staff to review the rosters of students for whom they were responsible for accuracy and to correct any errors in the student data.\textsuperscript{1199}

Florida’s Value-Added Formula

In response to the Student Success Act and the resulting statutory changes, Florida Commissioner of Education Eric J. Smith charged the Florida Student Growth Implementation Committee (SGIC) with recommending an initial growth model for FCAT Reading and FCAT

\begin{footnotesize}
\begin{enumerate}
\item \textsuperscript{1194} Fla. Stat. §1012.34(7)(e)(2011).
\item \textsuperscript{1195} Fla. Stat. §1012.34(7)(e)(2011).
\item \textsuperscript{1196} Fla. Stat. §1012.34(7)(e)(2011).
\item \textsuperscript{1197} Fla. Stat. §1012.34(8)(2011).
\item \textsuperscript{1198} Fla. Stat. §1012.34(8)(2011).
\item \textsuperscript{1199} Fla. Stat. §1012.34(8)(2011).
\end{enumerate}
\end{footnotesize}
Math data by the June 1, 2011 deadline outlined in the statute. The Florida Department of Education established the SGIC in December 2010. Upon SGIC’s creation, the Commissioner of Education appointed 27 individuals to the SGIC Board for four-year terms. Committee members included teachers, central office administrators, a university professor, parents, a business person, a union leader, and teachers from the elementary and secondary levels who represented “Florida’s diversity in culture, community, and region.” Sam Foerster, the Putnam County associate superintendent of schools, chaired the committee. In addition to chartering the committee, the FLDOE awarded a contract to the American Institutes for Research (AIR) to assist the SGIC with the development, evaluation, and implementation of a value-added model (VAM). In 2011, the SGIC’s role, in partnership with AIR, was to make a recommendation—not a final decision—for a value-added model to be used for the evaluation of teachers of reading and math courses assessed with the FCAT. To accomplish this, the

---


committee met both face-to-face and online in April and May of 2011. During the subsequent four years, the plan was for the committee to continue to provide feedback and additional recommendations to the FLDOE.

The SGIC’s first act was to discuss an AIR presentation about value-added models currently in use in the field of education. All of the proposed models could apply controls to various contextual factors. This was important because the placement of students in classes and the assignment of students to teachers was not random. Instead there were “purposive selection mechanisms” contributing to student placement in classes and student assignment to teachers. These mechanisms could include parent selection of schools, teacher selection of schools and subjects, parent selection of teachers, and principal discretion in assigning students to courses, classes, or teachers. While the updated Florida Statute forbade control for student gender, race, ethnicity, or socioeconomic status as variables impacting student growth, the SGIC was afforded the liberty of considering the impact of controlling for other contextual variables, such

---


as Students with Disabilities (SWD) status, gifted status, English language learner (ELL) status, and student attendance.\textsuperscript{1214}

To select a VAM, the SGIC considered two categories of growth models: typical learning path models and covariate adjustment models.\textsuperscript{1215} Both models differed on whether the effect of teachers on student growth was fixed or random. Researchers who posited a teacher’s effect on student growth was fixed believed the variance in growth can be attributed primarily to the teacher’s efficacy.\textsuperscript{1216} Those who claimed a teacher’s effect on student growth was random believed the population from which the students in the teacher’s class were drawn was the primary determinant of student growth and the performance of teachers varied primarily based on the traits of the students in their classes.\textsuperscript{1217}

Typical learning path models, also known as general longitudinal mixed-effects models, operate under the assumption that each student in a data set has a “typical learning path,”\textsuperscript{1218} or propensity to achieve, relative to the state average. There are three key traits endemic to typical learning path models. They do not directly control for prior achievement\textsuperscript{1219} (e.g., pre-test or


\textsuperscript{1215} Recommendations of the Florida Student Growth Implementation Committee, 6, (June 1, 2011), http://www.fldoe.org/core/fileparse.php/7503/urlt/0102687-value-added-model-white-paper.doc.

\textsuperscript{1216} Michael Borenstein, L.V. Hedges, a d H.R. Rothstein, Meta Analysis: Fixed Effects versus Random Effects 6, (Borenstein, Hedges, Rothstein, 2007), http://www.meta-analysis.com/downloads/Meta-analysis percent20fixed percent20effect percent20vs percent20random percent20effects.pdf.


\textsuperscript{1219} Recommendations of the Florida Student Growth Implementation Committee, 6, (June 1, 2011), http://www.fldoe.org/core/fileparse.php/7503/urlt/0102687-value-added-model-white-paper.doc.
prior year scores), they require multiple years of data, and they make adjustments to a student’s
tendency to achieve as additional years of data are compiled.\textsuperscript{1220} As each additional year of data
becomes available for a student, a more accurate prediction can be made regarding his or her
propensity to learn. An underlying assumption of this model is that teachers can alter the learning
paths of their students.

There are numerous variations within the category of typical learning path models. One
characteristic that varies among these different iterations is the assumption about the
“durability”\textsuperscript{1221} (duration) of a teacher’s impact on a student’s typical learning path.\textsuperscript{1222} Some
researchers argue that an individual teacher’s impact on a student’s achievement is permanent,
while others posit the impact is limited to as few as two school years.\textsuperscript{1223} Unlike typical learning
path models, covariate adjustment models directly control for prior student scores using pretest
or prior-year scores as a baseline for establishing growth.\textsuperscript{1224}

In April 2011, the American Institutes for Research presented eight models to the
SGIC.\textsuperscript{1225} These models varied across four domains: the use of statistical controls for factors
believed to be outside the control of teachers, assumptions about the durability of teacher impact

\textsuperscript{1220} Recommendations of the Florida Student Growth Implementation Committee, 6, (June 1, 2011),

\textsuperscript{1221} Recommendations of the Florida Student Growth Implementation Committee, 6, (June 1, 2011),

\textsuperscript{1222} Recommendations of the Florida Student Growth Implementation Committee, 6, (June 1, 2011),

\textsuperscript{1223} Recommendations of the Florida Student Growth Implementation Committee, 6, (June 1, 2011),

\textsuperscript{1224} Recommendations of the Florida Student Growth Implementation Committee, 6, (June 1, 2011),

\textsuperscript{1225} Recommendations of the Florida Student Growth Implementation Committee, 4, (June 1, 2011),
on student growth, the unit of measurement used to represent student achievement, and the statistical model employed to calculate value added.\textsuperscript{1226} The first four models followed typical learning path protocols, and the last four employed covariate adjustment formulas.\textsuperscript{1227}

Model 1 was a typical learning path model similar to William Sanders’ Tennessee Value-Added Assessment System (TVAAS). In terms of durability, Model 1 asserted teachers have a permanent impact on students, a school of thought often described as a “layered”\textsuperscript{1228} model due to the belief in the cumulative effect of prior teachers on student growth. Teacher effects are statistically modeled as random and assumed to be permanent in terms of durability.\textsuperscript{1229} No controls are applied for contextual variables,\textsuperscript{1230} factors believed by many to be outside the teacher’s control.\textsuperscript{1231} Similar to six of the other models, Model 1 used an interval scale (i.e., test scores) as the unit of measurement for student growth.\textsuperscript{1232}

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{1226} Recommendations of the Florida Student Growth Implementation Committee, 5, (June 1, 2011), http://www.fldoe.org/core/fileparse.php/7503/urlt/0102687-value-added-model-white-paper.doc.
\item \textsuperscript{1227} Recommendations of the Florida Student Growth Implementation Committee, 13, (June 1, 2011), http://www.fldoe.org/core/fileparse.php/7503/urlt/0102687-value-added-model-white-paper.doc.
\item \textsuperscript{1228} Recommendations of the Florida Student Growth Implementation Committee, 8, (June 1, 2011), http://www.fldoe.org/core/fileparse.php/7503/urlt/0102687-value-added-model-white-paper.doc.
\item \textsuperscript{1229} Recommendations of the Florida Student Growth Implementation Committee, 9, (June 1, 2011), http://www.fldoe.org/core/fileparse.php/7503/urlt/0102687-value-added-model-white-paper.doc.
\item \textsuperscript{1230} Recommendations of the Florida Student Growth Implementation Committee, 9, (June 1, 2011), http://www.fldoe.org/core/fileparse.php/7503/urlt/0102687-value-added-model-white-paper.doc.
\item \textsuperscript{1231} Recommendations of the Florida Student Growth Implementation Committee, 5, (June 1, 2011), http://www.fldoe.org/core/fileparse.php/7503/urlt/0102687-value-added-model-white-paper.doc.
\item \textsuperscript{1232} Recommendations of the Florida Student Growth Implementation Committee, 9, (June 1, 2011), http://www.fldoe.org/core/fileparse.php/7503/urlt/0102687-value-added-model-white-paper.doc.
\end{itemize}
\end{footnotesize}
Model 2 was similar to the RAND Corporation model developed by McCaffrey and Lockwood in 2008. Model 2 also shared traits with Model 1 in that teacher effects were modeled as random, no contextual control variables were used, and student achievement was measured via interval scales. Often called a “variable persistence model,” Model 2 differed from Model 1 in that the impact of an individual teacher was thought to decay over time.

Model 3 was a hybrid model. It was nearly identical to Model 2 in that it did not control for contextual variables, used a typical learning path statistical model, and applied test score intervals as the unit of measurement for student growth. It differed, however, in that it estimated teacher effects as fixed rather than random. Model 4 was also a hybrid model similar to Model 2. Model 4 used a typical learning path statistical model, measured growth using test score intervals as units, and modeled teacher effects as random. However, unlike Model 2, Model 4 controlled for contextual variables.

Model 5 was the first of the covariate adjustment models presented by AIR for SGIC’s consideration. Model 5 was similar to a model designed by Robert H. Myer in 1992 and later

---


updated in 2010. Like all covariate adjustment models, Model 5 did not need to account for any beliefs on the durability of teacher effects due to its use of prior student achievement as a variable. Teacher effect on student growth was modeled as random. Contextual student characteristic variables were also considered as part of Model 5, and test score intervals were used as the unit to measure student growth.\textsuperscript{1239} Model 6 was identical to Model 5 except that it excluded contextual variables from its formula.\textsuperscript{1240}

Model 7, otherwise known as the Differences Model, was a variation of a model implemented by AIR in Florida on behalf of the Foundation for Excellence in Education and its Excellence in Teaching Awards. It was described by AIR as the least complex and therefore the most transparent of the eight models. Because it was a covariate adjustment model, the durability of teacher effects did not need to be taken into consideration. Like most of the other models, it used test score intervals as the unit to measure growth. Unlike Models 5 and 6, Model 7 limited the weight of prior test scores to a coefficient of one\textsuperscript{1241} due to the fact that FCAT scaling resulted in abnormally large shifts in the mean performance of students at certain grades, which could result in significant statistical errors.\textsuperscript{1242}

Model 8, similar to a model used in Colorado, also sought to compensate for the uneven scaling of FCAT cut scores and the resulting statistical errors. This was accomplished by using

\begin{thebibliography}{9}


\end{thebibliography}
in-state percentile ranks,\textsuperscript{1243} norming students against their peers,\textsuperscript{1244} rather than on vertically scaled performance levels.\textsuperscript{1245} Model 8 was similar to all of the covariant adjustment models in that the durability of teacher effects did not need to be taken into consideration and prior student performance, in this case previous student percentile rank, was used to predict student growth. Model 8 did not control for any contextual variables.\textsuperscript{1246}

After the presentation of the eight value-added models, the SGIC discussed each model and identified concerns about the implementation of particular models.\textsuperscript{1247} The committee cited several issues associated with the typical learning path models. Generally, committee members were not comfortable with the “abstract nature of the control for student achievement”\textsuperscript{1248} without controlling for prior achievement. The committee also voiced trepidation about the impact of ongoing revisions to a student’s typical learning path over time.\textsuperscript{1249} Under such a scenario, a teacher could hypothetically be dismissed for low student growth, only to find a year


later that his or her student growth expectations had been lowered due to additional data. The SGIC also noted the imperfect measurement scales of FCAT and expressed concern over the heavy reliance on the scales under the typical learning path model.

When discussing the covariate adjustment models, the committee did not favor Model 8, the percentile rank formula. The committee cited its “inherently normative nature,” that suggested some students would always have to fail regardless of their actual proficiency as defined by scale scores. This, in turn would preclude the possibility of all teachers meeting the standard, even if 100 percent of students in the state met the proficiency cut scores.

Before the models could be tested, the SGIC needed to identify the contextual variables to be controlled during some of the simulations. In addition to the statute’s directive not to control for gender, race, ethnicity, or socioeconomic status, the SGIC added conditions of its own. For example, contextual variables used as controls in the value-added models are not within a teacher’s control and had to not be measureable by another variable (e.g., pre-test scores). As a result, several potential variables initially generated by SGIC members during brainstorming were eliminated.

---


Homelessness, for example, was found to be too closely linked to the statutorily forbidden socioeconomic status subgroup. The committee also noted migrant status could be measured within a school’s mobility. Variables such as school resources, school grades (state ratings), Response to Intervention (RtI) level, foster care status, and rural school status, although found to be valid, were eliminated from immediate consideration due to sufficient data not being available. Other factors, such as teacher experience and course complexity, were theoretically available but were either difficult to quantify for use in evaluation or inconsistently reported. Some committee members suggested student disciplinary data fell into the inconsistently reported category, but other committee members argued this was a variable within a teacher’s control. Homework performance was eliminated from the variable list because committee members believed it was subject to some degree of teacher control. While the committee included teacher attendance in the initial set of variables, the committee did not indicate this variable was under a teacher’s control. Instead the committee requested AIR study of the impact of teacher attendance on student growth ratings.\textsuperscript{1256}

After extensive discussion, the SGIC approved a list of nine contextual student characteristic variables for use in the model study. These variables were applied to some of the simulations and included the following: Students with Disabilities (SWD) status, English Language Learner (ELL) status, gifted status, attendance, mobility, difference from the grade level modal age, class size, homogeneity of entering test scores in the class, and school effect\textsuperscript{1257}

\textsuperscript{1256} \textit{Recommendations of the Florida Student Growth Implementation Committee, 17}, (June 1, 2011), \url{http://www.fldoe.org/core/fileparse.php/7503/urlt/0102687-value-added-model-white-paper.doc}.

\textsuperscript{1257} \textit{Recommendations of the Florida Student Growth Implementation Committee, 16}, (June 1, 2011), \url{http://www.fldoe.org/core/fileparse.php/7503/urlt/0102687-value-added-model-white-paper.doc}. 
School effect referred to the practice of attributing a portion of a teacher’s rating to the student growth of his or her school as a whole based on how the school’s growth differed from the statewide expectation.

The SGIC charged AIR with testing multiple variations of each of the discussed models. The variations of each model differed in two aspects: the use of one or two years of prior student data and the number of approved covariates applied. For the latter, three lists containing approved covariates were compiled. The first list contained no variables other than the number of subject-relevant courses in which a student was enrolled and prior achievement scores. The list second list added Students with Disabilities (SWD) status, English Language Learner (ELL) status, gifted status, and attendance. The final list included all of the aforementioned approved covariates.

AIR conducted simulations to answer four key questions posed by the SGIC:

1. What is the impact of including two years of prior achievement rather than one year?
2. What is the impact of including the common school component within the teacher value-added model?

---


3. What is the impact of including different subsets of covariates?

4. How do the more complex covariate adjustment models (Models 5, 6, and 8) compare with the differences model (Model 7)?

AIR added two “lenses” the SGIC could use to compare the models. AIR recommended SGIC should first analyze the models using empirical data, in this case, numerical evidence of correlation strength and statistical differences, bias, and error among the models. The second recommendation was that SGIC should evaluate the impact data, or real-world consequences, of each model. The SGIC discussed the results of the simulations using the guiding questions and AIR guidelines. Although not denoted on the list of approved covariates, SGIC’s discussions of the findings also included analysis of the impact of controlling for teacher attendance and the highest degree obtained by the teacher.


SGIC began its analysis by discussing a set of box-and-whisker plots\textsuperscript{1271} of teacher standard error for each model.\textsuperscript{1272} The SGIC noted these plots displayed more precise estimates of teacher effects when two years of prior test scores were used.\textsuperscript{1273} Simulation data generated by the relatively simple Differences Model (Model 7) demonstrated that it estimated larger teacher effects and generated larger standard errors.\textsuperscript{1274} This finding likely contributed to the decision to use one of the more complex and less transparent models.\textsuperscript{1275}

The SGIC’s final recommendation was to adopt Model 6,\textsuperscript{1276} with the application of two prior years of test scores, the number of subject-relevant courses in which a student was enrolled, and the full set of approved covariates: i.e., Students with Disabilities (SWD) status, English Language Learner (ELL) status, gifted status, student attendance, mobility, difference from modal age in grade (as an indicator of retention), class size, and the homogeneity of entering test scores in the class. The recommended Florida Value-added Measure (VAM) could be expressed

\begin{itemize}
\item Box-and-whisker plots, invented by statistician John Tukey, are designed to display both measures of central tendency and extreme outliers in a single graphic representation. The median forms the center line of a box which extends to the 25\textsuperscript{th} quartile on the right and the 75\textsuperscript{th} quartile on the left. Whiskers extend from the left and right of the box to the minimum and maximum values, showing the full range of data.

\end{itemize}

\textsuperscript{1271} Box-and-whisker plots, invented by statistician John Tukey, are designed to display both measures of central tendency and extreme outliers in a single graphic representation. The median forms the center line of a box which extends to the 25\textsuperscript{th} quartile on the right and the 75\textsuperscript{th} quartile on the left. Whiskers extend from the left and right of the box to the minimum and maximum values, showing the full range of data.


\textsuperscript{1272} \textit{Recommendations of the Florida Student Growth Implementation Committee}, 22, (June 1, 2011), \url{http://www.fldoe.org/core/fileparse.php/7503/urlt/0102687-value-added-model-white-paper.doc}.

\textsuperscript{1273} \textit{Recommendations of the Florida Student Growth Implementation Committee}, 21, (June 1, 2011), \url{http://www.fldoe.org/core/fileparse.php/7503/urlt/0102687-value-added-model-white-paper.doc}.

\textsuperscript{1274} \textit{Recommendations of the Florida Student Growth Implementation Committee}, 21, (June 1, 2011), \url{http://www.fldoe.org/core/fileparse.php/7503/urlt/0102687-value-added-model-white-paper.doc}.

\textsuperscript{1275} \textit{Recommendations of the Florida Student Growth Implementation Committee}, 21, (June 1, 2011), \url{http://www.fldoe.org/core/fileparse.php/7503/urlt/0102687-value-added-model-white-paper.doc}.

\textsuperscript{1276} \textit{Recommendations of the Florida Student Growth Implementation Committee}, 25, (June 1, 2011), \url{http://www.fldoe.org/core/fileparse.php/7503/urlt/0102687-value-added-model-white-paper.doc}.
by the equation below.\textsuperscript{1277} This equation illustrates the SGIC’s belief that a simple formula such as the Differences Model was not the best method for applying student growth to teacher evaluation:

\[
y_i = \mu + \sum_{g=1}^{M} \delta_g x_g + \sum_{j=1}^{K} \beta_j x_j + \theta_{(x)} i + \omega_{(m)} i + \varepsilon_i.
\]

In addition to the formula used to generate growth ratings, the SGIC also worked with AIR to determine the proportion of a teacher’s final rating derived from the common school (school effect) portion. This generated “significant discussion,”\textsuperscript{1278} with proponents of including school effect data citing the preservation of collaboration among teachers and opponents arguing such inclusion could create a disincentive for teachers to work in struggling schools. After a second follow-up meeting,\textsuperscript{1279} the SGIC decided to allocate 50 percent of the common school component to teacher value-added scores to be calculated using the formula below:\textsuperscript{1280}

\textit{Teacher Value Added Score} = \textit{Unique Teacher Component} + 0.50 \times \textit{Common School Component}

To meet the deadline established by SB 736, Florida Commissioner of Education Eric J. Smith granted tentative approval of the recommended model on June 1, 2011. However, as a condition of his approval, Commissioner Smith requested further clarification of the recommendation’s common school component. After receiving this additional information,


Smith granted full approval of the recommended Value-Added Model (VAM) on June 8, 2011.  

Florida Local School District Evaluation Policies Under the New Statute

The late adoption of the VAM left Florida’s sixty-seven county-wide school districts with only weeks to design, submit, and obtain state approval of their evaluation systems for implementation during the 2011-12 school year. The newly revised statute allowed school districts to submit a plan including the option of developing “an equally appropriate formula” similar to the state value-added model for application to teachers whose grade levels and courses were not assessed by the FCAT. School districts also had the freedom to use district-created assessments. Another option included setting measurable learning targets for teachers. However due to the short timeline and absence of state resources (e.g., funding required for AIR contracts), none of these alternatives proved to be feasible for school districts such as the Alachua County Public Schools (ACPS). Also a significant time between the finalization of the new statute and the deadline for school districts to obtain state approval of their evaluation systems elapsed during the summer, a time period when many school staff were on break and, therefore, unavailable.

---


The initial ACPS Student Growth Measures system was approved by the state and posted on the district website by September 30, 2011. The approved system followed state guidelines, including the FCAT-based Florida Value-Added Model, and applied these guidelines to all teachers, including those who did not teach grade levels or courses aligned to the FCAT. For teachers of non-FCAT subjects or grade levels, confidence intervals—statistics used to reflect the degree of certainty that a data set reflects a true average—were applied to teacher value-added estimates. The ACPS acknowledged the FLDOE Value-added model was “associated with some level of error.” As a result, confidence levels ranging from 19.7 percent to 68.3 percent were applied, thus indicating a low level of certainty the scores reflected reality. This, in turn, would allow a lower growth score to be considered proficient because, when applied, a confidence interval essentially lowers the cut score for “proficient.”

For teachers of grade levels not assessed with FCAT and third grade (due to the lack of prior

---


scores needed for growth), the ACPS also applied confidence intervals and aligned the entire 40 percent of the data component of evaluations to the common school component of the value-added rating.1295

To calculate the final growth ratings of ACPS teachers during the 2011-12 school year, nine data points were calculated for each teacher. Teachers were then awarded a possible total of ten points, one point for each data element above zero. A score above zero for four or more elements was considered “effective” for the teacher’s growth rating, and a score above zero on eight or more elements was required for a “highly effective” rating.1296 The growth rating, in turn, comprised the state-mandated 50 percent of his or her overall rating for teachers with three or more years of student data and 40 percent for those with fewer than three years.1297 Escambia and Hernando County School Districts adopted similar policies.1298

In December 2012, Dr. W. Daniel Boyd, Jr., the Superintendent of Schools for the ACPS, sent a tersely-worded letter to Ms. Kathy Hebda, the Deputy Chancellor of Educator Quality at the FLDOE. The letter referred to changes made to the school district’s evaluation plan,1299 most likely in response to criticism of the steep drop in the percentage of teachers rated as “effective”


or “highly effective.” The proposed alterations awarded a rating of “effective” to teachers with three or more elements above zero and “highly effective” to teachers with seven or more elements above zero. In the letter, Dr. Boyd asked for approval in two days, by December 12, so the school district could update their data files to provide baseline scores to teachers before their winter break. Those baselines would be used to generate teacher ratings for a school year that had already elapsed to its halfway point.

Florida Litigation in Response to the Statute and Policies

Before the 2012-13 school year ended, a group of seven teachers, supported by their respective unions, filed a lawsuit in the District Court of the United States for the Northern District of Florida on April 6, 2013. The seven plaintiffs included elementary, middle school, and high school teachers from three school districts: two from the Alachua County Public Schools, four from the Escambia County School District, and one from the Hernando County Schools. Kim Cook, the lead plaintiff, had taught for twenty-two years, most recently spending two years as a first grade teacher at W.W. Irby Elementary School in Alachua County. Cook had

---


spent much of her career teaching reading and writing to English Language Learners and was selected as the 2012-13 Teacher of the Year at Irby Elementary.\textsuperscript{1305}

Janine Plavac also taught in Alachua County, serving as a health sciences teacher at Gainesville High School. In addition to her teaching duties, she had served for eight years as the Director of the Academy of Health Professions, a selective-enrollment program at the school for students intending to enter medical professions.\textsuperscript{1306} Fellow health teacher Bethann Brooks had taught at Central High School for seven years and was selected as the 2012-13 Hernando County Teacher of the Year after winning Teacher of the Year for Central High School. Brooks was chair of her department at Central High school in addition to serving on the school’s leadership team and as a mentor for new teachers.\textsuperscript{1307}

Catherine Boehme also worked at the high school level, having taught biology at West Florida High School of Advanced Technology in Escambia County since 2002 as part of her thirty-year teaching career. She was a National Board Certified Teacher. In addition to her teaching duties, Boehme served on the FLDOE Teacher and Leader Preparation Implementation Committee and on the Teacher Certification Examination Committee.\textsuperscript{1308} Emily Jefferis was employed by Escambia County as well, having taught art for nine years at Ransom Middle


School in addition to having seven prior years of teaching experience. Jefferis was regularly asked by colleges to host student teachers, and her students had won a variety of awards.\textsuperscript{1309}

Fellow Ransom Middle School teacher Cathy McConnell was a music teacher who also served at Tate High School. Orchestra participation at the high school had grown during each of her nine years leading the program. During her fifteen-year career, she also participated in contests with her students, some of whom earned superior ratings in 2013 at the Florida Orchestra Association District 3 Music Performance Assessment. The fourth Escambia County teacher, Shauna Paedae, taught math in the International Baccalaureate (IB) program at Pensacola High School. During her six years at the school, her students had enjoyed high rates of success on IB examinations, including a 90 percent pass rate in 2012. Paedae had also served for ten years as a trainer for other teachers in Pensacola, Panama City, and Tallahassee.\textsuperscript{1310} Each of these teachers was joined by his/her respective unions as plaintiffs in a suit against Florida Commissioner of Education Tony Bennett and the school board of each school district.\textsuperscript{1311}

All of the teacher plaintiffs taught grade levels or courses not assessed by the FCAT and were therefore assigned to “instructional teams”\textsuperscript{1312} as described in Florida Statute §1012.34. This created a variety of evaluation scenarios having a unique impact on these teachers, whose


\textsuperscript{1311} Cook v. Stewart, (N.D.Fla. April 16, 2013) at 6, \url{http://www.meyerbrookslaw.com/documents/Cook_percent20vs_percent20Bennett/CookvBennett_Complaint_Filed.pdf} (last visited June 24, 2015).

\textsuperscript{1312} Fla. Stat. §1012.34(7)(e)(2011).
student growth ratings were based solely on school-wide data “without any attempt to determine the individual effects of the various teachers of different subjects who instruct”\textsuperscript{1313} students.

In some cases, teachers of grade levels and subjects not aligned to the FCAT were subject to the more extreme vagaries of the statute and the responses of each teacher’s respective school district. For example, teachers in kindergarten through third grade, along with teachers who taught grades eleven or twelve, such as first grade teacher Kim Cook\textsuperscript{1314} and junior/senior math teacher Shauna Paedae,\textsuperscript{1315} were evaluated during the 2012-13 school year based on students they did not teach that year.\textsuperscript{1316} In the case of Cook, the students on whose scores her growth rating was based did not attend her school, Irby Elementary, during the 2012-13 school year. Irby Elementary served only students in preschool through second grade, who then matriculated to Alachua County Elementary School. It was at the latter school—three years after they left her classroom—that students earned the scores on which Cook’s growth rating was established.\textsuperscript{1317}

The plaintiffs argued that despite the lack of direct correlation to individual teachers, these student growth ratings had the potential for significant employment consequences for each teacher. Forty to fifty percent of the plaintiff teachers’ evaluation ratings were derived from growth on test scores on which each teacher’s direct effect was not readily apparent.\textsuperscript{1318} Due to


\textsuperscript{1317} Cook v. Stewart, (N.D.Fla. April 16, 2013) at 18, \url{http://www.meyerbrookslaw.com/documents/Cook percent20vs percent20Bennett/CookvBennett_Complaint_Filed.pdf} (last visited June 24, 2015).

\textsuperscript{1318} Cook v. Stewart, (N.D.Fla. April 16, 2013) at 19, \url{http://www.meyerbrookslaw.com/documents/Cook percent20vs percent20Bennett/CookvBennett_Complaint_Filed.pdf} (last visited June 24, 2015).
the high weighting percentage assigned to student growth, this factor could have significant impact on a teacher’s overall rating.

Under §1012, as newly revised in response to the Student Success Act, not only was student growth incorporated into the teacher evaluation process, but traditional employment protections were weakened. Professional services contracts granted after three years\(^\text{1319}\) (known in some other states as tenure)\(^\text{1320}\) were one example. Just cause was still required to terminate a veteran teacher with a professional services contract, but under the revised statute, school officials were not only allowed, but were required, to terminate teachers if they received two consecutive “unsatisfactory” ratings or three “needs improvement” ratings within a two or three-year period. A teacher receiving an “effective” rating could also be impacted. Under the revised statute, a school district implementing a reduction in force process was required under the revised statute to use performance ratings rather than seniority as the criteria for the teacher retention.\(^\text{1321}\) In addition to changes in employment security, teacher compensation was also subject to the revised statute’s performance pay requirements.\(^\text{1322}\)

The employment and compensation consequences resulting from the student growth component of the new evaluation policies led the teachers to file the aforementioned lawsuit.\(^\text{1323}\)

Although the teacher evaluation systems at issue were state and even school district-specific, the


\(^{1322}\) Fla. Stat. §1012.22(1)(c)4-5(2011).

\(^{1323}\) *Cook v. Stewart*, 28 F.Supp.3d 1207 at 1208.
plaintiffs claimed those district systems and the state statute on which they were based violated their rights to substantive due process and equal protection\footnote{Cook v. Stewart, 28 F.Supp.3d 1207 at 1209.} under the Fourteenth Amendment to the Constitution of the United States.\footnote{U.S. Const. amend. XIV, § 1.} More specifically, the legal issue was the potential for the new evaluations systems to cause deprivation of employment or compensation as a property right under Section 1 of the Fourteenth Amendment, which states that “no State shall deprive any person of life, liberty, or property without due process of the law.”\footnote{U.S. Const. amend. XIV, § 1.} Due to the fact they were invoking federal rights under the U.S. Constitution, the plaintiffs argued federal courts had subject-matter and supplemental jurisdiction and were therefore the appropriate venue.\footnote{Cook v. Stewart, No.1:13-cv-000072-MW-GRJ, (N.D.Fla. May 17, 2013) at 2.}

Changes to Florida State Statutes in Response to the Litigation

Before Federal District Court Judge Mark E. Walker issued a ruling in \textit{Cook v. Stewart},\footnote{The Florida Senate: CS/CS/HB 7009, Florida State Senate, \url{http://www.flsenate.gov/Session/Bill/2013/7009} (last visited June 26, 2015).} the state legislature passed SB 1664 and its companion, HB 7009. Both bills were signed into law by Governor Rick Scott on June 28, 2013.\footnote{Fla. Stat. §1012.34(3)(2013).} HB 7009 revised Chapter 1012 of the Florida Statutes to assure teachers would be evaluated based on the performance of “students assigned to their classrooms.”\footnote{Fla. Stat. §1012.34(3)(2013).} This legislative change appeared to provide a remedy to one of the
concerns raised by *Cook v. Stewart*: i.e., teachers being evaluated based on the growth of students they did not teach.


Florida Education Association Response to Revised Legislation

The Florida Education Association described the changes made by SB 1664 as a “partial fix” and indicated it would continue to move forward with the *Cook v. Stewart* lawsuit. In a news release, union officials stated that “too many unanswered questions” remained. While the legislative change had addressed the issue of teachers being evaluated based on the performance of students they did not teach, the amended statute did not mandate teachers to be evaluated based on student growth in the subjects they taught because it did not require the creation of additional assessments for subjects other than reading or math or grade levels other than 3–10.

---


Court Grants Defendants’ Motion to Dismiss in Part

On July 8, 2013, the state defendants and Hernando County officials filed a motion to dismiss\textsuperscript{1334} the Cook v. Stewart lawsuit. Escambia County did the same on July 10,\textsuperscript{1335} followed by Alachua County on July 12.\textsuperscript{1336} On April 22, 2014, Federal District Court Judge Mark E. Walker granted in part and denied in part the state defendants’ motion to dismiss the complaint.

The court first addressed the issue of “standing”\textsuperscript{1337} pertaining to whether the plaintiffs had sufficient stake in the outcome of the controversy to warrant the intervention of the federal court.\textsuperscript{1338} In its 1992 ruling on Lujan v. Defenders of Wildlife, the United States Supreme Court established a three-pronged test for assessing a party’s standing. The Florida court applied this test to Cook v. Stewart.\textsuperscript{1339} First, the plaintiff must establish an “injury in fact”\textsuperscript{1340} to a legally protected interest that is concrete, particularized, and imminent, not hypothetical. Second, there must be a causal connection between the alleged injury and the defendant’s challenged action.


Third, it must be likely the injury would be redressed if the court ruled in favor of the plaintiff.\textsuperscript{1341}

The court concluded the plaintiffs’ complaint met all three prongs. The threat of possible lost compensation under the new evaluation system was sufficiently concrete and particularized to satisfy the first prong. Although it was difficult to establish direct causation, the suit also met the second prong. Judge Walker pointed out the standard was lower than that of “proximate cause”\textsuperscript{1342} and that a less direct effect could be “fairly traceable”\textsuperscript{1343} to the revised statute. The court also found the suit satisfied the third prong. If there were an injunction against the use of the new evaluation systems, it would prevent future harm.\textsuperscript{1344}

Next the court determined standard of constitutional review to be applied to the \textit{Student Success Act}.\textsuperscript{1345} For an equal protection complaint such as \textit{Cook v. Stewart}, the level of scrutiny applied needed to be determined by the court. The Supreme Court had established two primary tests for analysis in Equal Protection cases. The strict scrutiny test is used sparingly, most often for cases involving a suspect class such as race or national origin. Generally, all other claims are subject to the less rigorous rational basis test, requiring the government entity only to proffer a rational reason for the legislation, policy, practice, or procedure that furthered a legitimate


government interest. The plaintiffs had cited *St. Ann v. Palisi*, a New Orleans case involving students who were suspended from a school due to the behavior of their mother as the basis for arguing for the application of a strict scrutiny analysis, a process that would subject the Act to the most rigorous judicial examination. The *Cook v. Stewart* court stated the types of punishment for the conduct of others triggering a strict scrutiny review generally either involved criminal conduct or “a special case, such as the burdens placed on children for their parents’ conduct in *St. Ann*. ” The court concluded an employment consequence such as a teacher being evaluated based on on student performance or subjects he or she did not teach did not qualify for strict scrutiny. Therefore, the court held, the Student Success Act would be reviewed using the less stringent rational basis test that presumed a law is Constitutional so long as it shown to be rationally related to a legitimate government purpose.

---


Judge Walker then analyzed the plaintiffs’ facial challenge.\textsuperscript{1353} If the facial challenge passed, no application of the statute would have been considered Constitutional,\textsuperscript{1354} allowing the court to find the statute unconstitutional in its entirety.\textsuperscript{1355} The court addressed the key legal question relative to the plaintiffs’ facial challenge. Could the state legislature rationally believe it was furthering its legitimate interest in student learning by evaluating a teacher based on the performance of students he or she does not teach or his or her students’ test scores from a different subject? The court concluded the answer was “yes,” and the facial challenge failed.\textsuperscript{1356} Judge Walker wrote “it is entirely rational to believe that tying various facets of the terms of employment—salary, eligibility for promotions, etc.—to those evaluations might incentivize better teaching, thus advancing the state’s interest in increasing student learning growth.”\textsuperscript{1357} Judge Walker couched his ruling in caveats about the plaintiffs having made a “plausible showing that the Act is flawed and unfair,”\textsuperscript{1358} including “serious flaws identified by the

\begin{flushleft}
\textsuperscript{1353} Cook v. Stewart Order Granting in Part and Denying in Part State Defendants’ Motion to Dismiss, No.1:13-cv-000072-MW-GRJ, (N.D.Fla. April 22, 2014) at 12.
\end{flushleft}

\begin{flushleft}
\textsuperscript{1354} Alex Kreit, Making Sense of Facial and As-Applied Challenges, 18 William and Mary Bill of Rights Journal 658 (2010), http://scholarship.law.wm.edu/cgi/viewcontent.cgi?article=1168&context=wmborj.
\end{flushleft}

\begin{flushleft}
\textsuperscript{1355} Alex Kreit, Making Sense of Facial and As-Applied Challenges, 18 William and Mary Bill of Rights Journal 658 (2010), http://scholarship.law.wm.edu/cgi/viewcontent.cgi?article=1168&context=wmborj.
\end{flushleft}

\begin{flushleft}
\textsuperscript{1356} Cook v. Stewart Order Granting in Part and Denying in Part State Defendants’ Motion to Dismiss, No.1:13-cv-000072-MW-GRJ, (N.D.Fla. April 22, 2014) at 17.
\end{flushleft}

\begin{flushleft}
\end{flushleft}

\begin{flushleft}
\end{flushleft}
Ultimately, however, he stated, “This court cannot, and will not, allow its own critical views of a flawed initiative to color its analysis.”

The court then addressed the plaintiffs’ as-applied challenge, a type of challenge where a plaintiff argues that while an act is constitutional, it operates unconstitutionally when applied to his or her circumstances. In this particular case, the specific plaintiffs’ circumstances were dictated by district policies implemented in response to the statute. Ergo, Judge Walker ruled, “the proper object of judicial scrutiny is the policies themselves, not the Act” and the court granted the motion to dismiss the plaintiffs’ challenge to the Florida Student Success Act.

Granting the defendants’ motion to dismiss was denoted as partial because the court denied the motion to dismiss the plaintiffs’ challenges to the school districts’ evaluation policies. Judge Walker noted the defendants’ motion to dismiss failed to “discuss how or why the policies [were] rational except in the most general and conclusory terms” and attempted to “treat the Act and policies as one creature for purposes of the rational basis analysis,


ignoring the fact that they [were] separate legislative actions.”\textsuperscript{1366} Also, while Judge Walker dismissed the plaintiffs’ challenge to the state statute, he did not dismiss the state as defendants in the lawsuit, citing as reason the fact the state “played a vital role in the development and approval”\textsuperscript{1367} of each school district’s policies.\textsuperscript{1368}

Court Order on Cross-Motions for Summary Judgment

On May 6, 2014, Judge Walker issued an order on the cross-motions for summary judgment to resolve the plaintiffs’ substantive due process and equal protection challenges to the school districts’ evaluation policies.\textsuperscript{1369} Although the defendants did not respond to the plaintiffs’ motion for summary judgment, the court held it could not rule for the plaintiffs “on the mere fact that the motion was unopposed.”\textsuperscript{1370} Judge Walker noted during the April 9, 2014 hearing, representatives for the defendants indicated they had made a deliberate decision not to continue an aggressive defense to avoid legal expenses and preserve their limited resources. The court sympathized with the plight of the school districts, noting they were “attempting to implement an unfunded mandate.”\textsuperscript{1371}


\textsuperscript{1369} \textit{Cook v. Stewart}, 28 F.Supp.3d 1207 at 1209.

\textsuperscript{1370} \textit{Cook v. Stewart}, 28 F.Supp.3d 1207 at 1209.

\textsuperscript{1371} \textit{Cook v. Stewart}, 28 F.Supp.3d 1207 at 1209.
Judge Walker first addressed the plaintiffs’ facial challenge to the policies, beginning with the substantive due process challenge. Although the policies were not directly enacted by the state legislature, the court held they were subject to the rational basis test to the same extent as state statutes, citing Snowden v. Hughes, a federal case from the District Court for Northern Illinois pertaining to the actions of a State Primary Canvassing Board during an election. The court reasoned the school districts’ evaluation policies passed rational basis review because the school officials who wrote the policies and the Florida Department of Education that approved them could rationally believe the evaluation systems served a legitimate government purpose of increasing student learning.

When filing their equal protection claim, the plaintiffs posited the school districts’ policies violated their equal protection rights under the Fourteenth Amendment because they created “separate classes of teachers in Florida: those whose evaluations [were] based on student growth data for students assigned to the teacher in the subjects taught by the teacher, and those whose evaluations [were] based on student growth data for students and/or subjects they do not teach.” Due to the fact the complaint did not involve suspect classes such as race or national origin, the court applied the rational basis test rather than the strict scrutiny analysis. Again the court’s holding supporting the school districts’ policies paralleled its finding for the state statute. Judge Walker found the classification of teachers passed rational basis review because

---


“no invidious discrimination” was taking place, and the classification advanced a legitimate governmental interest in increasing student learning. In addressing the as-applied claims of the plaintiffs, the court held there was “nothing particular to Plaintiffs’ as-applied claims that would require an analysis different from the analysis of the facial claims.” The court issued a summary judgment in favor of the defendants, and *Cook v. Stewart* was dismissed in its entirety.

As he had when granting, in part, the defendants’ motion to dismiss, Judge Walker couched his findings with a caveat that the dismissal of the remainder of the case did not constitute his approbation of the statute. “Needless to say,” he wrote, “this Court would be hard-pressed to find anyone who would find this evaluation system fair to non-FCAT teachers, let alone be willing to submit to a similar evaluation system.” Judge Walker suggested that while a federal lawsuit challenging the constitutionality of the statute and resultant policies was not the appropriate recourse, there was a venue for change. “The Constitution presumes that, absent some reason to infer antipathy, even improvident decisions will eventually be rectified by the democratic process.” On September 5, 2014, the plaintiffs filed an appeal. The federal appeals court upheld the ruling of the lower court in a holding published on July 7, 2015.

---

1379 *Cook v. Stewart*, 28 F.Supp.3d 1207 at 1215.
1382 *Cook v. Stewart*, 28 F.Supp.3d 1207 at 1215.
Further Changes to Legislation

In between the appeal and the appellate panel’s ruling upholding dismissal of the lawsuit, HB 7069 was passed by both houses and signed into law by Governor Rick Scott on April 14, 2015. The new law required the administration of “statewide, standardized”\textsuperscript{1385} End-of-Course (EOC) assessments for high school courses such as geometry, algebra II,\textsuperscript{1386} biology,\textsuperscript{1387} and United States history.\textsuperscript{1388} The revised statute also granted permission to the commissioner to select assessments such as advanced placement (AP) and international baccalaureate (IB) exams to serve as state-mandated end-of-course tests.\textsuperscript{1389} These assessments had been added to the law annually since 2011,\textsuperscript{1390} prior to which, the only available state assessments were FCAT reading and math. For teachers of those high school courses, these additional state assessments could have the potential to serve as student achievement data for use in district teacher evaluation policies, mitigating at least some instances of teachers being evaluated on subjects or students they do not teach.

\textsuperscript{1384} The Florida Senate, HB 7069: Education Accountability, Florida State Senate, \url{https://www.flsenate.gov/Session/Bill/2015/7069} (last visited June 23, 2015).
A change entirely new to the 2015 statute required that thirty—rather than fifty—percent of a teacher’s performance evaluation be based on student growth. This change had the potential to reduce the potential employment consequences for teachers. A revision mentioned as one of the most significant in change summaries was a new requirement to “provide instructional personnel with information on student achievement of standards and benchmarks to improve instruction,” thereby implying this may not have been a consistent practice in the past.

The revisions also included the addition of using student achievement data as an alternative to student growth data. While the use of achievement rather than growth data has its own set of limitations, it would allow teachers lacking three years of trend data, including third grade teachers, to be evaluated based on the performance of their own students. For non-classroom teachers, the revised statute no longer included the requirement that teachers be evaluated using student growth on statewide assessments, opening the door for school districts to develop and use other assessments more germane to classroom instruction. In 2014, the language requiring school districts to award teacher performance pay based on evaluation ratings had been downgraded to a financial incentive for school districts. The revised 2015 statute removed the financial incentives.

---

dismissed, some of the issues raised by the plaintiffs were addressed, as predicted by Judge Walker, through the democratic process, as evidenced by HB 7069.

---

1397 Cook v. Stewart, 28 F.Supp.3d 1207 at 1215.
CHAPTER 3

LEARNING FROM HISTORY AND LESSONS FROM LAWSUITS

The 21st century is in many ways vastly different from Horace Mann’s mid-nineteenth century. When Mann introduced the first standardized test in the United States, the economy was just emerging from a primarily agrarian subsistence economy to a market economy. In the more than 150 years that elapsed between the Common School Movement and Race to the Top, the “waste sand dunes” ¹ in Northern Indiana and the open land in many other areas across the United States was converted by the Industrial Revolution into manufacturing powerhouses replete with steel mills and factories, only to decline decades later into a deserted “rust belt.” ² Trains, which were new and limited in scope during Mann’s day, eventually expanded to traverse the nation and create a new class of wealthy oligarchs, only to be subsequently replaced by automobiles and airplanes. As vastly different as historical eras may be, numerous similarities exist among the eras of education reform that punctuate them. Analyzing the common societal factors correlated to increased accountability for public education can be instructive for school districts.

New York and Florida are also dissimilar in many ways, including their history, climate, and the size of school districts. The litigation filed in each state in response to the use of

¹ John Franklin Bobbitt, The Elimination of Waste in Education, 12 The Elementary School Teacher 259 (1912).
standardized test scores as part of the teacher evaluation process differed as well. New York’s litigation was filed in the state courts by a union before any teachers were evaluated under the new system. In contrast, the Florida lawsuit was a consolidation of filings by several individual teachers who had already been evaluated under Florida’s RTTT rules. The case was heard in federal court and dismissed in its entirety. An analysis of the cases, however, reveals several commonalities that could be instructive for school officials wishing to avoid a visit to the courtroom.

Accountability is Not Poised to Pull a Vanishing Act

In the 2000s, many educators posited that federal accountability would “go away” when George W. Bush left office and the No Child Left Behind Act came due for reauthorization. During his 2008 election campaign, presidential candidate Barack Obama did little to disabuse voters of this notion. As noted in Chapter 2, during a speech, candidate Obama decried NCLB for its emphasis on high-stakes testing. Once he was elected, however, Obama’s Race to the Top program increased the length, rigor, and federal control over standardized tests and also raised the already-high stakes by requiring that achievement scores be utilized as part of teacher

---


evaluation calculus.\textsuperscript{4} Not only did President Obama’s actions in office belie the notion that federal accountability for schools was a passing Bush-era fad, but history also reveals patterns that call into question the alleged temporal nature of federal education reform. This is fitting, given the durability of both standardized testing and accountability in the United States.

Inexorable Geographic Expansion of Reform and the “Testing Machine”

As established in Chapter 2, the mass propagation of standardized assessments and accountability data did not, contrary to what many believe, spring fully formed from the womb of No Child Left Behind in 2000. Rather its genesis harks back almost as far as American public education itself. As described earlier, standardized testing in the United States began in 1837 when Mann first required a common end-of-year exam in the Boston public schools.\textsuperscript{5} While Mann’s work pertained only to one Massachusetts city, the pattern of standardized testing and accountability over the ensuing years has been one of steady expansion.

During the Efficiency Era outlined in Chapter 2, other major cities, including Cleveland, not only administered assessments, but also conducted in-depth qualitative and quantitative studies of their school systems.\textsuperscript{6} Disproving the assertion that schools only fairly recently came to devour data in its many forms, these early studies came complete with data visualizations.\textsuperscript{7}


\textsuperscript{5} William J. Reese, Testing Wars in the Public Schools: A Forgotten History 73 (Harvard University Press 2013).


Although they were hand-drawn and reproduced only on hard copy in black ink, they rival the complexity of those currently found on cloud-based portals.\(^8\)

As previously described, all of this testing, tabulating, and analyzing required resources. Soon all but one of the United States’ ten largest cities had research bureaus to oversee assessments and data,\(^9\) belying any claim that education has only recently become a data-driven profession. As for other uses of test scores, *Cleveland Study* researcher Leonard Porter Ayers likely would not have been shocked at Race to the Top’s required use of student academic test scores as part of the teacher evaluation process. On the contrary, were he alive today, Ayers might well have been surprised that it took so many years for it to come to fruition. After all, as noted in Chapter 2, he made prescient mention of merit-based promotions\(^10\) and the need for schools to develop systems to discern between average and excellent teachers.\(^11\)

Some might correctly retort that all of the aforementioned assessments, accountability, and reforms were primarily based at the municipal level. When cities participated in national-level research, such as the *Eight-Year Study*, it was at least quasi-voluntary. For those who long for what they feel is a fairly recent era of local control of education, the next phase of expansion—and the date of its genesis—is worth noting. From individual cities, testing, data, and accountability soon expanded to the state level. Just over a decade after the *Cleveland Study*,


Pennsylvania began a state-wide study.\textsuperscript{12} The next year Iowa launched its first statewide testing program as part of the “Brain Derby.”\textsuperscript{13} By the mid-point of the following decade, New York followed suit with a similar study.\textsuperscript{14}

As previously discussed, a mere two years later and more than fifty years before the aforementioned Charlottesville Summit, the voluntary growth of state-level testing and data analysis became exponential, when fifteen states participated in a 1937 conference facilitated by the Committee of Measurement and Guidance of the American Council on Education.\textsuperscript{15} The Council eventually identified states as the proper entity to oversee testing and the monitoring of education.\textsuperscript{16} Ironically, by 1940, the resulting Eight-Year Study mentioned in Chapter 2 had elevated data and education reform to the quasi-national level, as it analyzed data from individual districts and schools in multiple states.\textsuperscript{17} It is likely not a coincidence the geographic expansion of accountability and testing coincided with an era when states began shouldering a larger proportion of the funding for local school districts.\textsuperscript{18} By lining up at the state trough for a fiscal meal, school districts invited accountability from afar and unwittingly surrendered the first


\textsuperscript{13}E.F. Lindquist, \textit{The Iowa Testing Programs – A Retrospective View}, 91 Education 8 (1970).


\textsuperscript{17}Wilford M. Aikin, \textit{Adventure in American Education Volume 1: The Story of the Eight-Year Study} preface (Harper and Brothers 1942).

pieces of local control—and the geographic expansion of standardized testing and accountability continued.

Eighteen years later, after both a world war and a cold war, the term “quasi” no longer applied to national-level standardized testing, data analysis, and accountability. By this time, the federal government had thrown its hat—or more fittingly, astronaut helmet—into the ring. Following the 1957 Sputnik Soviet satellite launch, President Eisenhower signed the National Defense Education Act (NDEA) into law, ushering in a federal role in both funding and measuring public education. Like state funds, federal monies came with the proverbial strings attached, albeit fairly loose ones. From NDEA forward, federal funding of—and accountability for—public education were constants, with the only shift being back and forth between excellence and equity. NDEA focused on excellence in the face of perceived Soviet competition.

As noted in Chapter 2, the Elementary and Secondary Education Act, with its roots in the Civil Rights movement, was about equity. This did not, mean, however, that accountability and testing were absent. As previously described, states applying for federal funds under the Elementary and Secondary Education Act were required to collect evidence of efficacy in the form of student standardized test scores.

In some cases, state-wide tests were either developed or selected from among commercial assessments. To avoid running afoul of education reform’s “third R” of “Reds” by appearing like centralized communism, the quasi-national NAEP was administered only voluntarily to

---


a small sample of students. However, if one imagines local public school districts as the proverbial frog sitting in a pot of water, the heat had officially been turned up a notch.

As outlined in the previous chapter, ESEA remained largely unchanged from its early iteration through the 1970s and 1980s, continuing to provide millions of dollars in federal funding to local school districts in exchange for fairly moderate requirements for assessment and accountability. The strings would soon tighten. While increased federal accountability would not be fully realized for nearly two more decades, the signs were there in the early 1980s as difficult to miss as the big hair and shoulder pads that dominated fashion at that time. President Reagan delivered a speech on the *Nation at Risk Report* (NAR). Instead of a physical launch as in the case of Eisenhower’s Sputnik, Reagan referred to “unilateral educational disarmament” in the form of declining public schools and called for sweeping reforms.

These reforms would eventually be enacted by President George W. Bush. Bush not only led the next phase of federal education reform, but as previously described, even made it a lynchpin of his first presidential campaign in 2000. NCLB appears to have ended the partisan disagreement that had either halted previous federal reforms or stalled momentum by swinging the direction sharply from left to right. As noted in Chapter 2, NCLB’s marriage of equity and

---


excellence generated strong bipartisan Congressional support and led to an unprecedented level of federal accountability, mandating reforms that had previously been merely suggestions. NCLB required states not only to administer their own assessments to all students in grades three through eight and high school, but also required an accompanying affirmation of the rigor of their instruments by testing samples of students with the aforementioned federal NAEP. The latter requirement, while it sounds minor, subtly removed more control from states and placed it at the federal level. A state could still create its own tests, but if the results did not correlate to scores from NAEP, the content or cut scores had to be revised. With this development, not only local but in many aspects even state control was but an illusion. Education reform had officially been federalized.

In light of the aforementioned historical events, RTTT and its federal requirement that scores from standardized assessments be used as a part of the teacher evaluation process should not be a surprise. In fact, it was almost inevitable. The inexorable expansion of standardized testing, the gradual reduction in emphasis on innate intelligence, the century-old mention of the need to better measure teacher efficacy, and the geographic expansion in funding and accountability were all harbingers of what was to come in 2011. Figure 1 shows the changes in assessment and its relation to teacher evaluation throughout history.

---


Figure 1: Changes in assessment standards.

Shifting Locus of Responsibility: Students to Schools to Teachers

In addition to expanding geographically from city to state to federal levels, the types of tests used as part of accountability—and their implications for public education—have also shifted over time. Although copies of Mann’s standardized exams\textsuperscript{31} are not readily available

\textsuperscript{31}William J. Reese, Testing Wars in the Public Schools: A Forgotten History 73 (Harvard University Press 2013).
today, it is safe to assume they were achievement-based. During the Common School Era, Alfred Binet was still a half century away from developing the first aptitude test.32

During the early years of the Efficiency Era that overlapped with the eugenics movement and the proliferation of intelligence testing, data from both aptitude and achievement exams were used for the era’s large-scale studies.33 The inclusion of aptitude test data, in contrast to the other collected metrics, signaled the attribution of at least part of student learning to innate intelligence. Throughout the end of the era, IQ tests remained popular. However, the results of the New York research signaled a shift in perceived responsibility for student academic achievement. As described in Chapter 2, one of the many outcomes of the 1935 Regents’ Study was to repurpose an existing standardized assessment, the Regents’ Exam, from a graduation test measuring the results of students’ abilities and efforts to an assessment that sought to identify areas for improvement in both curriculum and pedagogy.34 This effectively removed at least part of the onus for learning from individual students and placed it firmly with schools.

Little had changed by mid-century. The standardized tests associated with NDEA included both achievement and aptitude tests.35 This maintained some emphasis on the contribution of students’ innate cognitive abilities to their learning. Less than thirty years later, however, the rhetoric of reform signaled a complete shift of all responsibility for student learning

---

32 Anya Kamenetz, The Test: Why our Schools are Obsessed with Standardized Testing—but You Don’t Have to Be 44 (Public Affairs 2015).


to schools. While he referenced Sputnik and NDEA when citing declines in educational gains during his NAR speech, President Ronald Reagan was careful to state that this was not due to a decline in student intelligence, signaling that increased achievement—not aptitude—testing was likely to follow, and that responsibility lay with the schools.

Neither NCLB or RTTT included any aptitude testing. Instead their focus was solely on the data generated by achievement tests. This, in turn, reflects the trend of increasing attribution of student performance to schools. Since the work of schools is implemented by principals and teachers, it is not surprising that RTTT included a requirement that states attribute growth on achievement tests to school staff and use it as the basis for evaluation.

Media Criticism: Fact of Life since Letterpress Days

Some educational leaders point to the Nation at Risk Report or NCLB as the starting point for media criticism of public education in the United States. In reality, the Fourth Estate has been slingimg mud at education since the beginning of education reform—and in fact, they may have made it possible. As noted in Chapter 2, Mann’s Boston reforms were politically feasible, in part, because of the news media’s enthusiasm for statistics that drowned out the anti-reform articles previously published by the popular Grammar Masters. A century later, a magazine series may have sparked the federalization of reform. As previously described, in


Sputnik’s wake, the life adjustment curriculum prevalent in American high schools\textsuperscript{39} was portrayed by the media as a random assortment of feel-good electives. Also, over 50 years prior to RTTT, journalists were already using questionable research to criticize public schools. In the \textit{Life} article series that blamed Soviet dominance on American education, the entire data set supporting the claim consisted of an informal qualitative analysis\textsuperscript{40} of two students: one rather dim Chicago high school student juxtaposed with a single scholarly Soviet teenager. As a result, the series almost reached the point of being a pedantic parody.\textsuperscript{41}

This comparison was hardly a fair—let alone statistically valid—and yet it was generalized and used in \textit{Life} magazine’s urgent call for changes to American high schools. In other words, any educational historian who believes that media criticism of public education began in the 1980s with \textit{Nation at Risk} (NAR) would be well served by procuring a copy of March 24, 1958 issue of \textit{Life} magazine. NAR was heavily publicized in the media, and since that time, media attacks on education have been fairly constant. In light of this, it should not be a surprise that the public in general was not upset by the RTTT requirement to use test scores for teacher evaluation

\begin{quote}
Prussians, Russians, and 170 Years of “Inferior” U.S. Schools
\end{quote}

From time immemorial—or at least from the Common School Era forward—a favorite tactic of both the media and educational leaders themselves has been to compare public education in the United States unfavorably with schools in other countries. The only variance has

\textsuperscript{39} Thomas D. Fallace, The Effects of Life-Adjustment Education on the U.S. History Curriculum 575 (William Patterson University Press 2011).

\textsuperscript{40} Sloan Wilson, \textit{Crisis in Education}, Life, March 2, 1958, at 25.

\textsuperscript{41} Sloan Wilson, \textit{Crisis in Education}, Life, March 2, 1958, at 25.
been the superior nation of choice. As described in Chapter 2, having spent his honeymoon abroad engaged in research on education, Mann found the performance of U.S. schools to be far inferior to those in Europe.\textsuperscript{42} One wonders how Mrs. Mann compared her husband’s performance to others after spending her honeymoon pondering Prussian pedagogy. One hundred years later, the superior nation of choice shifted from Prussia to Russia in the wake of Sputnik,\textsuperscript{43} which in turn likely generated support for the reforms of NDEA.\textsuperscript{44}

While he did not shy away from reference to the “Evil Empire” of Russia on other occasions, President Reagan had another potential foe in his sights when he delivered the \textit{Nation at Risk} speech. As noted in Chapter 2, Reagan cited rising global competition,\textsuperscript{45} particularly from Japan,\textsuperscript{46} as evidence that educational reform was needed. Beginning with President George H.W. Bush and continuing through all of his successors, including the Obama administration during RTTT,\textsuperscript{47} the federal call for reform no longer bothered to select an individual country superior to our own. Instead leaders referred to the need for our students to be able to compete in a global economy.\textsuperscript{48} This essentially implies that public education in the United States lags

\begin{flushleft}
\textsuperscript{43} Sloan Wilson, \textit{Crisis in Education}, Life, March 2, 1958, at 25. \\
\textsuperscript{44} Sloan Wilson, \textit{Crisis in Education}, Life, March 2, 1958, at 25. \\
\textsuperscript{46} Jim Horn & Denise Willburn, \textit{The Mismeasure of Education} 196 (Information Age Publishing 2013). \\
\end{flushleft}
behind most, or at least many, other nations. It is small wonder then that the use of test scores for
teacher evaluations as required by RTTT has enjoyed bipartisan support.

Private-Sector Involvement in Education: Business Has Never Minded Its Own Business

Some educators decried the unprecedented involvement of business in education during
the Race to the Top era as if it were part of a passing fad. However, attention to the needs of the
workplace in curriculum and assessment design, application of business principles to schools,
and interference of private sector leaders in education reform have a long history in the United
States. Like the encroachment of standardized testing, private sector influence in education
began indirectly and quasi-voluntarily.

The private-public enmeshment relied on principles before people. Before business
leaders began bloviating about the ills of public education, educators served as their proxies. As
referenced in Chapter 2, John Franklin Bobbitt borrowed heavily from engineer Frederick
Taylor’s work on business efficiency\(^49\) when he wrote “The Elimination of Waste in
Education,”\(^50\) the article that ushered in the scientific management—or efficiency—era of
reform.

The mere application of private-sector ideas to public schools with educational
researchers as conduits did not suffice for long. Principles were soon followed by people. At
first, business figures exerted their influence indirectly through funding. As mentioned
previously, another Efficiency Era reformer, Ayers, rose to prominence in education reform
when he authored a New York education study. Funding for the study was provided by railroad


\(^{50}\) John Franklin Bobbitt, *The Elimination of Waste in Education*, 12 The Elementary School Teacher 259 (1912).
magnate Russell Sage’s foundation. Ayers’ next such large-scale study, the Cleveland School Survey, was funded by The Cleveland Foundation, a consortium of philanthropists established by attorney, banker, and “mover and shaker” Frederick H. Goff. A few years later and one state east of Ohio, the next large-scale education reform research commenced. The Eight-Year Study, led by Ralph W. Tyler, evaluated high schools across the state of Pennsylvania. Tyler’s work was sponsored by the Carnegie Corporation, a philanthropic organization established by steel magnate Andrew Carnegie. Much the same as state and federal government, business leaders were also willing to pay for their influence, and local districts, particularly in large urban areas, happily sold them their freedom.

Soon private-sector leaders were advancing their agenda directly by personally participating in discourse surrounding education reform. As noted in Chapter 2, this occurred in the Cleveland area three years after its eponymous study, when E.S. Carman, one of the Frederick Taylor’s successors as President of the American Society of Mechanical Engineers,

participated in a panel alongside Cleveland school officials. Even the era of reform centered on civil rights and equity was not free of the influence of corporate leaders. John Gardner, president of the Carnegie Corporation, was appointed to the post of Secretary of Health, Education, and Welfare and played an instrumental role in the design of the Elementary and Secondary Education Act (ESEA). The appointment of a Carnegie Foundation leader was just the beginning of corporate influence over ESEA. As described in Chapter 2, additional accountability in the form of NAEP was later added to ESEA in response to government adoption of a private-sector process known as “program budgeting.”

In the 1980s, when the seeds for the next wave of education reform were planted under President Ronald Reagan, business leaders were once again at the table. The committee that released the *Nation at Risk Report* included business leaders. The influence of these private-sector leaders was readily apparent when NAR raised the specter of looming economic

59 11 The Journal of the Cleveland Engineering Society 267 (1918). In the interest of full disclosure, E.S. Carman was the great-grandfather of the author.


63 Neil J. Salkind, *Directions in Educational Psychology* 286 (Sage Publications 2011)


dominance by Japan. While much control over the details was left to state and local entities, that did not result in a lack of corporate influence over the implementation of NAR reforms. Some reform efforts, such as the Boston Compact, were spearheaded in part by private sector leaders from organizations like the Business Roundtable.

While significant changes between NAR and NCLB were not seen by the educational public, business continued to flex its muscles behind the scenes. As previously discussed, President Clinton’s Goals 2000 went a step further in connecting public schools to the private sector, actually prescribing direct business partnerships for school districts. President George W. Bush’s No Child Left Behind, which followed on the heels of Goals 2000, was supported by the Business Roundtable and other private-sector organizations. In fact, they not only helped draft the legislation, but also defended it several years later when it came under attack by a variety of other coalitions.

Given the ever-increasing level of business leader involvement in public education that has spanned over one hundred years, it is hardly surprising that private-sector leaders such as Eli Broad and Bill Gates participated in creating Race to the Top (RTTT). Nor is it a surprise that RTTT applied competitive business principles, such as the inclusion of student achievement data in teacher evaluations, to the classroom. The most notable difference from reforms a century earlier was that industrialist millionaires yielded the stage to Internet billionaires.

---


Causal Factors of Education Reform: It Is the Economy

The influence of business leaders in education can be at least partially explained by the correlation between education reform initiatives and the presence of significant economic change or decline, unfavorable international comparison, or an increasing need for higher education and workplace readiness. This correlation dates back to the United States’ first public school standardized test in 1845.\(^70\)

Mann assumed his position as Secretary of the Massachusetts Board of Education\(^72\) amid significant economic change. At that time, America was urbanizing at a rapid pace due to the newly-emerging U.S. market economy.\(^73\) The ruling Whig Party believed public education was poised to serve as the solution to both the demands resulting from the changing economy and class warfare. This ideal of public schools as a cure for social ills set a pattern that has persisted for over a sesquicentury. As noted earlier, however, Mann did not believe his public schools, then in the grip of the Grammar Masters,\(^74\) were fit to serve the purposes identified by the Whigs.\(^75\) Mann instituted his reforms, the implementation being enforced in part by administration of standardized exams.\(^76\) This established another precedent that persists to this day.

---

\(^70\) William J. Reese, Testing Wars in the Public Schools: A Forgotten History 73 (Harvard University Press 2013).

\(^71\) There was no hysterical braying about the workforce being replaced by robots during the Victorian era, of course.

\(^72\) Jim Horn & Denise Wilburn, The Mismeasure of Education 8 (Information Age Publishing 2013).

\(^73\) William J. Reese, Testing Wars in the Public Schools: A Forgotten History 51 (Harvard University Press 2013).


\(^76\) William J. Reese, Testing Wars in the Public Schools: A Forgotten History 73 (Harvard University Press 2013).
The next era of reform, the Efficiency Era, was the product not only of economic change as the industrial age commenced, but also of a decline in the form of financial panics.\textsuperscript{77} Once again, public education was called on to make changes designed to solve the nation’s socioeconomic ills. Worse than mere periodic panics was the Great Depression of the 1930s.\textsuperscript{78} This economic cataclysm ushered in yet another era of reform. During the era defined in Chapter 2 as the Technocratic Meritocracy, the educational leaders of the Depression era established statistics bureaus, created or revised tests, and conducted large-scale studies.\textsuperscript{79}

The final shift in assessment during the Technocratic Meritocracy, however, was the result not of an economic downturn but rather a seismic economic change. Veterans returned from World War II armed with college funding from the G.I. Bill of 1944.\textsuperscript{80} The first impact this had on assessment was the reorganization of the College Board\textsuperscript{81} to administer the SAT to a larger number of potential students more efficiently.\textsuperscript{82} World War II also brought about the

\textsuperscript{77} David F. Labraree, Someone Has to Fail: The Zero-Sum game of Public Schooling 88 (Harvard University Press 2012).


death of the Iowa “Brain Derby” and the replacement of its test with the ITED, and thereafter the ACT.

The advent of the SAT and the ACT reflected not only the impact of the Depression and the G.I. Bill but also the general shift in the economy toward upward mobility. It was African Americans’ wish to participate in this mobility which created demand for desegregated schools and led to the Civil Rights movement, and when this proved insufficient to funding and reform via the ESEA. As described previously, ESEA continued to provide accountability and funding virtually uncontested until 1980 when President Ronald Reagan asked, “What have we bought with all that spending?” It would take yet another era of economic uncertainty for the proposed reforms of NAR come to fruition just over two decades later. In 2002, as previously described, George W. Bush, the son of Reagan’s vice president, introduced No Child Left Behind.

Less than a decade later, the United States was facing both a recession and a transformation from the information age to the 21st Century economy. It would be disingenuous for any student of educational history to claim to have been surprised when Race to the Top

---


stepped into the vacuum as the a “piñata”\textsuperscript{89} that was NCLB reauthorization swung about wildly, shedding tissue paper and stale candy in the halls of Congress. RTTT fit a pattern established over 150 years earlier—and as states and districts held out one hand for funds, they surrendered more of their control with the other.

College & Career Readiness

As previously described, Arne Duncan referred to first-time reforms focused on what many feel is the 21\textsuperscript{st} Century concept of College & Career Readiness. \textsuperscript{90} It may be this type of rhetoric that makes some believe College & Career Readiness and the requirement to use test scores as part of the teacher evaluation process are merely the latest fads poised to go the way of other contemporary whims such as fidget spinners and nitrogen coffee. To use another source of caffeine as an analogy, College & Career Readiness can be read in the tea leaves of education reform as far back as the mid-nineteenth century. As previously discussed, Mann felt schools needed to center instruction around “the actual business of men and the affairs of life,”\textsuperscript{91} a Victorian-Era phrase for “Career Readiness.”

Efficiency Era source documents yield still more evidence that the concept of schools being responsible for preparing students for careers is a durable one. As noted in Chapter 2, the impetus for the entire era was the need, in an era of financial panics and eventually the

\textsuperscript{89} Jesse H. Rhodes, An Education in Politics: The Origins and Evolution of No Child Left Behind 169 (Cornell University Press 2012).


Depression, to help high school graduates compete for scarce employment opportunities. Test developer Edward Thorndike went so far as to devalue the teaching of any skills that lacked a direct application to the workplace. The Cleveland joint panel mentioned earlier provides further evidence. It addressed “The Problems of Education for the Industries,” another phrase that would likely be translated a century later as “Career Readiness.”

When the Efficiency Era transitioned to the Technocratic Meritocracy, a high school diploma and workplace readiness were no longer sufficient for those craving upward mobility. As discussed earlier, this phenomenon lead to the creation of the SAT, an aptitude-based college entrance exam, and the ACT, both of which would eventually be used to measure “College Readiness” for accountability models. The pendulum would continue to swing between preparation for careers and college, just as it did for excellence and equity in the years between NDEA and NCLB. In his Nation at Risk speech, President Reagan warned of a possible end of upward mobility for the next generation. To combat these issues, NAR demanded that schools

---


94 11 The Journal of the Cleveland Engineering Society 267 (1918). In the interest of full disclosure, E.S. Carman was the great-grandfather of the author.


produce highly skilled workers, including those who could function in a workplace replete with lasers and robots, signaling a renewed emphasis on “Career Readiness.”

In light of these findings, one can posit only that RTTT managed to marry workplace and college preparation in the term “College & Career Readiness.” Neither concept was new to education reform, and neither is likely to disappear any time soon. Since both types of readiness, and preparedness college in particular, are often measured with standardized assessments, it is no surprise that teachers are being held accountable for test scores—and likely will be for years to come.

Poor (and Immigrants, and Differing Abilities) Have Always Been Among Us

When discussing accountability, some will claim that the population of students attending public schools has changed dramatically, as if students from low-income homes, English Learners, and students with special needs only recently joined the data set. It is true that the disparities in achievement for low-income, English Learner, and special education students became more apparent when NCLB required that all students be tested and scores be disaggregated by subgroups. However, the presence of these students in the public schools of the United States can be documented as far back as the mid-nineteenth century.

---


As previously described, reference was made in source documents from Horace Mann’s Common School Movement not only to the poor but also to “foreign children who labor under many disadvantages” and came to school lacking fluent English. During the Efficiency Era, not only was the presence of immigrants noted in research, but at least one district was already offering what we might call English Learner classes today. In a report that foreshadowed NCLB’s subgroup disaggregation, the school success rates of students in one large-scale study were even compared by ethnicity. Although widespread special education was still decades away, reports from the era also noted teachers giving individual attention to students lagging behind their peers and those with behavioral or cognitive differences. Like today, there were those during the Efficiency Era who were not aware of history. Despite their documented presence more than fifty years earlier, these obtuse educational leaders referred to the challenge of educating students with “many new and different needs.”

100 Ronald G. Walters, American Reformers 215 (Hill & Wang 1997).
102 David F. Labraree, Someone Has to Fail: The Zero-Sum game of Public Schooling 87 (Harvard University Press 2012).
105 John Franklin Bobbitt, The Elimination of Waste in Education, 12 The Elementary School Teacher 266 (1912).
Although ESEA was initially linked to Civil Rights for African American students, its funding was provided to mitigate the impact of poverty for students of any ethnicity, establishing that the educational needs of low-income students weighed heavily on the minds of the leaders of that time. The NCLB subgroup data, then, was not a completely new concept in education. It was merely a logical outgrowth of reform from earlier eras. Nor is it shocking that the high standards and high stakes use of test scores of RTTT could exist during an era of increasing poverty, mass immigration, and the inclusion of students with special needs.

Litigation Slowed—But Did Not Stop—Race to the Top

Not only does history prior to RTTT support the durability of using test scores for teacher evaluation, but legislation and litigation in early-adopting states provide further evidence to this end. A New York court granted partial relief to the New York State Teachers’ Union as the result of litigation. However, that does not mean that test scores are no longer used as a data point in the evaluation of New York public school teachers. Although it can no longer be the sole factor in dismissing a New York public school teacher, legislation requiring the use of student achievement data in teacher evaluations remains encoded in state statutes.

Despite the law’s imperfections, legislators will likely be loath to repeal it, as this might convey a sense of low expectations for public school students. Even if the law were to eventually change, the RTTT repeal of the historic ban on using test scores in teacher

---

109 N.Y. State United Teachers, 929 N.Y.S. at 703.
evaluations would allow individual districts to continue the practice. In New York, the use of student achievement data as part of the teacher evaluation process did not, as many predicted, blow in and out like a tourist who enjoys the view from the Empire State Building, takes in a Broadway show, and caps it off with frozen hot chocolate from Serendipity before racing back to LaGuardia. Instead the use of assessment data to evaluate teachers has settled in for the long term. The holdings of the court did scale it back, denying it the co-op penthouse with Central Park views to which it had aspired. However, like a newly minted New Yorker who buys a nice little loft on the Upper East Side, the practice of using test scores for evaluative purposes has found itself what appears to be a permanent home in New York.

In Florida, not only has the law remained stable as it is in New York, but the fact that the case was heard in federal court could have national implications for future case law. Despite noting the problems with the statute, the judge dismissed the teachers’ complaint in its entirety. The court held that the termination of teacher employment due to test scores did not meet the criteria to apply the strict scrutiny test, subjecting it only to the less rigorous rational basis test. When the test was applied, the judge found the state had a rational reason to include test scores in the teacher evaluation process to further its legitimate government interest in

111 N.Y. State United Teachers, 929 N.Y.S. at 702 (citing Educ. Law §3012-c(2)(a)).
113 Cook v. Stewart, 28 F.Supp.3d 1207 at 1216.
educating children. In other words, according to the Constitution, government entities that fund education have a right to hold teachers accountable for students’ test scores so long as they have a rational reason to do so. The holdings also signal that a teacher’s job is not protected by the Constitution when employment is terminated based on student achievement. The durability of the use of test scores as part of the teacher evaluation process can be evidenced by the history of standardized testing and education reform and the actions of the courts in early-adopting Race to the Top states.

**Type of Test Matters**

The issues that triggered litigation in early-adopting Race to the Top states New York and Florida differed in several ways. One commonality between the RTTT plan in the two states, however, is the type of test used. Both states applied scores from high-stakes state tests to the teacher evaluation process. The purpose and rigor of state tests, along with the narrow subject matter and limited grade levels they assess, lack of transparency, and infrequent administration may have contributed to teachers’ negative reactions to their states’ plans.

In 2011, New York used the Regents’ Exam as its state test.\(^{118}\) As described in the previous chapter, the Regents’ Exam had initially served as a graduation-eligibility test and was repurposed following the 1939 Regents’ Study as an exam “designed to discover the weak spots in curriculum and teaching.”\(^{119}\) Both descriptions appear to indicate a high level of difficulty,


\(^{118}\) N.Y. State United Teachers, 929 N.Y.S. at 704.

and the latter even implies the Regents’ Exam may be what is often described as a “gotcha” test. In addition to a rigorous current exam, it was clear that educators in New York felt uneasy about preparing students to succeed on impending Common Core-aligned state tests.

The NYSUT (union) Truth in Testing website specifically addressed the need for professional development to better prepare teachers for Common Core testing. The union also demanded that sample test items be released, a reminder that teachers do not have ready access to state test content, meaning they would be judged on their students’ performance on unknown passages and questions. The anxiety was so significant that Governor Cuomo felt the need to run an advertisement in which he is shown sitting next to his daughter while she does her homework as he describes his willingness to delay Common Core testing until students are ready to meet the challenge. When an assessment becomes an issue in a gubernatorial campaign, it may be a portent of trouble if scores from high-stakes state tests are used for making employment decisions about teachers. Florida’s state test, the FCAT, was already aligned to similar rigorous new standards.

In both Florida and New York, the state tests used in the contested evaluation systems not only appeared to be rigorous, but were also limited in scope. As described in the previous chapter, Florida’s FCAT assessed only reading in grades three through ten, math in grades three

---


through eight, and science in grades five and eight. The New York Regents’ Exams were
arranged in a similar system at the elementary and middle school levels. At the high school level,
New York’s system covered more courses than Florida’s but was still far from comprehensive.
As a result, when scores from state tests were applied to the evaluation process, teachers of
specials (e.g. music, art, physical education) in grades three through five and instructors of many
middle and high school courses would be rated based on the performance of students in other
classes. The same would apply to all educators at the early childhood level, kindergarten, first
grade, and second grade. It is hardly surprising, in light of the incomplete grade and subject
coverage provided by state tests, that the attribution of student performance data to particular
teachers was one of the issues addressed in the Florida litigation or that all of the *Cook v. Stewart*
plaintiffs taught classes not assessed by a state test.

The difficulties of attributing student performance on state tests to teachers of courses not
measured using state assessments is exacerbated by the fact that these tests are typically
administered only once a year. Even if one believes in what Florida statisticians described as the
“durability” of a teacher’s impact on student learning, the situations created by the use of state
test scores as part of the teacher evaluation process were described as unfair even by the judge

---


who dismissed *Cook v. Stewart* in its entirety.¹²⁹ A kindergarten teacher, for example, would be evaluated by his or her past students’ achievement in fourth grade, the point at which two years of state data would be available to generate a growth score.

As described in the previous chapter, 23 states have attempted to increase the relevance and subject-area coverage of high school assessments by using either the ACT or the SAT—and in some cases the eighth, ninth, and tenth grade pretests from those systems—as part of their state testing systems.¹³⁰ While the use of the ACT and SAT as state tests has had the positive impact of increasing participation in college entrance exams, the history of these tests would indicate that they may not be suited to measure student growth as part of the teacher evaluation process. The historical SAT was deliberately designed to measure individuals rather than institutions.¹³¹ It was modeled after the intelligence tests previously created by its designers.¹³² While the ACT was based on the curriculum and designed to measure the efficacy of schools, its purposes also included serving as a means to “screen out the few who might least profit from college opportunities.”¹³³ It would appear that the use of scores from state-mandated administrations of college placement exams may not fill the gap in data needed for the growth component of the teacher evaluation process.


Creating new state tests for additional courses or grades is likely not a feasible solution either. The assiduous attention to detail described in the previous chapter demonstrates that assessment development is not a simple process. As described by creator E.F. Lindquist, the ACT staff took care in selecting passages and writing test questions. They also subjected them annually to a “thorough post-mortem.” Vintage-tinted color photographs of Lindquist and his colleagues resemble a nerdy version of the television show Mad Men. Suit-and-tie clad men peer at data through their horn-rimmed glasses as outsized early computing devices loom in the background. Post-test work also typically includes external validation, such as alignment to other assessments and college grades applied to the first administration of the SAT.

Race to the Top provides a more recent example of the effort required to develop large-scale standardized tests. RTTT consortia generated assessments for just eighteen courses. Those new assessments of English language arts and math in grades three through eleven cost $350 million for initial development alone. Even if test development were more simple and affordable, the NYSUT-proposed Truth in Testing legislation that would have banned any non-diagnostic state assessment below grade three appears to lay to rest any question as to whether

---


139 Steven Little & Angeleque Akin-Little, *Academic Assessment and Intervention* 166 (Taylor & Francis 2014).
teachers of courses currently not aligned to state assessments aspire to be evaluated using student performance on these tests.

Neither states’ testing system—New York or Florida—was unique in its lack of exhaustive course alignment. This non-comprehensive coverage is unlikely to change, both due to the uniform historical state testing requirements first imposed by NCLB\textsuperscript{140} and the sheer number of grade levels and subjects taught in United States EC-12 public schools. The size of the student data set yielded by state testing has recently been limited by another, relatively new, factor. As described in Chapter 2, with the advent of Common Core assessments, the United States has experienced a marked increase in the number parents opting their children out of state testing.\textsuperscript{141} The parent-supported refusal of students to participate in state tests has the potential to limit the validity of the student sample for even teachers of courses assigned to state assessments. If one assumes that parents who opt their children out of state testing are more involved and therefore likely to have children who are academically proficient, this could skim many top students from the results, skewing teachers’ student growth ratings based on state assessments.

Another potential parent-based challenge to using data from state tests for any purpose, including teacher evaluation, was foreshadowed in the original proposed Truth in Testing legislation. When NYSUT publicized the extent of student data—including birth dates, ID numbers, disability status, and other information submitted to systems managed by the state and assessment vendors in the proposed Truth in Testing legislation,\textsuperscript{142} many parents were likely

\begin{footnotesize}
\begin{enumerate}
\end{enumerate}
\end{footnotesize}
surprised by how much “Big Brother” knew about their children. While the data-privacy section of Truth in Testing was not included in the final legislation that was passed and enacted, parental questions about student data privacy could be raised again in the future.

Recent emphasis by educational leadership on data-driven instruction may have unintentionally created another obstacle—in this case teacher-based—to using scores from state tests as part of the teacher evaluation process. Some 21st Century teachers are data-savvy because they have been trained to analyze multiple data points either by district officials when making instructional decisions or by their unions when preparing for litigation. The plaintiffs in the Florida litigation and the evidence they provided in their complaint are illustrative of this. Music teacher Cathy McConnell was able to provide data to prove that her school’s orchestra program had grown during her tenure. McConnell’s students had also earned superior ratings at state-sponsored contests. Escambia County math teacher Shauna Paedae’s students had posted a 90 percent pass rate on International Baccalaureate exams. The connection of state tests to litigation in early-adopting RTTT states New York and Florida, the history and purpose of college entrance examinations required by many states, potential data privacy issues, and savvy teachers with access to multiple sources of achievement data indicate that the use of high-stakes state tests as indicators of student performance as part of the teacher evaluation process may not be prudent for districts wishing to avoid legal expenses.

Validity of Value Added is in the Eye of the Beholder

It was not particularly surprising that the teachers’ union in New York objected to the Regents’ Exam being counted as two assessments. A single test cannot become two exams simply by using achievement and unspecified growth data from the same assessment as separate measures. In contrast, Florida seemed to have made a considerable effort to arrive at a model that would be perceived as being valid by teachers. One might believe that models selected by a committee that received guidance from a well-known firm such as the American Institutes for Research (AIR) would be readily accepted by educators. However, a closer examination of the process reveals several potential challenges to the validity of professionally-developed growth models. Apparently, statistical growth modeling is not as scientific as one might perceive.

The first indicator that growth modeling is not an exact science became apparent at the Florida’s committee’s first meeting. AIR did not recommend a single growth formula as the most valid means of measuring teacher impact on student learning. Instead, as noted in Chapter 2, AIR presented eight models.\footnote{Recommendations of the Florida Student Growth Implementation Committee 13 (June 1, 2011),}\footnote{http://www.fldoe.org/core/fileparse.php/7503/urlt/0102687-value-added-model-white-paper.doc.} Within the report on the model options, the names of at least four researchers were mentioned: William Sanders,\footnote{Recommendations of the Florida Student Growth Implementation Committee 8 (June 1, 2011),}\footnote{http://www.fldoe.org/core/fileparse.php/7503/urlt/0102687-value-added-model-white-paper.doc.} Robert H. Myer,\footnote{Recommendations of the Florida Student Growth Implementation Committee 11 (June 1, 2011),}\footnote{http://www.fldoe.org/core/fileparse.php/7503/urlt/0102687-value-added-model-white-paper.doc.} Daniel McCaffrey, and J.R. Lockwood.\footnote{Recommendations of the Florida Student Growth Implementation Committee 9 (June 1, 2011),}\footnote{http://www.fldoe.org/core/fileparse.php/7503/urlt/0102687-value-added-model-white-paper.doc.} It is not beyond the realm of possibility that if a school district’s teacher evaluation committee were to select one of the eight models, at least two of the researchers behind the
competing models would be happy to testify in court that the model selected by school officials was invalid.

Furthermore, statistical growth models are not objective. As previously described, each formula is based on an underlying belief about the durability of teacher impact on student learning, thereby rendering each formula subjective from its very foundation. In addition to this idiosyncrasy, the Florida statute passed in response to RTTT added another layer of subjective value judgment. The Florida law expressly forbade the use of gender, ethnicity, or socioeconomic status as a variable in growth formulas although these factors are correlated to student achievement on standardized tests. Even without a statutory mandate to exclude particular student traits, any growth model would be limited as to the number of variables for which it could control or compensate.

An early hint that commercial growth models might be perceived as being overly complicated occurred when, as described in Chapter 1, John Oliver projected one on the screen, and his audience laughed. The growth model selected by Florida’s evaluation committee was awe-inspiring in its intricacy.

\[ y_i = \mu + \sum_{g=1}^{M} \delta_g x_g + \sum_{j=1}^{K} \beta_j x_j + \theta(\kappa)_i + \omega(\kappa)_i + \epsilon_i \]


One might imagine complexity would have afforded the model credibility. However, a popular media article quoted teachers as being stressed rather than impressed.\textsuperscript{152} Whether Oliver’s claim that such models are based on calculations pertaining to bovine reproduction\textsuperscript{153} is true, commercial growth models may not be the best method of measuring student learning in the context of the teacher evaluation process. There is no clear consensus among expert statisticians on how to best measure student growth. Additionally, models are subjective due to their basis in beliefs about the duration of a teacher’s impact on a student’s learning trajectory and permissible variables, and they are complicated to the point of serving as a comic foil for the general public.

Teachers Are More Than a Score: High Percentages May Mean Litigation

Both New York and Florida dictated in their post-RTTT statutes that a relatively high percentage of overall teacher evaluation ratings would be derived from students’ standardized test scores. However, it was not this high percentage expectation alone that triggered litigation. Instead the impact of the high percentage expectation that led to legal complaints was the potential to deprive a teacher of his or her job based solely on student test scores, regardless of performance in other components of the evaluation system.

In the case of \textit{Cook v. Stewart}, this was apparent in the debate over the plaintiffs’ legal standing in the United States District Court in Florida,\textsuperscript{154} when the court held that the plaintiffs’


\textsuperscript{153} \textit{Standardized Testing}: Last Week Tonight with John Oliver, (HBO, 2015), youtube.com, (February 23, 2017, 7:41 PM), \url{https://www.youtube.com/watch?v=J6lyURyVz7k}.

potential loss entitled them to legal standing.\textsuperscript{155} As described in the previous chapter, New York’s law specified that 40 percent of a teacher’s evaluation was to be based on student achievement as measured by state assessments. The primary issues raised by the plaintiffs in the New York litigation focused not on the percentage of evaluations attributed to student achievement but rather on the use of a data from a single test. Legally, the issue was described as statutory intent, or whether the regulations aligned to the spirit of the law.\textsuperscript{156} Nonetheless, the New York State Supreme Court ruled test scores could not constitute 40 percent of a teacher’s evaluation.\textsuperscript{157} The rationale behind this ruling appeared to run deeper than a mere number. It appeared the court’s holding was not due to the percentage \textit{per se} but rather due to the fact a teacher could be fired due to test scores even if he or she received perfect ratings on all other components of the evaluation process.\textsuperscript{158}

Florida attributed an even higher percentage—50 percent—of the teacher evaluation calculations to test scores. This, coupled with the final point system, could result in a teacher losing his or her job solely based on student performance on standardized tests. As described in the previous chapter, it was this potential property loss—i.e., the teachers’ potential termination of employment—that afforded the plaintiffs the necessary standing to have their case heard in federal court.\textsuperscript{159} While \textit{Cook v. Stewart} was eventually dismissed in its entirety,\textsuperscript{160} the court

\begin{itemize}
\item \textsuperscript{155} Cornell University Law School Legal Information Institute, \url{https://www.law.cornell.edu/wex/equal_protection} (last visited June 17, 2014).
\item \textsuperscript{156} \textit{N.Y. State United Teachers}, 929 N.Y.S. at 703.
\item \textsuperscript{157} \textit{N.Y. State United Teachers}, 929 N.Y.S. at 706 (citing Educ. Law §3012-c[2] [a]).
\item \textsuperscript{158} \textit{N.Y. State United Teachers}, 929 N.Y.S. at 706 (citing Educ. Law §3012-c[2] [a]).
\item \textsuperscript{159} \textit{Cook v. Stewart} Order Granting in Part and Denying in Part State Defendants’ Motion to Dismiss, No.1:13-cv-000072-MW-GRJ, (N.D.Fla. April 22, 2014) at 10.
\end{itemize}
noted that the defendant school districts had suffered significant losses as a result of legal fees and lost staff time.\textsuperscript{161} Regardless of outcome, once a complaint is filed in court, a school district has suffered a loss. Negative public perception in response to media stories, staff time, and legal fees all cost a school district dearly.

The high percentage of weight attributed to student test scores in the original RTTT rules in both New York and Florida appears to be correlated to the subsequent litigation. The words of the courts in New York and Florida appear to indicate that the ability of both systems to deprive a teacher of his or her job based solely on test scores was the primary problem—not the percentages themselves. This, in turn, seems to indicate the point system utilized in determining a teacher’s final evaluation rating—and the attendant weight it assigns to student test scores—could be equally troublesome to school officials if a teacher could be terminated due a low student growth rating based on student test scores when other areas of the teacher’s performance are rated at high levels.


\textsuperscript{161} \textit{Cook v. Stewart}, 28 F.Supp.3d 1207 at 1209.
CHAPTER 4
KEEPING THE MONKEY OFF A DISTRICT’S BACK

So long as states and school districts queue for federal funds and—much like the orphans in *Oliver Twist*—cry “Please sir, I want some more,”¹ this study’s findings would suggest that public education and accountability will continue to become increasingly enmeshed. Even with the demise of RTTT, it is important to note that states were required to encode the use of student test scores as part of the teacher evaluation process into their statutes. As a result, school district officials could find themselves continuing to use and revise their procedures for incorporating test scores into their teacher evaluation process. It would be helpful, then, for school district officials to learn from the lessons taught by prior litigation. Creating district or school plans for incorporating student achievement data into the teacher evaluation process that will withstand the test of the courts is insufficient. As established in *Cook v. Stewart*, even if litigation filed by teachers or unions is dismissed in its entirety,² prevailing in court does not necessarily constitute a win for a school district. The legal expenses alone could devastate a district, as evidenced by the fact that the defendant districts in Florida did not feel they had sufficient funds to even oppose the plaintiff teachers’ union-supported motion for summary judgment.³ If a legal complaint is filed, districts also face a hidden, but still dear, cost in the form of damage to their

---

¹ Nigel Bryant, *The Play of Oliver Twist* 97 (Heinemann, 1995).


public image. In Florida, this was limited to media coverage of the litigation.¹ In New York, the negative publicity was more widespread, impacting even the gubernatorial race² and involving parents concerned about data privacy due to information publicized by NYSUT on the Truth in Testing webpage.³ It would behoove districts, then, to develop plans for incorporating data from standardized tests into the teacher evaluation process that will enable them to avoid litigation altogether.

A litigation-resistant plan for including student growth data in the teacher evaluation process begins long before any discussion of tests or growth models commences. The factors correlated to increased accountability and testing identified in the previous chapter should be assiduously monitored by school districts. If districts are not proactive in not only analyzing but also responding to changes in the political and economic landscape, the media will likely control the narrative.

Administrators who keep apprised of and respond to societal shifts may be able to tell their own story and thereby have more latitude in designing their teacher evaluation growth data process. If a district does not maintain awareness of the world around it, it is likely to be subjected to greater scrutiny and have less freedom to design a litigation-resistant plan for including test scores in teacher ratings. There are four key societal factors—not unlike the Four

---


² Gift (Team Cuomo 2014), youtube.com, (October 21, 2014 7:41 PM), https://www.youtube.com/watch?time_continue=3&v=gJarD8z323o.

Horsemen of the Apocalypse— that appear to trigger increased accountability for public education in the United States.

Follow the Four Key Factors that Trigger Increased Accountability

The first factor of which school district officials should be aware involves political shifts at both the state and national levels. Since NCLB, federal accountability has been bipartisan and is unlikely to disappear. There are, however, shifts in priorities following changes to the executive or legislative branch at the federal level. During periods of federal stability, state elections can signal impeding changes to implementation regulations, particularly if the current federal system affords significant latitude to states. It behooves a district to track changes and adjust course accordingly.

Long before new policies are implemented, the campaign rhetoric of political candidates, including both written materials and speeches, can serve as a portent. If—as was the case with Presidents Johnson and Reagan—a candidate remains relatively silent on education during a campaign or, in the case of President Obama, speaks only to criticize the policies of the opposing party without detailing his or her plans, there is another means of predicting his or her future views: follow the money. History has demonstrated that donors influence the education policy of candidates. In the case of the Obama administration, for example, while candidate Obama decried what he perceived as an over-reliance on standardized testing during his campaign, the

---


fact that he was connected to entities such as the Gates Foundation\(^7\) would have served as an accurate predictor of the wave of the lengthier and more rigorous assessments his administration would unleash on K-12 public education.

Once candidates are elected, school district officials should assiduously follow and summarize the progress of education policy. This can be accomplished by using widely-available state and federal government tracking websites\(^8\) or joining an advocacy organization tailored to the size, location, and demographics of the district.\(^9\) To avoid devolving into mere sycophants, school district officials should thoughtfully analyze potential policies through the schema of two simple questions: “Is implementing this policy feasible?” and, most importantly, “Would it benefit our students?” If the answer to both questions is yes, a school district stands not only to be better prepared for future accountability but also positions itself to tell its own story by planning for early implementation of initiatives rather than having it told by the media.

The second factor of which districts should remain apprised—conditions attached to funding—is directly related to the first factor. As new policies are officially implemented, it is worth noting that they are often linked to grants. At the federal level, for example, from the National Defense Education Act\(^10\) to Race to the Top\(^11\) over half a century later, reform has been implemented voluntarily by attaching conditions to the acquisition of funds. In some instances,

---


\(^9\) About Us, Large Unit District Association Website (November 11, 2017, 1:28 PM), [http://www.ludaschools.org/about/](http://www.ludaschools.org/about/).


states make a decision to accept grants, but in others, a local school district may have the volition
to accept or decline participation in a grant.

When faced with a choice rather than growing starry-eyed at the prospect of increased
cash flow, district officials should approach grant opportunities with eyes wide open, thoroughly
reviewing all documentation related to voluntary funding, and according special attention to
accountability requirements. After a cost-benefit analysis to determine whether the grant funds
will truly add value to district efforts, district officials would be well-served by ascertaining if
any associated costs can be paid from the grant funds. Counter-intuitive as this may sound, this is
not always the case. For example, grants may forbid expenditure of funds on assessments or data
systems and may also limit the extent to which funds can be used to pay for administrative
expenses.

The public-sector trends are in turn often influenced by the third factor of which school
districts would be wise to keep apprised: the economy. Both economic shifts and downturns—or
the convergence of both as occurred during the Race to the Top era—influence the degree of
accountability foisted on the K-12 public schools in the United States. School districts may
benefit from monitoring economic conditions and telling their story before the government or
media tell it for them.

On the most basic level, school districts should be prepared by telling the story of, to
quote President Ronald Reagan, “What have we bought with all of that spending?”12 School
district officials may be well-served by implementing a cycle through which to review program
costs and effectiveness to operate as efficiently as possible without making cuts that could

12 The American Presidency Project: Ronald Reagan Remarks on Receiving the Final Report of the National
Commission on Excellence in Education , University of California Santa Barbara,
negatively impact students and by communicating frequently to the community about these efforts. Efficiency alone is not sufficient, however. History has illustrated that both economic downturns and shifts invite scrutiny of the skills students are gaining in schools to help them compete during challenging economic times.

There are steps school districts may wish to take to prevent becoming a scapegoat during economic downturns, and more importantly, to prepare students to compete in economic shifts. In regard to career readiness, a district might benefit from maintaining two-way communication with local businesses to gain knowledge of in-demand workplace skills and articulate the district’s accomplishments. An additional step would be to create partnerships that allow businesses to share their expertise and funds. When implementing the action plan generated by the aforementioned efforts, school district officials would be well-served by making decisions about course offerings and content based on the needs of students rather than staff convenience.

In addition to developing career readiness, diverse course offerings can also contribute to students’ college readiness. Standardized test scores, however, are perhaps the most publicly recognized aspect of college readiness. To ensure they are preparing students to enter credit-bearing college courses upon graduation, school district officials should regularly analyze and share data from the ACT or SAT suite of assessments with staff in a format that readily lends itself to designing instructional improvements. School districts can demonstrate a more holistic approach to college readiness by developing systems to track and analyze not only conventional metrics such as grades and test scores but also activities and service hours.

The impact of economic downturns and shifts on education can be magnified by the fourth factor of which school district officials should remain apprised: demographic changes to its student population. An astute district might respond to this reality by regularly reporting to
stakeholders not only student demographics pertaining to income levels, mobility, ethnicity, and native language but also the academic and social-emotional needs of any new or increasing groups. This process would empower district staff to plan improvements to ensure that all student populations are learning at high levels. To measure the efficacy of any changes made to instruction, district officials would also be well-served by analyzing existing assessment data for new student groups. If current assessments are not sensitive to the learning growth of all student groups, a district would be wise to devise new assessments or metrics.

Select the Right Assessments: Tailored, Timely, and Transparent

Following these steps to establish an instructional program, metrics, and messaging will provide a district with the best odds of securing the freedom to design at least some aspects of the student growth component of its teacher evaluation system. Latitude alone is not sufficient, however, for designing a litigation-resistant protocol for using standardized test scores as part of a district’s teacher evaluation system. Assiduous care must be taken in selecting the assessments that will generate data for the process.

First and foremost, school district officials should collaborate with a variety of stakeholders when designing or revising the student growth portion of the district or school’s teacher evaluation process. Lack of collaboration was particularly evident in New York, as it was an issue raised in the statutory intent portion of the litigation. While it was not stated explicitly in the legal complaint, a lack of stakeholder input was likely also an underlying issue in Florida. The high quality of teachers filing the suit, as measured by the accomplishments listed in the complaint, appears to indicate that Florida’s process design could also have been more collaborative.
In the context of assessment selection, participation from a broad group of stakeholders should include teachers from each grade level and secondary department. District leaders should listen carefully to concerns, not only about the majority of staff, but about outliers (i.e., to avoid a situation similar to Florida). However, the facilitator of the task force should communicate legal parameters clearly at the outset to avoid expending time admiring concerns not within the locus of control of the school district. The process should also include holding committee members accountable for communicating to and soliciting feedback from others within their grade level or department. Bearing in mind the results of the hasty process in New York, school districts would be wise to invest as much effort as needed to arrive at an acceptable solution earlier rather than spending time on litigation later. The assessments selected or created by the task force should be tailored, timely, and transparent.

Tailored, in the context of assessments, refers to alignment not only to content but also to the students who will take the test. As evidenced by the Florida litigation, teachers can be data-savvy and expect assessments that generate data for their evaluations to be aligned to the courses they teach. Attention should also be paid to a test’s sensitivity to the growth of different student groups, especially in regard to the aforementioned correlation of demographic shifts to increased accountability. If a school district decides to expand assessments used as part of teacher evaluation beyond the typical general skill areas of reading, math, and science, the task force should identify any grades or courses currently lacking common assessments and outline a plan to create or purchase them.

The mere existence of assessments, however, does not suffice to ensure that they will generate the most accurate growth data. The task force should also determine whether existing assessments contain items at a variety of rigor levels to be sensitive to the growth of a variety of
student populations. This applies not only to groups who may be lower-performing but also to high-performing students. Finally, assessments used to provide data for the teacher evaluation process should be available in a pre-post system, preferably by administering the exact same test to students twice, at least a few months apart. Another option for measuring growth would be to use parallel forms: tests that contain different items but are found via statistical analysis to be of equal rigor. However, using parallel forms instead of re-administering the same assessment could leave a school district vulnerable to litigation if the results do not demonstrate sufficient growth. Based on the varying opinions of statistical experts on other issues pertaining to student growth in Florida, it is reasonable to expect that the opinions of psychometricians could also vary regarding the relative rigor of parallel forms, leaving a school district vulnerable to litigation if the results do not demonstrate sufficient growth.

The second T of assessment selection or creation is timeliness. As evidenced by *Cook v. Stewart*, assessments that do not generate timely scores can generate problems, particularly in the context of attributing student growth to teachers. Ideally, an assessment to be used as part of the teacher evaluation process should yield both pre and post scores within the final evaluation rating deadlines of a single school year. This avoids any debate over the durability of teacher effect on student growth, such as occurred in Florida. When selecting or creating a timely test, district officials should consider both item format and scoring method. It is worth noting that the ACT and the SAT gained control of the college-entrance market in part by using multiple choice

---


items and employing the latest scoring technology. Forced choice may be the best format and automated scoring the most feasible method.

The third T of test selection or creating is transparency. The NYSUT demand for the release of test items and score samples illustrated teachers’ distaste for being assessed using unknown content or scoring methods. Transparent content and scoring also increases teachers’ chances of success, which in turn reduces the likelihood of litigation. Tests—including passages used for reading assessments—should be readily available for teacher review, ideally before the pre-test is administered or, at a minimum, after post-test scores are released.

Districts that choose to create tests locally would be well-served by devising a development cycle that includes teacher input on item selection, district-wide piloting, validity data sent to teachers when soliciting feedback, stabilization, and cut score setting. School districts might also want to consider scoring assessments using a system that allows teachers to view items while analyzing scores. If open-ended or performance items—those with no clear cut correct answer—are used, school districts should provide or procure clear scoring rubrics that can also be used as part of instruction along with scored work samples with ratings for teachers to review. Teachers should also dual-score this type of item with other teachers for inter-rater reliability.

If given a choice of assessment models for use as part of the teacher evaluation process, district officials may wish to avoid using state tests. Although using pre-existing assessments with perceived validity may be easier, it is not necessarily pragmatic for a district wishing to avoid litigation. Both New York and Florida faced litigation surrounding the use of test scores as part of the teacher evaluation process—and both used state tests for the student achievement portion of their teacher evaluation systems. Furthermore, many Common Core state tests are
controversial, not only with staff but with parents. Due to this, parent-driven opt-out practices can remove students—in many cases high-performing children—from the data set, leading to skewed results.

State tests can also present a challenge relative to transparency. In many cases, the passages used as part of state tests are subject to copyright and are thus unavailable. Working with the district stakeholder group to evaluate assessments against the Three Ts of Testing criteria described earlier may bring the group to this conclusion without leaders having to explicitly direct that state tests not be used as part of the teacher evaluation process. If the group still favors state tests after this evaluation, a quick review of litigation in RTTT states might be helpful as a next step.

Sorting tests, such as the ACT, the SAT or the Iowa tests, were not used as part of the teacher evaluation process in New York or Florida and thus are not directly correlated to RTTT-related litigation. Some may posit that high school students are likely to make an effort on the ACT or the SAT because they are tied to college entrance. However, as described in Chapter 3, their historical purpose, which would appear to indicate that it is impossible for 100 percent of students to receive passing scores, calls into question their suitability as teacher evaluation tools.

As was the case with state tests, working with the district stakeholder group to evaluate assessments against the Three Ts of Testing criteria may help bring the group to this conclusion without leaders having to explicitly direct that sorting tests not be used as part of the teacher evaluation process. If the group still favors state tests after completing the evaluation process described in the preceding paragraphs, a quick overview of the history of sorting tests may help the group understand the potential ramifications of using scores from these assessments as part of the teacher evaluation process.
When Designing a Growth Model, Perceived Validity Trumps Statistical Validity

Even if a district masterfully collaborates to select or create assessments that correlate to all grades and courses taught, it would all be for naught if the growth model used to rate teachers based on their students’ scores is not embraced by staff. As described in Chapter 3, Florida hired the American Institutes for Research (AIR)—a prestigious research and assessment firm—and conducted several meetings with a variety of stakeholders to select a statistically valid growth model. The resulting litigation, however, appears to indicate that perceived validity (by teachers) may be more important than statistical validity. This may be because, as illustrated in the descriptions of the potential Florida models, even “scientific” models begin with a researcher’s subjective belief about teacher effects on learning at their base.¹⁶

The fact that statisticians disagree on which models are best should be troubling to districts. If one statistician develops a model for a district, there are likely several others who would be happy to testify against it in court as expert witnesses, resulting in a veritable War of the Nerds in the courtroom. It is true that creating a growth model in-district will involve taking full responsibility with the district board for the results, which appears risky. However, based on Cook v. Stewart, there also appears to be significant risk of litigation when using an externally-generated growth model. If needed, task force facilitators could share a summary of the Florida white paper on growth models with staff or board members to disabuse them of the notion that commercially-developed growth models are scientific.

If there is no scientific best growth model for student scores on standardized tests, it might behoove a district to instead use a model that is comprehensible to teachers. Yes, statisticians would relish the opportunity to smugly rip it apart in court, but they are less likely to have that opportunity if teachers perceive it is valid and refrain from filing a legal complaint against their school district. Based on the Florida white paper describing the various growth models presented by statisticians, validity apparently is in the eye of the beholder after all. If the beholder is a kindergarten or English literature teacher, it might behoove a district not to use growth formulas that include Greek mathematical symbols.

Instead a school district may be well-served by developing a simple growth model using only the four basic mathematical operations: addition, subtraction, multiplication, or division. Establishing the mathematical formula is the first step, after which school district officials would be wise to create presentation materials that clearly communicate it to staff. Time permitting, school districts should consider piloting the model before it becomes high stakes, or, if using existing assessments, retroactively simulate a pilot using historical data to show teachers how they would have been rated.

Take Care When Attributing Scores to Staff

The step of attributing student growth scores to particular teachers should be rendered far easier if a district follows the steps above and considers societal factors, selects the right tests, and uses their assessments to generate simple growth scores with a district-developed growth model. However, districts still need to account for teacher self-efficacy and the climate needed to successfully maintain collaborative structures such as PLCs (Professional Learning
Communities) as well as staff and student mobility. The first step to holding educators responsible for student learning is to teach the teachers.

A key demand of NYSUT concurrent with the New York litigation was for teacher training on content related to the assessments being used as part of the evaluation system. Also, as stated in a previous section, teachers equipped to help their students succeed are more likely to be open to being held accountable for scores—and might be less likely to take a district to court. Most importantly, professional development for teachers may benefit students.

Expanding on the concept of selecting a transparent test, an effective first step for school district officials might be providing teachers with an opportunity to thoroughly review the standards to be assessed and the items—including any passages—used to assess them. To ensure that teachers—and more importantly students—are equipped to succeed, school district leaders should also solicit feedback from staff. This includes asking teachers what training and resources they need for their students to master the standards at the level of rigor represented on the assessment items. Providing support does not necessarily entail procuring the services of costly outside providers. School district officials can use data—both test scores and qualitative observations from principals—to identify in-district experts when possible. If not possible for some standards, school districts may need to find external providers, but they should consider building capacity within the district by using a trainer-of-trainers model.

Mobility is another important consideration for school district officials when attributing student test scores to particular teachers. Student mobility has three primary causes. Some student populations (e.g., low income, military, English learner) tend to have higher levels of mobility. In other cases, school districts matriculate students to new schools at certain grades, as was the case with the primary (grades K-2) and intermediate (grades 3-5) schools described in
the *Cook v. Stewart* complaint. In still other instances, staff may change grades, courses, or even schools from year to year.

If a district follows the advice in the Test Selection section and uses assessments that can deliver pre and post scores within a single year’s evaluation cycle, it should significantly reduce the problems generated by mobility. If a district must use a test for which pre and post scores are generated during two school years, it would behoove the district to link the pre and post scores student-by-student using a unique identifier as a primary key (as defined in Chapter 2). This is particularly important in schools with high student mobility or matriculation between grades.

In an era during which teachers often work collaboratively in PLCs (Professional Learning Communities), some districts may feel a school-wide or other group rating (e.g., by grade level or course) is preferable, as it prevents a sense of competition from paralyzing PLC work. A group rating provides the added benefit of a larger more reliable data set than an individual class. It could also make teachers more willing to be assigned students with learning challenges. However, in the Florida litigation, the group rating created problems. In summary, both individual and group ratings bring challenges when student test scores are applied to the teacher evaluation process.

The solution depends on a district’s configuration and culture. District officials would benefit from convening a representative group of stakeholders to share the following questions, take them back to their colleagues, and return to discuss the responses: Do we have tests of enough subjects to individually rate teachers? How many students are in the smallest class attributed to a teacher? What are the advantages of group metrics? If we use a group metric, how should the teacher groups be configured? What if a teacher is a member of more than one PLC?
If possible, district staff should report the projected results of each model (group and individual) and review with the stakeholder group before making a final determination.

Avoid Punitive Percentages if Possible

The litigation surrounding the use of student achievement data as part of the teacher evaluation process in New York\textsuperscript{17} and Florida\textsuperscript{18} differed in many ways. One important commonality described in the previous chapter, however, was the attribution of a high percentage of a teacher’s final evaluation rating to student academic growth as measured by tests. If possible, districts would be wise to avoid assigning a high percentage to the student growth portion of the evaluation. School districts bound to a high percentage by state school code could employ other strategies to reduce the odds of becoming embroiled in litigation.

If a school district has the latitude to select a higher percentage, some might ask why district officials would not employ that option to eliminate less effective staff. However, educational leaders should not depend on test scores to remediate or terminate under-performing teachers. If students are not learning from a teacher’s instruction, most current evaluation tools in the hands of a competent principal provide ample opportunities to account for deficiencies through gathering evidence and observing instruction. Furthermore, as established in previous chapters, attributing a high percentage of a teacher’s final evaluation rating to student test scores appears to be correlated to litigation.

\textsuperscript{17} \textit{N.Y. State United Teachers}, 929 N.Y.S. at 706 (citing Educ. Law §3012-c[2] [a]).

If the state offers districts a range of percentages of a teacher’s final rating that can be attributed to test scores, school districts would be wise to use the lowest percentage. If any stakeholders oppose this, a potential compromise would be to re-train principals on other components of the evaluation that could identify teachers in need of improvement. If a high percentage is unavoidable, a district may benefit from considering the strategies that follow.

Due either to state school code or stakeholder concerns, some districts may not have the option to adjust percentages. Depending again on school code, the final matrix that converts numerical ratings to text (e.g., satisfactory, needs improvement) could possibly be adjusted to prevent punitive final ratings for teachers. Litigation in New York and Florida, while correlated with high percentages of evaluation ratings attributed to test scores, was not about the numbers per se. Rather, a key Florida issue was a teacher’s job as a property right, and a central New York issue surrounded whether a teacher could be dismissed based solely on test scores. Faced with an inflexibly high percentage mandate, a school district has an option that could prevent the type of unwarranted teacher job loss that triggers litigation. The rigor of the tests and growth metrics used should be inversely proportional to the percentage of a teacher’s evaluation it comprises and the final employment consequences of the student growth data.

All of the aforementioned recommendations involve significant time and effort on the part of school district officials and other staff to customize the use of test scores as part of the teacher evaluation process. Some Florida districts made the mistake of taking the path of least resistance accepting the state model when they had the option to design their own. Most likely, in retrospect, they would have preferred arduous committee and data work to the litigation that ensued. It may benefit school district officials located in later-adopting states or those faced with the challenge of revising their process to follow the recommendations of this study. Otherwise,
they may one day find themselves shouting “Here comes the monkey!” as a process server
approaches bearing notice that they are being sued.

Areas for Further Study

This study analyzed the history of standardized testing, the progression of education
reform in the United States, state legislation in response to the federal Race to the Top (RTTT)
program, and litigation surrounding the use of test score data as part of the teacher evaluation
process in early-adopting RTTT states. The intent of the research was to identify best practices
for incorporating student scores on standardized tests into the teacher evaluation process. After
analysis of litigation in six RTTT states, the focus of the study was narrowed to suits filed in
Florida and New York. The cases in New York and Florida were moving more rapidly through
the courts as compared to the others. Litigation in the two selected states also encompassed the
key issues central to other complaints filed.

Extending the study to include the other RTTT states in which litigation has been filed
could provide further guidance to school districts as they use student growth data from
standardized tests as part of the teacher evaluation process. While the issues behind litigation in
Michigan, New Mexico, Tennessee, and Texas were similar to the ones addressed in this
research, the actions of the state or federal courts in those cases could prove to be different.
Furthermore, those states’ legislative responses to the litigation could also be a valuable addition
to the field of legal research.

Expanding the study to include RTTT states in which no litigation has been filed would
also be an informative inquiry. An analysis of not only the state statutes, but also the
implementing regulations and resulting policies of school districts within these states could
contribute to the body of best practices research. Additionally, the study could be furthered by monitoring RTTT states to ascertain if state statutes requiring the use of student test score data as part of the teacher evaluation process are maintained as the RTTT era fades away and takes its place in history alongside Horace Mann and the Common School Movement.