TEACHER QUALITY 2.0

AMERICAN ENTERPRISE INSTITUTE

SPECIAL REPORT 1

From Teacher Education to Student Progress

Teacher Quality Since NCLB

Arnold F. Shober | August 2012

Foreword

There is incredible interest and energy today in addressing issues of human capital in K–12 education, especially in the way we prepare, evaluate, pay, and manage teachers. States have been developing and implementing systems intended to improve these practices, with a considerable push from foundations and the federal government.

As we start to rethink outdated tenure, evaluation, and pay systems, we must take care to respect how uncertain our efforts are and avoid tying our hands in ways that we will regret in the decade ahead. Well-intentioned legislators too readily replace old credential- and paper-based micromanagement with mandates that rely heavily on still-nascent observational evaluations and student outcome measurements that pose as many questions as answers. The flood of new legislative activity is in many respects welcome, but it does pose a risk that premature solutions and imperfect metrics are being cemented into difficult-to-change statutes.

AEI's *Teacher Quality 2.0* series seeks to reinvigorate our now-familiar conversations about teacher quality by looking at today's reform efforts as constituting initial steps on a long path forward. As we conceptualize it, "Teacher Quality 2.0" starts from the premise that while we've made great improvements in the past ten years in creating systems and tools that allow us to evaluate, compensate, and deploy educators in smarter ways, we must not let today's "reform" conventions around hiring, evaluation, or pay limit school and system leaders' ability to adapt more promising staffing and school models.

Before we can begin to address these issues, we have to reflect on how we got to our current position. In this first installment of the series, Arnold Shober, associate professor of government at Lawrence University, sets the historical context for this 2.0 dialogue by tracing the evolution of our conception of teacher quality throughout the past decade. Shober describes the changing political pressures, milestone policy changes, and notable research findings that have molded today's approaches to boosting educator effectiveness. He argues that we have seen a shift in paradigm from trust to measurement, adjusting our view of teaching from something immeasurable to something we can parse and compare. Although these developments have created challenges of their own, they have encouraged schools and school districts to think about teacher effectiveness in a new way.

I am hopeful that you will find Shober's paper to be as enlightening and thought-provoking as I have in helping make sense of the road we've paved to reach our current state. For further information on the paper, Shober can be reached at arnold.shober@lawrence.edu. For additional information on the activities of AEI's education policy program, please visit www.aei.org/hess or contact Lauren Aronson at lauren.aronson@aei.org.

—FREDERICK M. HESS
Director of Education Policy Studies
American Enterprise Institute

Executive Summary

The concept of "teacher quality" has undergone a profound transformation in the last decade. Through the late 1990s, most policymakers assumed that educator effectiveness was immeasurable and that our only hopes to increase it were tied to classroom experience and academic credentials. Yet since 2001, through a series of notable research findings, changing political pressures, and landmark policy changes, we have come to view teacher quality as independent of licensure and individually measurable. We now approach evaluating the quality of our teachers by measuring their ongoing performance in the classroom. The key markers in the transformation:

- **Declining trust in teacher education:** Data from the Higher Education Act of 1998 showed that teacher education programs graduated students of widely variable ability.

- **Ineffective credential requirements:** Despite federal requirements for high-quality teachers, by 2006, students in high-poverty schools were still more likely to have lower-quality teachers than their peers in other schools.

- **Bipartisan agreement:** By 2009, major federal politicians in both political parties, including Democratic politicians heavily supported by unions, agreed that low teacher quality was a barrier to high-quality education.

- **New era of measurement:** No Child Left Behind created a treasure trove of individual student data that, though imperfect in many ways, led to new opportunities for research to link teachers and students.

No consensus has emerged about how to improve teacher quality, but policymakers continue to experiment with teacher preparation programs, recruitment incentives, tenure provisions, and differential pay.

From Teacher Education to Student Progress

Teacher Quality Since NCLB

Arnold F. Shober

It is hard to imagine a time when teacher quality was not a central piece of American education policy, yet the meaning of "quality" has undergone a profound transformation in the last decade. Through the late 1990s, policymakers and district personnel trusted teacher credentials as a marker of quality in the education system, and they paid little attention to the variation in classroom effects despite the common credential. Since then, the combination of a renewed public emphasis on improving academic achievement and new research on teacher effectiveness has prompted policymakers to question the trustworthiness of linking certification to quality.

Equating quality with qualifications was a reasonable assumption before the 1990s as the learning process was not well understood, and an emphasis on training and retaining expertise was appropriate under such circumstances. In effect, earning a credential through training meant that teachers could be trusted to make the right decisions on their own in an uncertain educational environment. Further, in the mid-twentieth century, the National Education Association (NEA) and the American Federation of Teachers (AFT) took the view that teachers were professionals first, who needed job security to reinforce public trust in their decisions, and that teachers could be trusted to self-regulate, claiming that "what teachers want is what children need." To preserve teacher experience, the nation's major teachers' unions prevailed upon state legislatures to extend collective bargaining to them in the name of teacher quality. Later, in the 1980s, many states addressed teacher quality by requiring new teachers to have an official teacher-mentor. In their view, certified teachers could be trusted to do what was right for children without outside supervision—and, by implication, were interchangeable—and they largely succeeded. Once certified, one elementary teacher was as good as any other elementary teacher. At century's end, both the 1998 reauthorization of the Higher Education Act (HEA) and the 2001 passage of No Child Left Behind (NCLB) accepted the view that quality was the combination of training and experience.

Arnold Shober (arnold.shober@lawrence.edu) is Associate Professor of Government at Lawrence University.

But by the late 1990s, it was abundantly clear that this definition was fiction. Teacher *qualifications* did not guarantee teacher *quality*. Thirty years of data showed that students systematically learned more in some classrooms than others and that disparities in learning could be tied to disparities in teacher quality. Certified teachers were not, as it turned out, interchangeable; they were individuals with strengths and weaknesses. Trusting credentialing to produce teachers with equal abilities has hindered recruiting the best and brightest college graduates into teaching, prevented schools from rewarding the best teachers, and hurt low-income, low-performing students as teachers sort themselves into more desirable schools. Policymakers began to look back further in the policy chain. If credentialing could not be trusted to produce reliable inputs to K–12 education, then what could?

Even as the policy problem crystallized, "boosting teacher quality" was not a well-defined set of policy prescriptions in 2001. Foreshadowing forty years of research, James Coleman's *Equality of Educational Opportunity* (EEO) reported in 1966 that "the quality of teachers shows a stronger relationship to pupil achievement [than school facilities], . . . [but] the results are not at all conclusive regarding the specific characteristics of teachers that are most important."[1] As late as 1999, Dan Goldhaber and coauthors suggested that 97 percent of teacher quality was "intangible." For researchers and policymakers, then, identifiable components of good teaching were a mystery.[2]

But the lack of a clear theory of action did not stop policymakers from constructing a causal narrative about what some thought *should* make a good teacher: improving teacher education and increasing teacher credentials. In light of the dramatic uncertainty surrounding teaching, these were some of the few items that policy could control. Federal policymakers adopted this approach in HEA and NCLB, following a path well trodden by state departments of education.[3] But, as with earlier teacher quality policies, this status-quo trust of credentials for teacher quality was long on gut feeling and short on

empirical evidence. Increasing funding for colleges of education, even with additional oversight, did not upset existing legislative-political alliances with universities. Increasing requirements for licenses tinkered with the front end of teachers' careers and did not threaten teachers' unions' membership.

Thirty years of data showed that students systematically learned more in some classrooms than others and that disparities in learning could be tied to disparities in teacher quality.

Yet since 2001, state and federal policymakers have accomplished a revolution in the conceptualization of American teacher quality in ways that were unanticipated by either HEA or NCLB. The change in policy is twofold. First, policymakers subtly shifted the framing of teaching policy. Instead of treating teaching as difficult to parse and compare, policymakers now suggest that the outputs of teaching are understood well enough to quantify, measure, and manipulate them. When policymakers regarded credentials as a reliable input, they considered the autonomous decisions that teachers made in uncertain circumstances generally trustworthy. Teachers often supported this view by arguing that they needed to tailor their methods to different students' unique needs, which could not be anticipated by uniform policy prescriptions. Thus, little effort was made to measure the value of a credential to teachers' performance in their classrooms. Now, as policymakers believe they understand the process of teaching and learning better, they seek to *verify* the value of teachers' credentials by setting a minimum, predefined, and assessable standard of quality. NCLB aided this change through its emphasis on measurement, particularly on reading and math performance, and its narrow focus on student assessment, thereby allowing scholars to link teacher quality to student outputs with little lag time. This was a major advance from the 1990s, when scholars were still analyzing thirty-year-old EEO data, picking through data from Tennessee's class-size-reduction experiment in the 1980s, and drawing weak claims about quality using cross-sectional administrative figures.

Second, politicians in both major parties adopted this framing. More to the point, major *Democratic* politicians argued that trusting credentials to ensure student performance undercut educational equity. Thus, overwhelmingly Democratic teachers' unions have struggled to halt evaluation based on student performance, differential pay, and alternative certification because their fellow partisans in government champion these policies.[4] Credentials were meant to ensure that teachers were highly qualified as a group, but credentials have lost legislators' trust. Legislators and policymakers want to measure the value of individual teachers.

As policymakers and scholars debate the reauthorization of the Elementary and Secondary Education Act (ESEA) in 2012, reconceptualizing teaching as a measurable endeavor—partnered with new evidence about effective teaching from NCLB data—has made possible meaningful, long-term improvements to teacher quality through federal law. This essay traces the development of this new focus, considers potential causes for the change, and suggests how major teacher effectiveness proposals fit into this paradigm.

From Qualifications to Quality

Lawmakers passed the ESEA in 1965 in the interest of advancing equity for low-income, nonwhite children, primarily through supplementing local and state school spending with federal dollars. The focus of the act, implicit in the title of Coleman's companion study, was creating educational *opportunity*, with less focus on educational outcomes. Policymakers recognized that they had little evidence regarding effective practices within schools, but they fell back on a reasonable guess: if low-income and nonwhite children had access to the same resources as middle- and high-income white children, their academic performance would follow. Unfortunately for the act's original supporters, evidence from the National Assessment of Educational Progress (NAEP) in the subsequent thirty years suggested that the academic performance of nonwhite and low-income children showed little improvement given the scale of federal involvement. Simultaneously, a handful of state court decisions in the 1980s and 1990s—notably *Rose v. Council for Better Education* (1989) in Kentucky—argued that untargeted spending would do little to overcome these educational disparities.

The combination of ESEA's stated concern for educational equity and stubbornly low performance left policymakers in a bind. Clearly, the focus on input measures

would neither make the resources in low-income schools equal to their upscale counterparts nor substantively improve students' academic performance. This conundrum opened the window for cross partisan experimentation and presaged the shift in the conceptualization of teaching. While ESEA would continue to focus on equity, the act's emphasis would shift to equality of outputs. The 1994 reauthorization of ESEA, the Improving America's Schools Act (IASA), called for academic state standards, academic testing, and wider participation in NAEP, edging federal concern toward results. Seven years later, NCLB added teeth to IASA's requirements, but the law only nudged federal policy to redirect its focus toward the central component of classroom-based learning: teachers.

The central role accorded teachers at the turn of the century marked a significant shift from 1965, when policymakers believed financial resources were at the core of education achievement. Nonetheless, policymakers in 1998 and 2001 had little more information about teachers' direct influence on students than policymakers had in 1965. Scholarly evidence available in 1998 suggested that strong teachers could improve student academic achievement compared to other school factors, but the *how* remained unclear.[5] Teachers were the closest to the learning process, and the uncertainty surrounding effective teaching practices encouraged policymakers to maintain a veil over the practice of teaching. They continued to target only the entry points into teaching: certification and colleges of education.

Teacher Quality as Teacher Education. The opening salvo in the fight leading from IASA to HEA and NCLB came from the National Commission on Teaching and America's Future (NCTAF). Its 1996 report, *What Matters Most: Teaching for America's Future,* reported that 27 percent of newly hired teachers did not have standard credentials when beginning their work—almost half of these had no credentials at all.[6] More damning was that these teachers were concentrated in the highest-poverty, least-white schools. For example, the report documented that only 54 percent of math teachers in predominantly nonwhite high schools had a major in their field, while 86 percent of teachers in predominantly white schools did.[7] NCTAF also noted that the requirements for earning a teaching license varied widely among the states and that teacher preparation programs in institutions of higher education were uninspiring, fragmented, superficial, and backward-looking. In short, teacher preparation programs were "caught in a vicious circle of mediocre practice modeled after mediocre practice."[8] These delinquencies prompted Rep. George Miller (D-CA), then the second-ranking Democrat on the House Committee on Education and the Workforce, to argue that teacher preparation programs were "perpetrating a fraud on the public."[9]

Among other things, the NCTAF report recommended that standards for *both* teachers and students be raised; in particular, it urged increased standards for licensure and a thorough overhaul of teacher preparation programs. The bipartisan cast of authors who produced the report, including Ted Sanders, Albert Shanker, Jim Edgar, and James B. Hunt, gave it significant sway in the Clinton administration and the Republican-controlled Congress. President Bill Clinton and Senator Jim Jeffords (R-VT) echoed the report's conclusions: the United States needed more teachers; those teachers needed more subject-matter training; teacher education was divorced from real-world classrooms; and some traditional teacher certification programs degraded overall teacher quality.[10]

The central role accorded to teachers at the turn of the century marked a significant shift from 1965, when policymakers believed financial resources were at the core of educational achievement.

None of these proposed reforms substantively challenged the core framing of teaching as an output. Teaching could still be controlled through licensing, and once licensed, teachers would still act as a policy input, free to adapt strategies as they saw fit. In a 1998 interview, Leila Vickers, dean of the school of education at North Carolina A & T State University, argued that teacher preparation programs should revel in this uncertainty:

> I would like to see the state make policies that allow us teacher educators to develop a "pluralistic" approach to assessment. If we are going to have professional teachers model effective behaviors, to set high expectations for all students, we have to begin in the preparation program to say, "All individuals do not learn in the same way." . . . We

> must employ strategies that use the modalities through which they learn.[11]

Despite maintaining an input view of teaching, the 1998 reauthorization of the HEA was a fundamental departure in federal education policy. Teaching had been mentioned in a disorganized, catchall Title V since the act's origin, but in 1998 Congress created a new Title II that focused exclusively on teacher quality. For the first time, the federal government would require any college of education that received federal money to publicly report how its teacher preparation programs were assessed, to publish how many of its teaching candidates passed state licensure exams, and to construct their exams to measure "teacher candidate knowledge and skills."[12] This was a compromise, as Rep. Miller's vigorous fight for federal teacher preparation standards ultimately failed.

For the first time, the federal government would require any college of education that received federal money to publicly report how its teacher preparation programs were assessed

Teacher Quality as an Individual Characteristic. Work in Congress on NCLB had already begun when HEA passed, and NCLB's Title II standards for "highly qualified teachers" (HQT) were a direct outgrowth of the teacher preparation debate for HEA. Observers lamented that NCLB appeared to pay little attention to bolstering teacher quality—only $2.8 billion of the Department of Education's $32 billion K–12 appropriation—but this time, Rep. Miller was instrumental in ratcheting up accountability for teacher quality.[13] HEA demanded transparency from *institutions* responsible for preparing teachers but remained silent on how states and school districts were to address the existing inequitable distribution of teacher quality. In contrast, NCLB placed the onus on *teachers* by requiring that every teacher in schools receiving Title I money be "highly qualified" by 2005–2006. Under NCLB, a highly qualified teacher was one who had a bachelor's degree, full state certification, and demonstrated knowledge of his or her subject matter. The subject-matter requirement, drawn from the NCTAF report, was a small step toward requiring *highly effective* teachers, yet NCLB continued to assume that teachers were essentially equal in what they would bring to the classroom as long as they met HQT. Indeed, Stanford University professor Linda Darling-Hammond argued in the fall of 2001 that certification was the "one lever" that states had to improve teacher quality.[14]

NCLB's dramatic approach to assessing student achievement—requiring all students to meet a fixed standard on a fixed timeline—ran interference for policymakers who sought to reframe teacher quality in the same way. If schools and districts could be held accountable for their students' academic performance, then colleges of education should be held equally accountable for their graduates' performance. US Secretary of Education Rod Paige was an eager participant in hammering teacher preparation programs. Using data from HEA-required reports in 2002, he showed that only *one* state, Virginia, set a passing score even close to the national average for individual certification exams. (That is, in most states, an aspiring teacher could meet the requirement by scoring significantly below the fiftieth percentile.) The rest were below. The Education Trust thought he was too charitable; a report from the group vigorously attacked states for providing "inconsistent, incomplete, and utterly incomprehensible" data.[15] With a showing this bad, Paige argued, why should teacher candidates attend colleges of education at all? He suggested instead that neither teachers nor colleges of education should be trusted simply on the basis of their credentials or accreditation. Predictably, many schools of education and the NEA blasted the Department of Education's report. But the shift in framing toward measurement was undeniable, and the political tide had begun to shift.

Although NCLB's Title II could have been used to push merit pay, targeted professional development, or new rules for teacher tenure, Bush administration officials were hesitant to push for teacher effectiveness measures beyond those geared for colleges of education. First, research on increasing teacher effectiveness was thin. As late as 2005, authors of an American Educational Research Association report lamented that they could find no research linking teacher accreditation programs to student learning. Second—and more important politically—Title II funds were seen as the "consummate pot of local-control money," as later characterized by Secretary Margaret

Spellings.[16] Administration officials did not want to further antagonize the bipartisan chorus of state and local officials who had already decried NCLB for violating one of the last state-centric policy areas.

As late as 2005, authors of an American Educational Research Association report lamented that they could find no research linking teacher accreditation programs to student learning.

Nevertheless, Title II did unintentionally abet advocates of assessing the outputs of teaching. As states drafted HQT plans, it became apparent that they were circumventing the spirit of HQT if not the letter itself. Lawmakers knew that the 1998 HEA would affect only new teachers, so any significant improvements in the quality and distribution of the 2.9 million–strong teacher corps would be years away.[17] Eager to provide quicker improvements, lawmakers required all teachers to meet NCLB's HQT definition, but states could use an alternative High, Objective, Uniform State Standard of Evaluation (HOUSSE) to help existing teachers satisfy HQT.

As it turned out, HOUSSE proved to be neither uniform nor objective in many states. In 2004, New York's HOUSSE allowed a teacher to provide evidence of *either* five graduate courses *or* the combination of supervising a student teacher and a bachelor's degree in education. Georgia equated two professional development conferences with a doctoral degree. At the other end of the spectrum, Illinois required twenty-four credit hours of coursework in a teacher's content area.[18] In 2005, Wisconsin claimed that 99.5 percent of its core classes were taught by teachers who met HQT; Massachusetts claimed 93 percent, Alabama 81.8 percent, and California 74 percent.[19] In 2007, data from the states showed that HOUSSE still left low-income, nonwhite students with less experienced teachers: 13 percent of highly qualified teachers in schools with more than 75 percent nonwhite students had fewer than three years' experience. For schools with less than 25 percent nonwhite students, the figure was only 5 percent. High-poverty schools showed a similar disparity.[20]

These disparities were exacerbated by states' slow progress toward the law's 100 percent benchmark for 2005–2006. To the chagrin of Rep. Miller and Sen. Ted Kennedy (D-MA), Secretary Spellings relented to pressure from teachers' groups and state departments of education in early 2004 and weakened HQT.[21] Now, teachers would only have to demonstrate "broad-field" competence (e.g., "science") rather than a subject-specific test (e.g., physics). Also, states could design a single procedure to test for competence rather than requiring multiple assessments. By the official deadline, some forty states had not met the requirement, and political reality forced the Department of Education to grant an additional year to those states making a "good-faith effort."[22]

Spellings recognized the tremendous loophole that HOUSSE had created and the incongruence of holding students accountable for performance while holding teachers accountable only for their qualifications. In May 2006, the Department of Education announced that it would no longer allow states to use HOUSSE except in narrowly defined situations (rural teachers of multiple subjects in high school, special education teachers of multiple subjects, and limited-term foreign teachers), with the support of Sen. Michael Castle (R-DE) and Miller. This time, Miller complimented Spellings for taking HQT "seriously."[23] But both the Council of Chief State School Officers and the NEA argued that Spellings had no legal authority to override the language of NCLB, and by fall 2006, the department partially retreated from its plans. In a letter to chief state school officers, Spellings chastised the states for permitting teachers to use HOUSSE's "substantially less rigorous" standards to demonstrate competency, but she revised the department's earlier push by noting that it would seek to end HOUSSE in the reauthorized bill rather than through the rule-making process.[24]

Teacher Quality as a Bipartisan Enterprise. A draft bill to reauthorize NCLB in 2007 continued to take small steps away from a focus on qualification toward a focus on quality. Similar to the 1998 HEA provisions, the discussion draft of Title II circulated by Reps. Miller and Buck McKeon (R-CA) would have required states to report average teacher tenure and the number of first-year teachers, as well as disparities between high- and low-poverty schools.[25]

While these and other attempts have died in Congress, President Barack Obama's Department of Education

has completed the federal transition from the conception of teaching as a credential-based policy input to an evaluated policy output. Instead of overturning the Bush-era accountability focus as some of Obama's supporters hoped, Secretary Arne Duncan amplified it. First, despite strides made by colleges of education over the previous decade, Duncan still called teacher preparation programs in 2010 the "Bermuda Triangle of higher education . . . no one knows which students are succeeding as teachers, which are struggling, and what training was useful or not."[26] Unlike Secretary Paige, who channeled similar criticisms into a charge for alternative certification programs, Secretary Duncan pushed states to link student performance data to individual teachers and principals in the department's *Blueprint for Reform*.[27] His views appeared in the 2009 stimulus bill in the form of requirements for states to report teachers' scores on certification tests upon entering teaching.[28] Second, the Obama administration put teacher effectiveness at the center of educational equity in its goals for a reauthorized ESEA. Although the administration's main push, Race to the Top (RTT), was touted as a departure from the punitive policies of NCLB, fully 28 percent of points in the RTT rubric were earmarked to ferret out effective teachers and principals. But the actual percentage was much higher because the rubric included verbiage in other sections about identifying "effective" teacher preparation programs, "effective" instructional strategies, and "effective" schools identified by student-level data. By one calculation, 53 percent of the rubric was tied to data-driven, "effective" education.[29] Teachers' performance output would be monitored throughout their careers.

Although some Democrats had been supporters of measuring teacher quality back in 2001, the renewed push to measure teacher effectiveness from the Obama administration seriously undercut teachers' unions' political firewall against the evaluation of teachers' individual performance. In 2006, the NEA adopted an ambitious plan to overhaul NCLB after it failed to check the legislation in court. The NEA recognized that it was politically "lonely," in the words of an executive committee member, and argued that teachers had been left out in writing the original bill.[30] Its proposal explicitly delinked teachers from student performance because "teaching is not an individual, isolated profession [but] . . . dependent on the entire network of teaching professionals."[31] It also suggested loosening the definition of "highly qualified" to include an academic minor in a subject and permitting one state's definition of "highly qualified" to meet every other state's definition. Yet, members of Congress charged with revising the law were hesitant to give the NEA a prime place at the table—both Republicans and Democrats on the committees classed the NEA as one of a "wide array" of stakeholders.[32]

Yet, by 2010, it became apparent to the leadership of the teachers' unions that the federal government was not going to back down. In addition to zeroing in on teacher effectiveness, both President Obama and Secretary Duncan supported charter schools, school reconstitution, and versions of merit pay, all highly questionable from the unions' perspective. This was not unnoticed. In a major shift, both the AFT and NEA acknowledged that teacher effectiveness could be measured. AFT president Randi Weingarten remarked: "Deliberately or not, President Obama, whom I supported, has shifted the focus from resources and innovation and collaboration to blaming it all on dedicated teachers."[33] Yet she sought to portray the AFT as open to measuring teacher effectiveness. In a speech to the National Press Club in January 2010, Weingarten criticized standardized testing but agreed that for evaluating teachers, "Student test scores based on valid and reliable assessment should *also* be considered—*not* by comparing the scores of last year's students with the scores of this year's students, but by assessing whether a teacher's students show real growth while in his classroom."[34] Although Weingarten indicated that teacher performance measures should also consider classroom observation, self-evaluations, and portfolio reviews, the admission of test scores to the list underscored a dramatic change from a decade earlier. For the AFT, teachers were individuals with distinct strengths and weaknesses.

Yet the AFT has long been more open to experimenting with education reform, so perhaps this change was less striking. The NEA faced a more difficult fight. At its July 2011 national convention, the union endorsed Obama for reelection, but its delegates were "appalled" by Secretary Duncan and the administration's failure "to respect and honor the professionalism of teachers," according to an adopted resolution.[35] NEA president Dennis Van Roekel recognized that for the NEA to "get into that arena," it would have to concede that student performance was part of teacher effectiveness.[36] Though significantly weakened from an initial draft, delegates approved language acknowledging that student exams could be used to evaluate teacher effectiveness if they were "developmentally appropriate, scientifically valid, and reliable for the purpose of measuring both student learning and a teacher's performance."[37] Whether any exam would ever meet the NEA's standard is uncertain,

but the organization no longer wanted to be kept outside the bargaining room when both parties, major think tanks, and many policy researchers suggested that teacher effectiveness was tightly linked to learning.

Whatever the cause, once the national teachers' unions admitted that teachers could be evaluated independently of their credentials, it was clear that teaching could no longer be undergirded with trust alone. Policymakers across the spectrum demanded measurement. In 2012, policymakers—and the unions—have moved to draft state accountability systems to include teacher performance. Under advice from the Department of Education, many states intend to evaluate teachers with up to half of their measure coming from student test scores.[38] When the Obama administration granted eleven states waivers from having to meet Adequate Yearly Progress (AYP) toward 100 percent student proficiency in early 2012, it was in return for bolstering teacher effectiveness measures and increasing the number of students potentially covered by assessment systems.[39] Although the push to evaluate the outputs of teaching did not alter existing law or even dominant practices, this new conception gave reform-minded policymakers political cover for substantive changes to the relationship among teachers, students, and schools.

New Research, New Technology, New Frame

The reimaging of teaching took a decade to percolate through the back rooms of major stakeholders in American education policy, but it did not circulate simply because it was trendy, politically popular, or even particularly well thought out. Indeed, in 2001 teacher quality measures garnered very little attention in the initial debates over NCLB, were subject to strong pushback from teachers' organizations, and generated a host of critics both for and against.

Instead, the secular change from teacher-as-input to teaching-as-output was driven by two major external developments. First was the emergence of credible research in the late 1990s and early 2000s that challenged long-held assumptions about the relationship of licensure to quality. Second was that NCLB's mandatory testing requirements created a treasure-trove of data on the individual student level that, though imperfect in many ways, created research and political pressure to take the next step and link student performance to teacher performance. These developments occurred in parallel but interacted in ways policymakers had not originally envisioned. New policies create new politics.

This new conception gave reform-minded policymakers political cover for substantive changes to the relationship among teachers, students and schools

The motivation for new research was driven by a combination of the Clinton-era "reinventing government" and educational excellence movements in the 1990s. Under the guidance of Vice President Al Gore, the Government Performance and Results Act of 1993 (GPRA) sought to "systematically hold . . . agencies accountable for achieving program results" because "waste and inefficiency in Federal programs undermine the confidence of the American people in the Government."[40] Although American education was not the focus of Gore's efforts, GPRA was a manifestation of the results-oriented policy space in which trust-based teaching could be undermined. As NAEP results weakened policymakers' confidence in ESEA's original theory, seeking to hold schools "accountable for results" was a small step. The educational excellence movement provided the appropriate "results": high (or at least improving) test scores. As such, a handful of education reformers and scholars in the early 1990s suggested that school choice and market-like incentives could revolutionize the quality of American education, where "quality" meant "academic performance."[41] These reforms suggested that outsiders, whether state panels, parents, students, or voters, could evaluate school and teacher quality better than practitioners themselves. At the time, there was little research to support or oppose the claims, and most arguments were thought experiments or analogies to private schools or desegregation. In 1996, Eric Hanushek characterized the situation as frustrating: "We know neither what forms of incentive systems are best nor what results we might expect . . . evaluation is central" but largely unavailable.[42] The call for evaluation of American education found supporters in Congress, and the 1994 IASA expanded NAEP to measure students' academic progress. This data

supplemented older data that had been repeatedly mined for hints about teacher effectiveness with little satisfaction. NAEP would provide contemporary, if cross-sectional, data.

The Educational Testing Service report signaled that reformers could target *specific teaching practices* to improve learning instead of trying to attract better teacher candidates with higher pay.

This newly available data became the foundation of studies questioning the long-standard claim by teachers' unions that certified and highly paid teachers benefited students. In 2000, the Educational Testing Service released a study by Harold Wenglinsky suggesting that teachers' content knowledge and specific pedagogical practices (such as frequent quizzes and hands-on instruction) had a significant effect on students' NAEP math and science scores. Wenglinsky found little relationship between those scores and either teacher pay, retention, or general professional development.[43] Although the study's conclusions were weakened by the limitations of 1996 NAEP data (they were cross-sectional rather than longitudinal; they had no reliable information about students' socioeconomic status; they had no school-level information; and nonscore information was self-reported), the report signaled that reformers could target *specific teaching practices* to improve learning instead of trying to attract better teacher candidates with higher pay. Further, the study found that many traditional professional development programs—run by colleges of education and teachers' organizations—had little or even negative effects on math and science scores.

Concerns about the variability of teaching moved beyond the closed world of education policy experts. The private-sector Teaching Commission, headed by former IBM CEO Lou V. Gerstner, recommended substantial changes to teacher preparation to enhance the quality of new teachers in 2004. Although the panel's recommendations were not unique, the high-profile nature of the commission suggested that the private sector would use data to analyze effectiveness. By middecade, outspoken Democrats in Congress, independent researchers, the US Department of Education, and the business sector had all weighed existing teacher quality policy and found it wanting. With that, efforts to replace trust with measurement became a political juggernaut.

These reports questioned whether the black box of teacher quality was actually opaque, but they only foreshadowed more critical reports built on the avalanche of data that NCLB generated, the second development leading to the erosion of trust in credentials. NCLB's AYP data were not intended to measure teachers, a fact that critics of the law are not shy to highlight, but the data were more detailed, more uniform, and more widely available than any previous educational achievement data. The temptation became too much to resist. The National Council on Teacher Quality produced a scathing preliminary report in 2004 using state-provided data, but its conclusions were confirmed by a 2007 federal report using NCLB data, *Teacher Quality under NCLB: An Interim Report*. One of the lead researchers accused states of undermining teacher quality, saying, "The high compliance rate suggests there were states that set the bar low, and, in a way, grandfathered in a lot of teachers."[44] Both reports criticized HOUSSE, as had many others, but they delivered a damning blow because they used *federal* data. A decade earlier, researchers would have had to design a research protocol and synthesize disparate data. HEA and NCLB consolidated the data, bringing the policy issue into the sunlight. If HQT was not the strong medicine that some of its initial boosters had hoped to give to slow-moving states and colleges of education, policymakers could have written its failure off as another case study in the difficulty of reforming loosely coupled systems like schools, or why federal policy flounders in the face of limited state capacity.[45] HQT did not itself contribute to reframing teaching, but the proliferation of student assessment in the wake of NCLB created new opportunities for research to link teachers and students. That is, NCLB had not improved teacher quality in the way it had intended. HQT was meant to bolster trust in teachers' ability by requiring *more* credentials, but HQT only accelerated the movement to measure teachers using data collected from *students*.

Middecade, a series of reports indicated that teachers' credentials had little to no value in predicting student achievement. Research completed in 2006 by Thomas Kane and his coauthors showed that math performance by students in classrooms with alternatively certified teachers

was indistinguishable from that in classrooms with traditionally certified teachers in the Los Angeles Unified School District (Kane's research is particularly notable because AFT president Randi Weingarten singled him out as a trustworthy researcher in a 2010 address).[46] Hanushek and his colleagues found similar noneffects of certification and master's degree requirements in Texas.[47] Likewise, in 2007, Robert Pianta and his coauthors found wide variation in student performance in a multistate study of middle-class students—so wide that the findings "did not appear congruent with the high performance standards expected for students or for teachers as described by most state teacher certification and licensure documents."[48] Further, outside the academic world, NCLB's student data continued to show classroom-level gaps between white and nonwhite student groups—even in states such as Wisconsin, Montana, Oklahoma, and Connecticut that reported upwards of 98 percent highly qualified teachers. By the end of the decade, policymakers had an indisputable research base showing that teacher licensure itself, used to justify treating teachers as equivalent inputs, was unrelated to teachers' effectiveness in the classroom. The "one lever" Darling-Hammond said states had over the teaching workforce turned out to be futile tinkering toward utopia.

As research confirmed the political reframing of teaching, critics of NCLB's absolute benchmark for AYP sought to replace it with a student- or classroom-based value-added measure. Value-added measures were something of a halfway covenant between the US Department of Education and practitioners, and they have generated lukewarm interest in teachers' union affiliates.[49] Value-added models gave the Department of Education a political "out" when it became clear that a large number of schools and districts would be identified as "in need of improvement" using the fixed-target AYP. In 2005, Secretary Spellings allowed nine states to pilot value-added models of academic growth. A study that year by the National Association of State Boards of Education praised the decision, saying that the change "move[d] the discussion about teacher quality to where it belongs: centered on increasing student learning as the primary goal of teaching."[50] The Obama administration made value-added models a de facto policy in 2009 by requiring RTT applicants to use them to measure student achievement, a move that Secretary Duncan later said would "[hold] us all accountable for the quality of education we provide to every single student in America."[51] For teachers' union affiliates, value-added approaches have been embraced as a way to mentor new teachers and identify students needing extra help.[52] In a simultaneous effort beginning in 2005, the Data Quality Campaign began to push states to create data systems to track individual student scores to make value-added evaluation feasible.

Yet for both the administration and unions, the shift to value-added models middecade acknowledged the changing framing of teacher quality. The fixed AYP standard implied that students should meet the standard *regardless* of their teachers, but value-added standards suggest that different students may respond differently to different teachers. Teachers are then not interchangeable but individuals who differ in their effectiveness.

> The proliferation of student assessment in the wake of NCLB created new opportunities for research to link teachers and students.

Elixirs and Effectiveness: Finding Quality Teachers

The reframing of teaching in the 2000s opened a policy window that had been closed a generation before by ESEA's intensive focus on school resources. Policymakers do not talk about teacher quality in 2012 the same way they did in 2001. Today, the debate is better informed by research, better linked to student learning, and thoroughly national. For policymakers, the central takeaway message from the decade is that the next generation of teacher quality must be tightly *targeted* to improving results. No serious policymaker now suggests that requiring more hours of general professional development will improve student outcomes, and many of them doubt that traditional teacher certification is a marker of quality. Sensing the opportunity for a dramatic overhaul of teaching, policy entrepreneurs swept into the legislative marketplace with elixirs to remedy American education through teaching reform. Yet California governor Jerry Brown singled out "experts and academics and foundation consultants" for their exuberance in January 2012, saying, "it is salutary and even edifying that so much interest is shown in the next generation, [but] in a state with six million students,

300,000 teachers, deep economic divisions, and a hundred different languages, some humility is called for."[53] As with California, so with the nation. None of these medicines can cure the patient in the way their boosters suggest, but each represents a small, positive step toward understanding teachers as individuals rather than as an undifferentiated class of employees. In broad strokes, policy entrepreneurs propose to remedy low teacher quality by addressing teacher preparation, teacher recruitment, or teacher tenure.

Teacher Preparation. One elixir, and the least disruptive to the status quo, suggests that overhauling teacher education is the key to closing the teacher quality gap. Proponents acknowledge that colleges of education produce teachers of widely varying skill but prefer to emphasize improving "skills" of teachers rather than rating them on student results. According to Darling-Hammond, "Developing more-skillful teaching is a sine qua non for attaining higher and more equitable achievement for students in the United States."[54] Darling-Hammond and others suggest that colleges of education develop residency-like programs to minimize lower-quality teaching that is characteristic of first- and second-year teachers. The residency model got a boost from Secretary Duncan in 2012 when he suggested that states could require evidence of effective teaching *before* granting a license.[55] One critic of the approach conceded that "no one is even against high-stakes evaluation" anymore.[56] Others have suggested increasing content-specific courses. Along these lines, the National Council for Accreditation of Teacher Education is boosting standards for teacher education programs. Colleges of education have been slow to report on their actual practices, and officials are worried about how a residency-like model would affect the cost and course structure of college. Still, these institutions know they remain under the microscope for improving student performance.

A second elixir some propose is to require existing teachers to gain advanced credentialing, especially National Board certification. Unlike traditional teacher preparation programs, National Board certification is intensive and content-based, and it requires reflection on pedagogy. It also has a rigorous exam (the pass rate for the Praxis II certification exam in a given year is approximately 90 percent, while the first-time National Board pass rate is near 50 percent).[57] Districts have used National Board certification as a distinction for teacher quality. Between its inception in 1993 and 2011, 97,281 teachers had become so certified, although 48 percent of these are in Florida, North Carolina, South Carolina, and Washington, states that give financial bonuses for certification.[58] As a boost to teacher effectiveness, this reform shows mixed results. It does appear that National Board certification is a fair indicator of quality teaching. Two studies of North Carolina teachers with board certification, published in 2007 and 2010, found statistically significant improvements in student test scores versus non–board certified teachers, although not necessarily uniformly across the curriculum. In particular, board certification seemed a better marker in lower grades and for low-income students.[59] Yet as a way to *improve* existing teachers' effectiveness, National Board certification appears to have little to offer. One of the studies just mentioned found that National Board teachers were no more effective after receiving the certification than before, and a 2009 study of Florida teachers found no consistent benefits for any grade or student exam.[60]

By the end of the decade, policymakers had an indisputable research base showing that teacher licensure itself was unrelated to teachers' effectiveness in the classroom.

Teacher Recruitment. Other policy entrepreneurs suggest that the root problem of low teacher effectiveness is not a problem of preparation but of supply. Traditionally certified teachers tend to have lower academic scores in college than graduates as a whole, and some academic areas are routinely difficult to fill, especially math, engineering, and natural sciences.[61] The solution, according to these entrepreneurs, is to make teaching more attractive to potentially excellent teachers who forgo traditional teacher preparation because the opportunity cost of an education degree is too high.

Alternative certification would bring in new, high-quality teachers with real-world experience or a passion for teaching. Because research suggests that general pedagogical courses or professional development have little impact on teaching effectiveness, alternative certification could bolster the ranks of content-expert teachers or high-performing college graduates who would not otherwise have considered teaching—especially in low-performing, low-income

districts. The most highly publicized alternative to traditional teacher preparation is Teach For America (TFA), a nonprofit organization formed in 1990 to address teacher shortages in low-income schools. TFA's "corps members" make an explicit two-year commitment to a designated school. Surveys of TFA teachers suggest that about 60 percent of them remain in teaching for more than two years and 28 percent over five.[62] Those numbers, however, are not for teachers remaining in their original school, and other studies of TFA suggest very high turnover rates. Using data from 2003 to 2008, for example, 79 percent of TFA teachers had left the New York City public schools after four years compared with 31 percent of new but traditionally certified teachers.[63] As a solution to teacher shortages, then, TFA does not appear promising. Turnover is problematic because it creates a school environment with little community, increases personnel costs as districts seek to replace teachers, and potentially weakens student achievement as new teachers with little experience replace departing ones.

Alternative certification does not necessarily assume that teachers will enter a lifetime of service. Consistent with the new framing of quality, alternatively certified teachers are expected to be *effective*—as measured by outside reviewers. TFA teachers are unquestionably highly qualified in terms of degrees and college entrance exam scores; the same New York study found that high school TFA math teachers scored 110 points higher on SAT math than traditionally certified high school math teachers.[64] Whether that academic prowess translates into classroom success is an elusive answer, however. Louisiana TFA teachers appear to consistently out-perform similarly experienced non-TFA teachers and hold their own with more experienced teachers, but Houston teachers produce comparable or worse results than non-TFA teachers.[65] New York TFA teachers produce results slightly better than comparably experienced teachers in some cases and worse in others—but generally better than other alternatively certified teachers.[66] And therein lies the rub: unless schools and districts can overcome high turnover among alternatively certified teachers, alternative certification is unlikely to transform American education in the neediest of districts.

Teacher Tenure. A last group of policy entrepreneurs argue that student performance should directly reshape the teacher workforce by calling into question the assumptions of traditional teacher tenure. And, unlike either teacher training or teacher recruitment, workforce policies seek to overhaul the existing teacher corps in pursuit of effective teaching outputs. Critics of traditional teacher tenure argue that treating similarly qualified teachers as equal denies voters the right to hold teachers accountable for their performance—and by extension, their use of taxes. There are two basic approaches to reforming tenure. The first, differential pay, severs the link between years of experience and pay. The second abolishes tenure entirely and ties teachers' employment to demonstrated effective teaching.

Proponents of differential pay, also known as merit pay, assume that teachers respond to financial incentives. The assumption itself is not controversial: teachers' unions have argued since their inception that higher pay will increase the quality of teachers and that the level of teacher pay discourages potential teachers. One researcher estimated that a 10 percent increase in potential non-teaching wages reduced the likelihood that the top quarter of female college students would go into teaching by 6.4 percent. The same 10 percent increase only reduced teaching's attractiveness by 3.7 percent for women in the twenty-fifth to fiftieth percentile.[67] Yet the dominant approach to teacher pay works against using pay as an incentive. Most districts pay teachers based on years of experience (step) and their level of education (lane). Experience and education would be appropriate measures of quality if policymakers still understood teaching as a black box because both measures reward teachers for dedication to their careers; that may in fact help teacher supply. But research has amply demonstrated that educational credentials have no correlation with teacher quality, and student achievement is most sensitive only to the first few years of a teacher's career, thereby undermining the logic of rewarding teachers solely for their credentials and experience.

Merit pay addresses this disconnect. Instead of rewarding teachers for years of service, merit pay offers higher salaries or bonuses based on demonstrated student performance. Incentives like this have suggestive results. A first evaluation of Denver's ProComp program found that students with new teachers had consistently higher scores than those with Denver teachers overall. High-poverty schools were also able to retain more of their new teachers, although they still had higher turnover than suburban schools.[68] A competing study in Nashville, however, found that simply offering bonuses had no significant effect on teacher performance, even bonuses of $15,000.[69] Incidentally, the Denver study found similar results for existing teachers who opted into the merit system. The results of these studies suggest that merit pay can, in fact, improve *recruitment* of higher-quality teachers, but offering bonuses to existing teachers by itself will do little to

refocus teachers' existing practice.

Most dramatically, a second group of policy entrepreneurs suggests that students' short time horizon in school dictates that school districts should fire or reassign poorly performing teachers now. Although scholars continue to struggle with identifying "good" teachers, principals and other teachers have little problem reliably identifying poor teachers in their schools. This approach is most visibly associated with Michelle Rhee's IMPACT teacher evaluation system in the District of Columbia Public Schools, but less publicized variations have been tried in other districts, including the Houston Independent School District and the Hamilton County (Tennessee) school district.[70] Initial evaluations of laying off the least effective teachers indicate not only that student achievement would improve but also that districts would ultimately lay off fewer personnel than under a seniority-based layoff system.[71] In the short term, firing the worst teachers may produce strong, positive results, but the approach still leaves unaddressed why some teachers enter the system ill prepared. Hiring teachers only to fire them shortly thereafter is a destabilizing, expensive process for schools that, in the long term, could undermine short-term gains.

Conclusion

When James Coleman and his coauthors released EEO in 1966, they despaired about the ability of schooling to rectify pervasive inequity in American education. Yet, since then, researchers have confirmed that teaching is a central driver of student achievement. In the last fifteen years, policymakers have moved from trusting teachers' credentials to measuring teaching to verify success in the classroom. New data and better research have opened a policy window for creative and meaningful changes to teacher accountability. All of the top-line reforms are plagued by uncertainty—When can good teachers be identified? When is experience more important than training? How much should student achievement contribute to a teacher's evaluation?—but all of the questions address the core enterprise of schools: student learning.

The change from trust to measurement occurred not so much from a failure of teachers but from a recognition that teacher education programs were cookie-cutter solutions for infinitely variable education problems. Identical credentials granted at the beginning of teachers' careers cannot adequately predict how well teachers will fare in widely and predictably different classrooms. Teachers, too, are not uniform individuals, a fact that even the stepped credentialing popular in the 1990s hid. Measurement allows policymakers to remedy poor teacher performance for students' sakes and to place teachers in classrooms where they can expect to thrive.

For policymakers, the central take-away message from the decade is that the next generation of teacher quality must be tightly *targeted* to improving results.

Nevertheless, the small steps of the 2000s toward the next generation of teacher quality face technological and political difficulties going forward. Technologically, remote instruction has challenged the connection between teaching and learning. Unlike the cable television instruction promoted by some states in the 1980s, contemporary Internet-based education can provide self-paced learning for students who are comfortable with—and expect—interactions with computers. As distance learning expands in K–12 education, the link between student performance and teacher quality will become increasingly tenuous. For students, the advantage of remote education is that it can be tailored much better to their own education level and needs because students interact with the curriculum independently and individually. The quality of online, "content" teachers could be measured minutely by tracking the time it takes remote teachers to respond to questions, the number of attempts students take on any particular assignment, or even the number of interactions a teacher has with a particular student. On-site teachers will provide less primary instruction and more task management for students; the qualities of these teachers will differ from those needed by "content" teachers. Assessment of student learning will be harder to appropriately assign.

Politically, the bipartisan consensus around teacher quality may be challenged by renewed concerns about American federalism. Although state governments eagerly adopted the Common Core Standards Initiative, their eagerness sprang at least as much from a desire to defuse federal efforts as from a desire to create uniform standards. The Obama administration has not been shy about using federal power to trump state policymakers in

immigration, health care, and education policy. To the extent Republicans (and state-level Democrats) characterize teacher quality initiatives as intrusive federal policy, the next generation of teacher quality may become a political albatross for its supporters in the states.

Despite these challenges, the next generation of teacher quality will continue to positively reshape American education. In a world that demands economic competitiveness both locally and internationally, the presumption that all teachers are equal in talent thwarts the American dream of equal educational opportunity. For teachers, the de facto loss of trust in credentials suggests that they must place greater emphasis on student learning. This does not need to be a negative thing. Although Title II does not seem to have dramatically boosted teacher quality in the way its early supporters hoped, schools and school districts have had to think about teacher effectiveness in a new way. Teachers may endure greater scrutiny in their classrooms, but that may also mean that they will receive better, individual, targeted pedagogical training, consistent and meaningful evaluation of their practice, and better understanding of the use of evaluation data. If so, they will benefit from these changes.

For students, especially low-performing students, policymakers' focus on the results of teaching has improved their chances of receiving a better education. Untargeted spending is politically easy and often explicitly supported by teachers' unions, but it perpetuates inequality as teachers are rewarded for *not* taking challenges as much as they are *for* taking challenges. Boosting teacher quality through targeted reforms will not cure American education once for all, but it will reward teachers who take challenges and succeed at them—because their students will, too.

Notes

1. James S. Coleman, Ernest Q. Campbell, Carol J. Hobson, James McPartland, Alexander M. Mood, Frederic D. Weinfeld, and Robert L. York, *Equality of Educational Opportunity, Report OE-38001* (Washington, DC: US Department of Health, Education, and Welfare, Office of Education, 1966), 22.
2. Dan Goldhaber, Dominic Brewer, and Deborah Anderson, "A Three-Way Error Components Analysis of Educational Productivity," *Education Economics* 7, no. 3 (1999): 199–208.
3. For example, Jeff Archer, "State Teacher Policies Tied to Student Results," *Education Week*, January 12, 2000; Jeff Archer, "Teacher-Quality Bill Comes Down to Wire in New Mexico," Education Week, March 21, 2001; and David J. Hoff, "New Law Allows Ky. Districts to Pilot Differentiated Pay," *Education Week*, April 17, 2002.
4. "Licensure" and "certification" are used interchangeably in this essay.
5. For example, Rob Greenwald, Larry V. Hedges, and Richard D. Laine, "The Effect of School Resources on Student Achievement," *Review of Educational Research* 66, no. 3 (Fall 1996): 361–96; and Ronald F. Ferguson, "Paying for Public Education: New Evidence on How and Why Money Matters," *Harvard Journal of Legislation* 28, no. 2 (1991): 465–98.
6. National Commission for Teaching and America's Future, *What Matters Most: Teaching for America's Future* (Woodbridge, VA, 1996), 15.
7. Ibid., 16.
8. Ibid., 32.
9. Stephen Burd, "Liberal Democrat Is Unlikely Foe of Teacher-Education Programs," *Chronicle of Higher Education*, April 24, 1998.
10. Penelope M. Earley, "Finding the Culprit: Federal Policy and Teacher Education," *Educational Policy* 14, no. 1 (January 2000): 25–39.
11. Mary E. Diez, *Changing the Practice of Teacher Education: Standards and Assessment as a Lever for Change* (Washington, DC: American Association of Colleges for Teacher Education, 1998), 34.
12. Public Law 105-244, Section 207.
13. US Department of Education, "Education Department Budget History Table, FY1980–FY2012," August 5, 2011, www2.ed.gov/about/overview/budget/history/edhistory.xls.
14. Jeff Archer, "Foundation Stirs Debate with Report Questioning Research on Licensure," *Education Week*, October 17, 2001.
15. Bess Keller and Michelle Galley, "Paige Uses Report as a Rallying Cry to Fix Teacher Ed.," *Education Week*, June 19, 2002.
16. Stephen Sawchuk, "Spending of Federal Teacher-Quality Funds Questioned," *Education Week*, November 19, 2008.
17. National Center for Education Statistics, *Digest of Education Statistics 2010* (Washington, DC: US Department of Education, April 2011), table 68.
18. Christopher O. Tracy and Kate Walsh, *Necessary and Insufficient: Resisting a Full Measure of Teacher Quality* (Washington, DC: National Council on Teacher Quality, Spring 2004), 5, 9–12.
19. Kevin Carey, *Hot Air: How States Inflate Their Educational Progress under NCLB* (Washington, DC: Education Sector, May 2006), 18.
20. Beatrice F. Birman, Kerstin Carlson Le Floch, Amy Klekotka, Meredith Ludwig, James Taylor, Kirk Walters, Andrew Wayne, Kwang-Suk Yoon, Georges Vernez, Michael Garet, and Jennifer O'Day, *State and Local Implementation of the No Child Left Behind Act; Volume II—Teacher Quality under NCLB: Interim Report* (Santa Monica, CA: RAND, 2007), 41.
21. Erik W. Robelen, "Federal Rules for Teachers Are Relaxed," *Education Week,* March 19, 2004.
22. Bess Keller, "States Given an Extra Year on Teachers: 'Highly Qualified' Delay Has Strings Attached," *Education Week*, November 1, 2005.
23. Bess Keller, "No State Meeting Teacher Provisions of 'No Child' Law," *Education Week*, May 24, 2006.
24. Margaret Spellings, "Letter to Chief State School Officers Regarding Efforts to Ensure That All Core Academic Subjects Are Taught by Highly Qualified and Experienced Teachers, and Encouraging States to Eliminate Use of HOUSSE Procedures," September 5, 2006, www2.ed.gov/policy/elsec/guid/secletter/060905.html.
25. David J. Hoff, "Draft Retains Quality Rules for Teachers," *Education Week,* September 12, 2007.
26. Stephen Sawchuk, "Duncan Cites Shortcomings of Teacher Preparation," *Education Week*, October 23, 2009.
27. US Department of Education, *A Blueprint for Reform* (Washington, DC, March 2010), 14, www2.ed.gov/policy/elsec/leg/blueprint/blueprint.pdf.
28. Stephen Sawchuk, "Stimulus Reflects Push for Teacher Effectiveness," *Education Week,* February 5, 2010.
29. US Department of Education, "Race to the Top Scoring Rubric, Corrected," January 2010, www2.ed.gov/programs/racetothetop/scoringrubric.pdf.
30. Vaishali Honawar, "NEA Opens Campaign to Rewrite Federal Education Law," *Education Week*, July 12, 2006.
31. National Education Association, "ESEA: It's Time for a Change!" July 2006, 24, www.nea.org/assets/docs/HE/TM-NEAPositiveAgendafortheESEAReauthorization.pdf.
32. Honawar, "NEA Opens Campaign."
33. Steven Brill, "The Teachers' Unions' Last Stand," *New York Times Magazine*, May 17, 2010, www.nytimes.com/2010/05/23/magazine/23Brill.html.

34. Randi Weingarten, "A New Path Forward: Four Approaches to Quality Teaching and Better Schools" (speech, Washington, DC, January 12, 2010), http://aft.3cdn.net/227d12e668432ca48e_twm6b90k1.pdf, 4.
35. National Education Association, "New Business Item C," July 4, 2011, www.nea.org/grants/33354.htm.
36. Stephen Sawchuk, "NEA Relaxes Evaluation Policy, Endorses Obama," *Education Week*, July 13, 2011.
37. National Education Association, "New Policy Statement on Teacher Evaluation and Accountability," July 5, 2011, www.nea.org/grants/46326.htm.
38. See Stephanie Banchero, "States Test Education Law," *Wall Street Journal*, July 19, 2011; and Wisconsin Department of Public Instruction, "Accountability Reform Overview," January 27, 2012, 8, http://dpi.wi.gov/esea/pdf/summary.pdf.
39. Some have questioned whether the waivers will encourage states to create a two-track accountability system for students. See Sean Cavanagh and Alyson Klein, "Broad Changes Ahead as NCLB Waivers Roll Out," *Education Week*, February 9, 2012.
40. Office of Management and Budget, "Government Performance and Results Act of 1993, Section 2," www.whitehouse.gov/omb/mgmt-gpra/gplaw2m.
41. Most famously in John Chubb and Terry Moe, *Politics, Markets, and America's Schools* (Washington, DC: Brookings, 1989), chapter 5, but also in Eric Hanushek, "The Productivity Collapse in Schools," in *Developments in School Finance, 1996* (Washington, DC: National Center for Education Statistics, 1996).
42. Eric Hanushek, "Improving School Performance While Controlling Costs," in William J. Fowler, ed., *Developments in School Finance, 1995* (Washington, DC: National Center for Education Statistics, 1996), 118.
43. Harold H. Wenglinsky, *How Teaching Matters: Bringing the Classroom Back into Discussions of Teacher Quality* (Princeton, NJ: Educational Testing Service, 2000).
44. Debra Viadero, "NCLB Teacher Rules Unevenly Enforced, Major Study Finds," *Education Week*, August 29, 2007.
45. See, for example, Paul Manna, *School's In* (Washington, DC: Georgetown University Press, 2006).
46. Weingarten, "A New Path Forward."
47. Robert Gordon, Thomas J. Kane, and Douglas O. Staiger, *Identifying Effective Teachers Using Performance on the Job* (Washington, DC: Brookings, 2006); and Steven G. Rivkin, Eric A. Hanushek, and John F. Kain, "Teachers, Schools, and Academic Achievement," *Econometrica* 73, no. 2 (March 2005): 417–58.
48. Robert C. Pianta, Jay Belsky, Renate Houts, and Fred Morrison, "Opportunities to Learn in America's Elementary Classrooms," *Science* 315, no. 5820 (March 30, 2007): 1795–96.
49. The AFT has sponsored workshops on how value-added models work and when they might be useful, although the union's official position is that such models should never be used for pay or employment decisions.
50. Lynn Olson, "'Value Added' Models for Gauging Gains Called Promising," *Education Week*, October 25, 2005.
51. *Winning the Future with Education: Responsibility, Reform, and Results, Before House Comm. on Education and the Workforce*, 112th Cong. (March 9, 2011) (statement of Arne Duncan, US Secretary of Education), www.ed.gov/news/speeches/winning-future-education-responsibility-reform-and-results.
52. Daniel Koretz, "A Measured Approach," *American Educator* (Fall 2008): 18–39; and Matthew Di Carlo, "What Is Value-Added?" (presentation, 2001 TEACH Conference, Washington, DC, July 2011), www.aft.org/pdfs/teachers/teach11materials/m215_istherevaluepres.pdf.
53. Jerry Brown, "Text of Gov. Jerry Brown's State of the State Address," *Los Angeles Times,* January 18, 2012.
54. Linda Darling-Hammond, "A Marshall Plan for Teaching," *Education Week*, January 10, 2007.
55. Stephanie Banchero, "Plan Offers $5 Billion to Improve Teaching," *Wall Street Journal*, February 15, 2012.
56. Stephen Sawchuk, "Teacher-Quality Group to Revamp Education School Review," *Education Week*, February 23, 2011.
57. Dan Goldhaber and Emily Anthony, "Can Teacher Quality Be Effectively Assessed? National Board Certification as a Signal of Effective Teaching," *Review of Economics and Statistics* 89, no. 1 (2007): 134–50.
58. National Board for Professional Teaching Standards, "NBCTs by State," www.nbpts.org/resources/nbct_directory/nbcts_by_state.
59. Goldhaber and Anthony, "Can Teacher Quality Be Effectively Assessed?"; and Charles T. Clotfelter, Helen F. Ladd, and Jacob L. Vigdor, "Teacher Credentials and Student Achievement in High School," *Journal of Human Resources* 45, no. 3 (2010): 655–81.
60. Goldhaber and Anthony, "Can Teacher Quality Be Effectively Assessed?"; Douglas N. Harris and Tim R. Sass, "The Effects of NBPTS-Certified Teachers on Student Achievement," *Journal of Policy Analysis and Management* 28, no. 1 (2009): 55–80.
61. Michael Podgursky, Ryan Monroe, and Donald Watson, "The Academic Quality of Public School Teachers: An Analysis of Entry and Exit Behavior," *Economics of Education Review* 23 (2004): 507–18.
62. Morgaen L. Donaldson and Susan Moore Johnson, "TFA Teachers: How Long Do They Teach? Why Do They Leave?" *Phi Delta Kappan*, October 4, 2011.
63. Donald Boyd, Pam Grossman, Karen Hammerness, Hamilton Lankford, Susanna Loeb, Mathew Ronfeldt, and James Wyckoff, "Recruiting Effective Math Teachers: How Do Math Immersion Teachers Compare? Evidence from New York City," NBER Working Paper 16017, May 2010, 40, www.nber.org/papers/w16017.
64. Boyd et al., "Recruiting Effective Math Teachers," 27.
65. George H. Noell and Kristin A. Gansle, "Teach for America Teachers' Contribution to Student Achievement in Louisiana in Grades 4–9: 2004–2005 to 2006–2007," Technical Report (Baton Rouge, LA: Louisiana State University, October 27, 2009); Margaret Raymond, Stephen H. Fletcher, and Javier Luque, *Teach for America: An Evaluation of Teacher Differences and Student Outcomes in Houston, Texas* (Stanford, CA: Hoover Institution, Center for Research on Education Outcomes, 2001); and Linda Darling-Hammond, Deborah J. Holtzman, Su Jin Gatlin, and Julian Vasquez Heilig, "Does Teacher Preparation Matter? Evidence about Teacher Certification, Teach for America, and Teacher Effectiveness," *Education Policy Analysis Archives* 13, no. 42 (October 2005), http://epaa.asu.edu/epaa/v13n42.
66. Boyd et al., "Recruiting Effective Math Teachers."
67. Marigee P. Bacolod, "Do Alternative Opportunities Matter? The Role of Female Labor Markets in the Decline of Teacher Quality," *Review of Economics and Statistics* 84, no. 4 (November 2007): 748.
68. Edward W. Wiley, Eleanor R. Spindler, and Amy N. Subert, *Denver ProComp: An Outcomes Evaluation of Denver's Alternative Teacher Compensation System* (Boulder, CO: School of Education, University of Colorado at Boulder, 2010).
69. Matthew G. Springer, Dale Ballou, Laura Hamilton, Vi-Nhuan Le, J. R. Lockwood, Daniel F. McCaffrey, Matthew Pepper, and Brian M. Stecher, *Teacher Pay for Performance: Experimental Evidence from the Project on Incentives in Teaching* (Nashville, TN: National Center on Performance Incentives, Vanderbilt University, 2010).
70. Stephen Sawchuk, "Spending of Federal Teacher-Quality Funds Questioned," *Education Week*, November 19, 2008.
71. Dan Goldhaber and Roddy Theobald, *Assessing the Determinants and Implications of Teacher Layoffs* (Washington, DC: National Center for Analysis of Longitudinal Data in Education Research, Urban Institute, 2010).