



So Long, Lake Wobegon?

Using Teacher Evaluation to Raise Teacher Quality

Morgaen L. Donaldson June 2009



So Long, Lake Wobegon?

Using Teacher Evaluation to Raise Teacher Quality

Morgaen L. Donaldson June 2009

Executive summary

In recent months, ideas about how to improve teacher evaluation have gained prominence nationwide. In April, 2009, Secretary of Education Arne Duncan proposed that districts report the percentage of teachers rated in each evaluation performance category. Michelle Rhee, chancellor of the Washington, D.C. public schools, has proposed evaluating teachers largely on the basis of their students' performance. Georgia and Idaho have launched major efforts to reform teacher evaluation at the state level. Meanwhile, researchers have noted that a well-designed and implemented teacher evaluation system may be the most effective way to raise student achievement.¹ And teacher evaluation reaches schools and districts in every corner of the country, positioning it to affect important aspects of schooling such as teacher collaboration and school culture, in addition to student achievement.

Historically, teacher evaluation has not substantially improved instruction or expanded student learning. The last major effort to reform teacher evaluation, in the 1980s, petered out after much fanfare. Today there are reasons to believe that conditions are right for substantive improvements to evaluation. Important advances in our knowledge of effective teaching practices, shifts in the composition of the educator workforce, and changes in the context of public education provide a key opportunity for policymakers to tighten the link between teacher evaluation and student learning. What's more, some districts have already instituted rigorous teacher evaluation programs that affect instruction and learning.

This report will explore how best to implement teacher evaluation. The first section examines the structure of teacher evaluation and the role of student learning in assessments of teachers' effectiveness. Across a variety of approaches to evaluation, students' learning and achievement typically play a small role in the evaluation of their teacher. The approaches reviewed—observation, performance-based assessments, portfolios, and value-added analysis—have distinct strengths and clear weaknesses. For this reason the paper argues that the most robust approach likely combines these methods to capitalize on their benefits and minimize their drawbacks.

In the second section, the paper draws on research to examine the reasons why teacher evaluation has generally had little effect on instruction, learning, and achievement. Teacher evaluations often suffer from the “Lake Wobegon effect”: Most if not all teachers receive satisfactory evaluation ratings.² It is possible that all teachers are above average in some schools, but there is generally more variation in teacher effectiveness within schools than

between them. Thus, *any* school—low -performing or high -performing, wealthy suburban or under-resourced urban—is likely to employ more under-performing teachers than its evaluation ratings suggest.³ In fact, principals *and* teachers believe that teachers are less effective than evaluation ratings would indicate.⁴

Multiple factors, often working in tandem, produce this effect. External constraints decrease evaluators’ inclination to evaluate rigorously—vague district standards, poor evaluation instruments, overly restrictive collective bargaining agreements, and a lack of time all contribute to this problem. Internal constraints, such as the absence of high-quality professional development for evaluators, a school culture that discourages critical feedback and negative evaluation ratings, and a district culture that offers little oversight and few incentives for administrators to evaluate accurately, also contribute to inflated ratings.

Evaluation has few negative or positive consequences, which is a reality that reduces evaluators’ will to evaluate accurately and thoroughly and teachers’ motivation to take evaluation seriously. Evaluators rarely provide teachers with substantive feedback, which further reduces evaluation’s impact on teaching and learning. Across all these factors, the teachers union, the structure of evaluators’ jobs and training, and the culture of schools strongly influence the quality of teacher evaluation and whether it improves teaching and learning.

The report’s third section assesses the current prospects for teacher evaluation reform, concluding that the time is right for major change. Traditional public schools now face pressure from without by charter schools, voucher programs, and the growing home-school movement and from within by accountability measures. They can no longer do business as usual, or merely assert that their teachers are “highly qualified.” Increasingly public schools must demonstrate that their teachers are effective.

Retirements in the educator workforce are changing the face of the classroom and main office in today’s schools. New educators, both teachers and principals, are more receptive to differential treatment of teachers than were prior generations. Seventy percent of new teachers in a representative sample said that the fact that teachers do not get rewarded for superior effort and performance is a drawback.⁵ Eighty-four percent of these teachers said that making it easier to terminate unmotivated or incompetent teachers would be an effective way to improve teacher quality. The influx of so many new educators also provides an opportunity for supervisors to evaluate teachers more rigorously now, before these individuals gain tenure.

Lastly, we now have more developed collective knowledge about good teaching and the infrastructure to support pedagogical change. The National Research Council’s work on how children learn has produced a growing body of knowledge on how best to teach. Inquiries into pedagogical content knowledge are helping researchers and policymakers develop an increased understanding about effective teaching approaches. State curriculum frameworks and assessment systems have been widely implemented over the last 20 years, providing

the infrastructure to promote systemic improvements in teachers' pedagogy. States have also developed databases to track student (and, in some places, teacher) progress over time. Finally, substantive improvements to teacher evaluation are underway in sites around the country. The paper specifically examines Cincinnati's Teacher Evaluation System, or TES, a program that could affect teacher and student learning. With TES, Cincinnati has systematically addressed many of the problems that plague teacher evaluation.

With conditions ripe for reform, districts and states need the will to make the structural and—more importantly—cultural changes necessary to improve teacher evaluation in substantive and meaningful ways. The paper concludes by offering seven recommendations to districts and states that seek to reform teacher evaluation to increase its impact on teaching, learning, and achievement.

Teacher evaluation in the hot seat

Students of educational policy know that we have been down the road to better teacher evaluation before. In the wake of 1983's *A Nation at Risk*, interest in reforming teacher evaluation flourished in connection to merit pay.⁶ Evaluation and merit pay drew support as a strategy to address the “rising tide of mediocrity” in the nation's schools and, by extension, within the teacher workforce.⁷ However, when the economy contracted and lay-offs occurred in many schools, seniority, rather than evaluation results, generally drove decisions about who would receive pink slips. Similarly, interest and experimentation with teacher evaluation subsided over time, as money to support pay and evaluation reform ran out and new priorities, such as curriculum standards, rose to the fore.

Will current interest in teacher evaluation buck historical trends to change evaluation in ways that improve teaching, learning, and achievement? There are reasons to believe so. Important advances in our knowledge of effective teaching practices, shifts in the composition of the teacher workforce, and changes in the context of public education may tighten the link between teacher evaluation and student learning. Some districts have already instituted rigorous teacher evaluation programs that affect instruction and learning.

Teacher evaluation has the potential to improve instructional effectiveness and student learning by enabling teachers to receive high-quality guidance and feedback, thus improving their instruction. Evaluation could also enable principals to better identify the most and least effective teachers. This would allow them to dismiss ineffective teachers, which would raise the average quality of remaining teachers, and reward highly effective teachers, which would motivate everyone to strive for greater effectiveness.

More vulnerable teachers, more robust evaluation

The intensity and frequency of evaluation vary according to teachers' tenure status.⁸⁹ Nontenured teachers often receive the most frequent, intense evaluation—at least twice per year according to one study of Midwestern districts.¹⁰ In Miami-Dade County, FL, for example, non-tenured teachers are observed at least twice a year and “may be dismissed without cause” during their first 97 days on the job.¹¹ After the end of this probationary period—which is present in many districts—an administrator must follow a delineated process to dismiss teachers.

Once teachers have tenure (after year three in most states, but after year two or four in some), evaluation is typically less frequent and intense. Tenured teachers in approximately half of the Midwestern districts studied by Brandt and colleagues were evaluated once every three years; others tended to be evaluated more frequently.¹²¹³ Most districts use evaluations to identify teachers for remediation. A few districts, including Denver, grant or withhold pay raises based on evaluation results.

State tenure policy—reflected explicitly or implicitly in local evaluation policy—makes it more difficult for principals to dismiss tenured teachers compared to their non-tenured counterparts. State tenure laws set a higher standard for dismissal of tenured teachers in part because courts have found that teachers have a property right to their job under the 14th Amendment to the Constitution. Evaluation policy usually specifies procedures by which tenured teachers may contest the process—as opposed to the substance—of their evaluation.^{14,15}

Student learning is not always on the evaluation map

Student learning plays a varied role in teacher evaluation.¹⁶¹⁷ By far the most common method of evaluation involves observing a teacher’s instruction. Pure observation does not consider student achievement as measured by standardized test scores, but may consider student learning as viewed during the observed class period.

In 60 percent of the districts Brandt and colleagues studied, observation played a role in teacher evaluations.¹⁸ Formal observations are generally structured around an observation instrument that measures teachers’ performance on an assessment metric. For example, Philadelphia has a dichotomous satisfactory/unsatisfactory rating scale, while Tulsa’s includes four categories. Evaluators may observe for an entire class period or “walk through” a classroom to observe only briefly. “Walk-throughs” are increasingly popular, and some states, including Texas and Tennessee, provide training and instruments to guide their implementation. Informal observations are typically less structured.

Many districts and states have moved toward performance-based assessments as a means for teacher evaluation. This method involves observation, but assesses the teacher’s instruction against an articulated set of performance standards translated into detailed descriptors within a rubric. Charlotte Danielson’s *Framework for Teaching*, aligned with the Interstate New Teacher Assessment and Support Consortium, or INTASC, standards is one of the most common systems. It has served as the foundation for evaluation systems in cities such as Cincinnati, OH and Reno (Washoe County), NV and in states such as Iowa and Idaho.¹⁹

Some argue that performance-based assessments deliberately include teacher behaviors that, according to research, promote student learning. Thus these evaluations provide an indirect assessment of student outcomes. Additionally, some performance-based evalu-

ation includes consideration of student work.²⁰ Ultimately, however, the performance-based assessment method is focused more on teacher inputs than student outputs.

Some districts evaluate teachers based on a portfolio. Many portfolios include a performance-based element and some required evidence of student learning.²¹ The National Board for Professional Teaching Standards portfolio, one of the best-known examples, includes video of a candidate teaching, samples of student work, and evidence of impact beyond the classroom in working with colleagues, community members, or families. Additionally, National Board candidates must pass content- and grade-specific examinations. Candidates' entries are scored by a minimum of 12 teachers who have been trained in how to evaluate these portfolios. Connecticut has used a content-specific portfolio system coupled with mentoring in its licensure decisions since 2000 in its Beginning Educator Support and Training Program.

Other new and promising programs are using portfolios to assess teachers' work. As part of its three-tier licensure system, New Mexico requires teachers to submit a Professional Development Dossier, which includes their lesson plans, student work, and reflections alongside their superintendent's verification of the dossier's accuracy and a recommendation by their superintendent regarding whether they should move to the next level of licensure. Two reviewers outside of the teacher's district rate portfolios. Portfolios will also play a prominent role in tenured teachers' evaluations in Oregon districts revising their approaches to teacher quality in an initiative led by the Chalkboard Project. Although portfolios do include student work, in most cases teachers choose which work to include, thus introducing self-report bias.

Value-added analysis, a body of statistical approaches that enable researchers to estimate a student's achievement *growth* in a specified time period, focuses squarely on student outcomes. Many districts and states are now laying the groundwork to base teacher evaluation at least partially on a teacher's impact on her students' achievement. Ten states have regulations supporting the use of student achievement data in teacher evaluation in a general way.²² Texas and Tennessee have led the way in developing the data systems to track student achievement over time and link these data to individual teachers. Delaware specifies that 20 percent of a teacher's evaluation be based on student achievement and Texas bases one-eighth of its evaluation on value-added estimates.

Value-added analysis has clear benefits when compared to more traditional methodologies for teacher evaluation. Compared to administrator observations, student test scores may be less vulnerable to administrator bias or favoritism. Compared to portfolios, value-added analysis does not suffer from self-report bias. Finally, if the ultimate goal of teaching is student learning, evaluation should privilege that outcome.

Despite these compelling arguments, the notion that value-added analysis adequately captures a teacher's contribution to her students' learning is a hard sell in many quarters. Some argue that value-added analysis should be rejected because a large number of

teachers—as high as 69 percent in an estimate for Florida—do not teach subjects that are tested on state-mandated assessments.²³

Others accept the validity of student tests as a measure of teacher effectiveness, but raise questions about whether the value-added measure isolates a teacher's impact on her students' test scores. Neither students nor teachers are randomly assigned to classes, and it's likely that value-added estimates contain some bias because of systematic patterns of student and teacher assignment. Moreover, researchers who conduct value-added analyses are not in agreement about how to estimate these models. For example, some researchers control for students' background characteristics in their value-added models while others do not.

The research on various approaches to teacher evaluation suggests that they all provide some information on student learning. Jacob and Lefgren found that principals' ratings of teachers' ability to raise student achievement were correlated with the teachers' actual value-added student gains.²⁴ These principal ratings were not formal evaluation results, but they were likely influenced by observation-based evaluation. The researchers found that principals successfully identified teachers who produced the highest and lowest value-added student gains, but they were not as good at identifying those in the middle.²⁵

The research on performance-based evaluation suggests that teachers' ratings have a small to moderate correlation with value-added estimates of student achievement. Anthony Milanowski found that ratings were correlated with value-added estimates in math (.43), reading (.32), and science (.27) in Cincinnati.²⁶ Similarly, H. Alix Gallagher identified a relationship between evaluation scores and student achievement in reading (.50) in a Los Angeles charter school.^{27,28}

Much of the research on portfolios has investigated the relationship between National Board certification and student achievement. This evidence is mixed. On the one hand, the certification process does appear to select the most effective teachers from the applicant pool.²⁹ Moreover, Goldhaber and Anthony found that students taught by National Board certified teachers in North Carolina elementary schools experienced significantly larger gains in math and reading than did similar students taught by non-Board certified teachers.^{30,31} However, a more recent study of National Board certified teachers in Florida between 2000 and 2004 in grades 3-10 found that future NBCT's were not more effective than non-Board certified teachers as measured by most value-added estimates.³²

Research has shown that portfolios can be valid and reliable.^{33,34} One study on Connecticut's BEST program found that students of teachers who scored highest on the portfolio assessment gained three additional months of learning over one year than those taught by the lowest scoring teachers.³⁵

Value-added evaluations implicitly provide the most direct evidence on student achievement. However, they also suffer some limitations. Value-added estimates for the same teacher are likely to vary from year to year. Goldhaber and Hansen found modest cor-

relations (.30 in reading, .52 in math) in the year-to-year teacher effects estimates in their study of fifth-grade teachers in North Carolina.³⁶ Value-added estimates for the same teacher in the same year also vary across different kinds of student tests. Papay found small to moderate correlations (.17 to .51) among value-added estimates based on three reading achievement tests administered in one district.³⁷ Moreover, any student achievement test reveals only a small slice of what a teacher imparts to her students.

All of these teacher evaluation methods provide some information about the link between teaching and learning, but they also suffer important limitations. The wisest approach may be to blend different approaches to get a more robust and multi-faceted understanding of a teacher's influence on her students' learning and achievement.

Who has a say in teacher evaluation?

Principals and assistant principals evaluate teachers in the great majority of schools. Teachers, most often department heads or instructional supervisors, sometimes evaluate other teachers. Some note that teachers may have subject or grade-specific knowledge that makes them especially good at evaluating their peers³⁸

Formal peer assistance and review occurs in some districts. In these systems, teachers assume primary responsibility for evaluating, supporting, and, in some cases, dismissing their non-tenured and struggling tenured peers. Peer review started in Toledo, OH in 1981, and is now used in Cincinnati and Columbus, OH; Minneapolis, MN; Montgomery County, MD; Rochester, NY; and, at a small scale, Chicago, IL and New York City. Although peer assistance and review has aroused substantial interest, experts estimate that no more than 50 districts nationwide have developed peer review systems.³⁹

Evaluation in some places incorporates feedback from parents, students, or other school workers such as custodians and secretaries.⁴⁰ Blaine County, ID uses 360-degree evaluation techniques, popular in the private sector, to assess teachers and principals by collecting data from parents, students, peers, and supervisors. A range of districts, including Charlottesville, VA and Newton, KS, incorporate parent feedback into teachers' evaluations. Many states have created committees to design teacher evaluation systems, and Utah and Florida encourage parent participation while Colorado, Kentucky, Louisiana, and New York encourage citizen and student participation.⁴¹

Why has evaluation failed to improve teaching and learning?

There are a number of reasons why teacher evaluation has generally failed to influence teacher quality and student learning. Across these factors, the interests of teachers unions, the structure of evaluators' jobs and training, and the culture of schools contribute to whether evaluation improves instruction and achievement.

The Lake Wobegon effect

Most evaluators give most teachers positive ratings on summative evaluations. Over four years, nearly 100 percent of Chicago teachers were rated “satisfactory” or above.⁴² Ninety-six percent of San Bernardino, CA’s teachers met or exceeded expectations for the 2002-03 and 2003-04 school years.⁴³ This phenomenon is not unique to urban districts. Between 1995 and 2005 only 1 in every 930 teachers (.1 percent) in Illinois received an unsatisfactory rating.⁴⁴ Ninety-nine percent of Oregon teachers are rated satisfactory each year.⁴⁵

It is possible that all teachers are above average in some schools, but there is generally more variation in teacher effectiveness within schools than between them. Thus, *any* school—low-performing or high-performing, wealthy suburban or under-resourced urban—is likely to employ more under-performing teachers than its evaluation ratings suggest.⁴⁶ In fact, principals *and* teachers believe that teachers are less effective than ratings indicate.⁴⁷

This is problematic because every time an evaluator issues a satisfactory rating to a teacher who is teaching unsatisfactorily it becomes more difficult to fire her based on performance.⁴⁸ Moreover, when teachers who are performing effectively see their ineffective colleagues receive satisfactory evaluations, it may diminish their motivation to do well. Rating inflation is also problematic because it makes it difficult to reward teachers whose performance is truly outstanding. Some teachers strive for extrinsic recognition as excellent teachers and are discouraged when such recognition is not granted.

The road to Lake Wobegon

Evaluators fail to accurately assess teachers because the infrastructure, resources, and incentives of evaluation and the culture of schools rarely support differentiation among teachers.

These factors interact, creating a classic “chicken-and-egg” problem regarding teacher evaluation’s design and enactment, on the one hand, and its consequences, on the other.

Few consequences typically result even when well-designed evaluation instruments are implemented carefully. Principals are reluctant to push for consequences when they know their instrument or its implementation is weak. They are reticent to push for stronger instruments or implement them more carefully when they fear the consequences won’t stand. This suggests that efforts to reform the structure and content of teacher evaluation must also take into account the culture of the school and district in which it exists.

Instrument failure

The design and implementation of evaluation processes in many cases run counter to evaluators’ efforts to differentiate among teachers. Until recently, many instruments failed to reflect the complexity of teaching. Instead, they included checklists of items that were easily measurable (the neatness of classroom bulletin boards, for example) but not central to instruction. Some instruments included only a few categories (satisfactory and unsatisfactory), which made it difficult for evaluators to differentiate among teachers.

A district’s evaluation standards do not always align with its instructional focus.⁴⁹ Thus, evaluators sometimes assess teachers based on a tool that ignores or even contradicts the instructional practices the district promulgates through professional development. All of these factors combine to make the instrument less meaningful than it should be. Indeed, among Chicago principals, 73 percent said that the evaluation instrument did not “accurately or meaningfully assess performance.”⁵⁰

External constraints to differentiation

Like the specific evaluation instruments, district standards regarding teacher evaluation are often inadequate. Brandt and colleagues found that district policy was more likely to specify the evaluation process and procedures (who does it, how often, when) than the substance of the evaluation (what it should focus on, the criteria and standards for evaluation, and consequences).⁵¹ Based on their analysis of case studies of human capital management in five large, urban districts, Koppich and Showalter found that “even when a common set of [evaluation] standards might be used [within a district], few if any documents define what the standard is intended to connote, and there are no performance rubrics for any of the standards.”⁵²

Some district policies explicitly inhibit evaluators’ attempts to differentiate among teachers. Chicago principals must follow extra procedures to rate a teacher lower than she was rated in a previous evaluation. Fifty-one percent of principals said that the contract limited

their ability to downgrade a teacher's ratings.⁵³ The infrequency of tenured teachers' evaluation—sometimes only once every five years and sometimes for as little as one class period—also limits its impact on professional growth and instructional quality.⁵⁴

Another key external constraint is time. Administrators must evaluate not just teachers but also custodians, guidance counselors, and secretaries.⁵⁵ Lack of time is a major obstacle to high-quality, accurate evaluation. One study estimated that principals spent 7 to 10 percent of their professional time (100 to 150 hours) evaluating 20 teachers.⁵⁶ Principals whose school employs many non-tenured teachers likely spend much more than 10 percent of their time conducting multiple evaluations. As the Baby Boomer Generation retires, teacher evaluation will constitute an increasing portion of most principals' work in years to come.

Internal constraints to differentiation

Accurate evaluation requires considerable knowledge and commitment on the part of the evaluator. Evaluators need to know and be able to identify the tenets of good instruction. Yet this is no small task, given the complexity of teaching and the fact that administrators evaluate teachers of subjects or grades with which they are not familiar.⁵⁷ Without high-quality professional development, evaluators will not evaluate accurately and the evaluation will likely have little impact on teaching or learning.

Evaluation training for principals is not universal despite its critical role. Only 75.2 percent of districts nationwide reported offering professional development in evaluation or supervision to principals in 2003-04.⁵⁸ Ninety-nine percent of Connecticut districts offered such training, whereas 52.3 percent of Colorado districts and 44.7 percent of Montana districts did so. This says nothing regarding the quality of professional development, which most research suggests has been lackluster.⁵⁹

Evaluators must also have the will to differentiate among teachers. Of all obstacles to differentiation, this is the most intractable. Generally, evaluators' ratings have been subject to scant oversight.⁶⁰ They typically have few incentives to give teachers unsatisfactory ratings and, in fact, have compelling reasons to deem most teachers satisfactory.

Principals must weigh the benefits and costs of issuing unsatisfactory evaluations. Most schools have been governed by the norms of teachers, who constitute the adult majority in any school building. Cultural norms, sometimes reinforced by contractual language, have limited intrusions into individual teacher's rooms, even for the purpose of evaluation by a principal or lead teacher.⁶¹ A principal who "cracks down" by giving critical feedback and an unfavorable evaluation rating in some ways acts counter to cultural norms and risks losing the cooperation of teachers on whom she or he depends to keep the school running.

Moreover, even if a principal follows the prescribed process, issues unsatisfactory ratings, and “evaluates” an underperforming teacher out of his or her school, there is no guarantee that the replacement teacher will be an improvement over the dismissed one. In many places, principals do not have full discretion over hiring and could be forced to assume a substandard replacement teacher.⁶² Some principals likely prefer the proverbial “devil they know” rather than gambling on the “devil they don’t.”

Evaluation does not improve teachers’ skills and knowledge

An evaluation can be summative, simply conferring a status, or it can be formative, used to inform a teacher’s future practice. The effect of formative evaluations hinges on feedback, the quality of which is a central concern of teachers.^{63 64} And despite the importance of feedback to teachers, the evaluation process often provides little of it. 28.8 percent of San Bernardino, CA, teachers in 2002-03 and 32 percent of those in 2003-04 reported that they received no feedback or suggestions during their evaluation process.^{65 66} This may explain why only 26 percent of teachers in one recent study reported that evaluation was “useful and effective.”⁶⁷

Some evaluators may provide little feedback because they view evaluation primarily as an opportunity to motivate teachers. Teachers are, at heart, optimists, and almost all evaluators have been—or are—teachers.⁶⁸ Some principals argue that if all children can learn, all teachers, with enough support, can ultimately teach. Kimball and Milanowski found that, in an effort to motivate teachers, evaluators they interviewed tended to frame evaluation as formative, whether or not it actually was, and allowed teachers substantial input in defining the evaluation focus and evidence to be gathered.⁶⁹

Evaluation has few consequences, positive or negative

Any effort to improve teacher evaluation’s impact on teaching and learning must specify and reinforce its consequences. In fact, some advocate dismissing a substantial proportion of new teachers based on their performance. From their work in Los Angeles, Kane and Staiger have concluded that a district could drop the lowest performing two-thirds—measured by value-added score—of first-year teachers and achieve a 3-4 point increase in student achievement over time.⁷⁰

To date, evaluation has resulted in few consequences, negative or positive. The statistics on actual firings are quite low, but some teachers who have received unsatisfactory evaluations may resign to avoid being fired. In 2003-04, districts across the nation dismissed or did not renew on average 3.1 teachers of approximately 209.7 teachers in total.⁷¹ Fourteen of 11,000 teachers in Philadelphia were fired in 2003-04; 12 of 2600 in Pittsburgh were dismissed or non-renewed.⁷² Despite the lower bar for firing non-tenured teachers,

districts do not seem to take advantage of this relative freedom.⁷³ Of the 3.1 teachers fired—on average—in 2003-04 in each district, only 1.2 of these were teachers with three or fewer years of experience.⁷⁴

Evaluators' emphasis on motivation at the expense of critical feedback also breeds a reluctance to issue negative evaluations and move to dismiss teachers unless they are extremely ineffective. The belief "that teachers deserve a chance to improve" prompted 51 percent of Chicago principals to refrain from not renewing non-tenured teachers. This belief led 39 percent of these administrators to fail to pursue dismissal of tenured instructors.⁷⁵

Teachers unions certainly limit the power of principals to fire underperforming teachers. They must protect teachers' due process rights and represent them fairly. Some principals avoid the dismissal process because they fear the teachers' union will battle them every step of the way.

Yet dismissing tenured teachers is possible. Although the Duty of Fair Representation requires union leaders to fairly represent all teachers in their district, the AFT and NEA cite private sector case law to conclude that "a union has the discretion to refuse to process or pursue a member's grievance, even a dismissal grievance, without violating its duty of fair representation."⁷⁶

Although the negative consequences of evaluation receive ample attention, evaluation has also generally failed to identify outstanding teachers who are then rewarded. Forty-nine percent of teachers in a recent survey reported that district and school officials "do not reward outstanding teachers; the reward is solely intrinsic."⁷⁷ Failure to reward outstanding performance may suppress some teachers' motivation to work hard, pursue learning opportunities, and improve their instruction.

Changes in the educational context, workforce, and technologies

The obstacles to improving teaching and learning through teacher evaluation are considerable, but there are reasons to believe that we can make improvements. The changing context of American public schools, the shifting composition of the educator workforce, and the increasing sophistication of our knowledge of teaching and learning suggest the time is right for substantive evaluation reform.⁷⁸ Some districts already practice rigorous evaluation that leads to substantial professional growth and real consequences for teachers whose performance lacks and for those whose instruction is truly outstanding.

Increased accountability and competition have raised the stakes for American public schools in recent years.⁷⁹ The expansion of charter schools, voucher programs and the home-school movement, on the one hand, and elevated accountability on the other mean that “regular” public schools can no longer do business as usual. They must demonstrate a commitment to raising teacher quality and producing learning results. Drawing on private sector models, superintendents and human resource directors have begun to develop comprehensive approaches to human capital development at the district level. Many districts have begun by reforming hiring practices, but leaders intend to systematically improve all aspects of human capital development, including evaluation.⁸⁰ In an increasingly competitive market, public schools can no longer assume that teachers are high quality; they must demonstrate this fact.

There is some evidence that recent public pressure has resulted in improvements to evaluation. Halverson and colleagues found that citizen demands led their case study district to implement a standards-based evaluation system with a greater focus on student learning, increased frequency of observation of experienced teachers, and an improved remediation program to give principals greater latitude to dismiss poor performers.⁸¹

There is also evidence that teachers unions are increasingly willing to collaborate with districts on matters of teacher quality. Johnson and colleagues found that the union presidents they interviewed were pursuing reforms in teacher evaluation, compensation, and assignments.⁸² These leaders stated that principals do not take full advantage of their capacity to “evaluate teachers out” of their school and expressed a willingness to accept the process as long as it followed the requisite legal procedures.⁸³ They expressed a commitment to teacher quality, noting that defending clearly substandard teachers sullies their reputation and reduces the stature of the teaching profession. The broader

educational backdrop, they said, meant that failure to work with management on teacher quality could mean the end of public schools, let alone teachers' unions.

Unions are also raising the bar for teaching quality through peer review. Some scholars argue that peer review allows evaluators to assess teachers more accurately, provide better feedback, and recommend teacher dismissal more often than their administrator counterparts. These peer reviewers operate differently because they are often full-time evaluators, they must justify their evaluations to a panel of teachers and administrators, and they receive substantial start-up and ongoing training.

Perhaps most important, peer assistance and review brings teachers' unions to the table from evaluation's outset, potentially reducing union opposition to a teacher's dismissal.⁸⁴ These arguments have borne out in Montgomery County, MD. A teachers' unions official reported that one person was fired for performance in the decade before peer assistance and review was instituted; in the first six years it was in place, 400 teachers were dismissed.⁸⁵

Changes in the educator workforce may also tighten the relationship between evaluation, teacher quality, and student performance. According to some estimates, more than one-third of the U.S. teacher workforce is projected to retire in the next four years.⁸⁶ As Baby Boomers exit, members of Generation X and the Millennial Generation will form a larger portion of the teacher workforce. These people expect to be assessed based on their performance and receive rewards, if they perform well, or sanctions, if they do not. Seventy percent of new teachers in a representative sample said that the fact that teachers do not get rewarded for superior effort and performance is a drawback.⁸⁷ Eighty-four percent of these teachers said that making it easier to terminate unmotivated or incompetent teachers would be an effective way to improve teacher quality.

New teachers are also more likely to view reforms like differential rewards or compensation more favorably than veterans. In a recent study, 65 percent of teachers with less than five years of experience wanted the union to negotiate to pay teachers based on their performance, compared to 45 percent of more experienced teachers. Fifty-eight percent of newer teachers but only 39 percent of veterans agreed that teachers in their school "who do a stellar job" should be rewarded. Some research suggests that new teachers generally support individual, performance-based pay based on objective assessments.⁸⁸

These changes in teachers' attitudes have implications for the principal workforce, which is also experiencing massive turnover due to retirement. For instance, in Chicago's 700 schools, 355 principals and 329 assistant principals retired between 2002 and 2007.⁸⁹ To the extent that new principals share new teachers' views, there is even more potential to link evaluation more tightly with teaching and learning.

Lastly, with so many educators retiring and an increasing number of non-tenured employees, districts and schools have a growing opportunity to change evaluation habits.

Evaluators have the opportunity to evaluate new hires rigorously and fairly, moving to dismiss those who do not demonstrate sufficient growth. Of course, this contention assumes that replacement teachers will be available and more effective than those who are dismissed.

In addition to changes in the politics of public education and the composition of the educator workforce, our approach to teaching and learning has become more sophisticated. The National Research Council's efforts to describe how teachers should teach based on how students learn have advanced our collective understanding of effective teaching.⁹⁰ Research has further developed the idea of standards-based teaching and, through the concept of pedagogical content knowledge, improved our notion of good teaching at particular grade levels and specific subject areas.⁹¹

The infrastructure to disseminate effective teaching practices has developed considerably since the last major effort to reform teacher evaluation in the 1980s. State curriculum frameworks and assessments are now the norm, in contrast to two decades ago. Early-career teachers welcome these developments, viewing curriculum standards as a source of support and guidance rather than an infringement on their professional autonomy.⁹² With common curriculum and assessment more widespread than in the past, instruction may be the next frontier in improving the quality of public education.

At the same time, advances in computer technology have enabled states to collect data on students and teachers and track performance over time. For example, New York, New Mexico, and Colorado passed laws in 2007 to make their data systems more robust. New York's law enables student data to be tracked from pre-school through college, while New Mexico plans to track teacher candidates into their jobs, and Colorado is developing a teacher identifier system. These advances suggest more states will incorporate student performance data in teacher evaluation in the future.

Proof Positive: Cincinnati⁹³

Models of effective teacher evaluation already exist. Cincinnati's Teacher Evaluation System, or TES, appears to influence teaching and learning.⁹⁴ Based on Charlotte Danielson's Framework, TES grew out of the 1997 collective bargaining agreement between the Cincinnati Board of Education and the Cincinnati Federation of Teachers. With careful study and advice from experts in research and practice, Cincinnati has developed a program that reduces many of the challenges outlined above.

TES rests on an instructional framework composed of 16 standards arrayed within four domains: (1) Planning and Preparing for Student Learning, (2) Creating an Environment for Student Learning, (3) Teaching for Student Learning, and (4) Professionalism. TES incorporates teachers as evaluators to reduce reliance on administrators and increase the quality of evaluation. Teacher Evaluators conduct the majority of tenured teachers' observa-

tions during the comprehensive evaluation cycle, while administrators do annual observations. TES also intersects with the district's Peer Assistance and Evaluation Program, or PAEP, which assigns Consulting Teachers to evaluate new teachers and struggling veterans.

TES requires that teachers be observed much more frequently than is typical. New teachers receive at least two formal and two informal evaluations before December of their first year in the district. If they do not meet expectations, they receive four additional formal observations within the year. Only the first of these observations is announced. New teachers who meet expectations then receive one annual evaluation that is also unannounced. At year four, they receive a comprehensive evaluation based on four formal observations of their teaching. If teachers succeed and receive tenure, they receive one annual evaluation per year by their administrator and a Comprehensive Evaluation based on four formal observations every five years.⁹⁵

This does not prohibit a tenured teacher from being observed more frequently. If an administrator, school-based union representative, or even a peer believes a teacher is ineffective, he or she can recommend that individual receive remediation. The principal then gathers evidence by observing the teacher twice in a period of 10 to 30 days and offering assistance based on the observation. If the teacher fails to demonstrate improvement, the principal officially refers the teacher to the PAEP joint panel, composed of equal numbers of Cincinnati Federation of Teachers members and administrators. This Board reviews the case and decides whether to put the teacher on intervention, which consists of intensive assistance and observation provided by a Consulting Teacher, or CT. The CT then recommends whether the teacher should be dismissed or remain on intervention for another year. If the teacher's evaluations demonstrate growth, she moves out of intervention status. If a teacher consistently underperforms, a principal can initiate this process in September, refer the teacher to intervention in December and, after intervention by the CT, dismiss the teacher by March of the same school year. Typically, the process takes one and a half years from referral to dismissal.

TES systematically addresses many of the challenges most teacher evaluation systems face

Challenge: Lack of time

TES Solution: TES reduces the time constraint many evaluators face by dividing evaluation responsibilities between administrators, CTs, and Teacher Evaluators, or TEs. In practice, the role of CT and TE is often played by the same teacher, who is released from regular classroom teaching for three years to evaluate his/her peers. Each full-time evaluator has a case load of 18-25 teachers and is thus able to conduct higher quality evaluations and provide better feedback to teachers.

Challenge: Lack of evaluator skill

TES Solution: Cincinnati devotes substantial time to providing professional development on its evaluation standards to evaluators and teachers. New CTs and TEs receive 10 to 11 days of training before starting their role. Continuing evaluators receive five days of training each summer. Over the course of the school year, CTs and TEs receive two hours of training every other week to review evaluation standards and calibrate scoring. After three years in their role, CTs and TEs cycle back into full-time teaching, which keeps their knowledge of teaching and learning current. All evaluators, CTs, TEs, and administrators, must pass an evaluator certification test that requires them to assess instruction using the TES rubrics.

Challenge: Lack of evaluator will

TES Solution: One way that Cincinnati has increased evaluators' will to evaluate accurately is by having multiple evaluators for each teacher. This increases the accountability for individual evaluators, as they know their

work will be reviewed by others well versed in the evaluation system. Cincinnati is also beginning to evaluate principals on the basis of student performance, which holds promise for motivating principals to assess truthfully and place underperforming teachers on intervention early.

Challenge: Formative evaluation fails to improve instruction

TES Solution: Evaluators receive substantial training not only on the framework and how to use it, but also on how to coach teachers to improve their practice. Teachers participate in professional development that focuses on TES standards and expectations. The heavy emphasis on coaching and assistance, particularly in the non-tenured years, provides teachers with considerable feedback.

Challenge: Evaluation has few consequences

TES Solution: As part of TES, Cincinnati has created advanced career levels for teachers. To achieve Lead Teacher status, a teacher must receive a "Distinguished" (4) score on a written application based on Domain 4 (Professionalism) of TES and an "Advanced" or "Accomplished" overall rating on a Comprehensive Evaluation. Thus, teachers who perform at the highest level on their evaluations are rewarded.

The consequences of poor evaluation performance are also clear. A teacher who is placed on intervention before December 1 of a school year could be dismissed within three months.

So long, Lake Wobegon?

Cincinnati is clearly leading the way on evaluation reform, but it is not alone in its efforts to improve teaching and learning through teacher evaluation. Across the country, districts are reforming their evaluation systems to tighten the link with instruction and learning. Montgomery County, MD, has drawn on Jonathan Saphier's work to cultivate a robust understanding of teaching and learning among all educators in the district and raise the quality of teaching in the process. Williamsburg-James City County, VA, public schools have instituted a portfolio in addition to administrator observations. Over two years, the proportion of teachers receiving the top rating declined from 75 percent to 43 percent and most teachers and principals viewed the portfolio as fair and accurate.⁹⁶ Oregon's Chalkboard Project and partner districts are spending the entire first year of their initiative on training teachers and evaluators in evaluation standards and processes.

New York City recently launched a pilot program, supported by Schools Chancellor Joel I. Klein and the United Federation of Teachers, that rates fourth- to eighth-grade teachers on the basis of their students' test scores. Principals will likely use these ratings to determine teaching assignments and professional development needs. New York has also instituted a form of peer review for tenured teachers at risk of being deemed incompetent. A teacher who is not a member of the UFT or an employee of the district coaches a referred teacher for 10-12 weeks, at which point she recommends whether or not the teacher should be fired. Principals consider this recommendation and, if the teacher appeals the case, the recommendation is admissible as evidence at the appeal hearing.

Any effort to tighten the link between teacher evaluation and teaching and learning should attend to the challenges that plague evaluation and build on the features successful programs embody. Evaluation programs have the best chance of improving teaching quality and student learning if they include:

- **An extended development phase:** Cincinnati's Teacher Evaluation System and Denver's ProComp have benefited immensely from time spent gathering input from stakeholders and experts, and designing, piloting, and revising the programs. Teaching is complex and schools are multifaceted. It is naïve and unwise to believe that a robust and effective evaluation system be developed overnight and in isolation.

- **Valid, reliable instruments:** The instrument must reflect what teachers and evaluators consider to be the essential elements of instruction *and* yield reliable results. If it fails on either of these accounts, teachers and evaluators will not take it seriously. This suggests that all stakeholders should be involved in some capacity in developing the instrument. It also suggests that the evaluation instrument should reflect the district's vision of effective instruction and align with the content of induction and professional development programs.
- **Multiple measures:** Any one measure of teacher effectiveness is limited. The best programs consult multiple data sources over time to draw conclusions about teachers' effectiveness. Multiple evaluators bring different expertise to the task and provide checks and balances for others' assessments.
- **Robust professional development for evaluators and teachers:** Evaluators and teachers must develop a shared understanding of what constitutes good teaching. Professional development for evaluators and teachers must go beyond the managerial aspects of evaluation—the timeline and required forms—to focus on the instructional components of the task—what is good teaching, how do you know it when you see it, and how do you reduce unfounded biases.⁹⁷ Districts that continually discuss these questions achieve a more robust evaluation system *and* more substantial instructional progress.
- **Accountability, incentives, and support for evaluators:** District-level officials should provide more support for school administrators by evaluating them thoughtfully and thoroughly and offering assistance on non-instructional matters so that they may focus on instruction and evaluation. When principals dismiss teachers, the district should not undermine principals by failing to follow through on their decision or by forcing them to take a sub-par replacement. They should also provide administrators incentives for thorough evaluation by offering them rewards for detailed feedback. Lastly, they should pressure administrators to evaluate accurately by reviewing evaluation reports and by incorporating an analysis of principals' evaluations of teachers into district-level evaluations of principals. The fact that principals' evaluations are increasingly tied to student performance may prompt principals to evaluate more accurately and thoroughly.⁹⁸
- **Integration within a human capital system:** Lastly, reform of teacher evaluation policies and practices is less likely to endure, let alone improve teaching and learning, if it conflicts with the instructional emphasis of concurrent district reforms. Teacher evaluation exists within a system of efforts to influence what and how teachers teach and students learn. A district could have a strong evaluation system, but if hiring, induction, and professional development do not reflect the instructional vision at the core of that evaluation system, it will not succeed. As Heneman and Milanowski conclude, “any particular HR innovation occurs within the context of a broader HR system,” which means that any innovation “needs to be meshed with planned changes in other HR practices in order to drive meaningful and sustained teacher quality improvement.”⁹⁹

Conclusion

Officials who seek to increase evaluation's influence on teaching and learning have substantial work ahead of them. School administrators charged with carrying out new evaluation programs will confront long-standing cultural barriers to making accurate judgments about teachers' instruction and attaching consequences of real import to these assessments. They will do so in an economic climate that dictates they will have fewer resources in terms of time, money, and personnel, on which to draw.

However, it's clear that the stakes for children have never been higher. If we are committed to expanding learning and increasing achievement, especially for low-income children and those who are under-performing, we must improve teachers' instruction. Teacher evaluation holds great promise for achieving this aim. To increase the impact of evaluation on teaching and learning, all stakeholders—the teachers union and the school board, district-level administrators and building-level leaders, teachers and the public—must be involved. Without broad, authentic involvement, we risk squandering the momentum around teacher evaluation that currently exists. With careful deliberation, support and accountability for evaluators and teachers, schools may make real progress in raising teacher quality and enhancing student learning through teacher evaluation.

Endnotes

- 1 Donald O. Staiger, Robert Gordon, and Thomas J. Kane, "Identifying Effective Teachers Using Performance on the Job" (Washington: The Brookings Institution, 2006).
- 2 For an example, see Eric A. Hanushek and others, "The Market for Teacher Quality" Working Papers 11154 (National Bureau of Economic Research, Inc., 2005).
- 3 So named after Garrison Keillor's fictional Minnesota town, which he regularly describes as a place where "all the women are strong, all the men are good looking, and all the children are above average."
- 4 Principals and teachers believe that teachers are less effective than ratings would suggest. In Chicago, 56 percent of veteran principals said they had assigned a higher evaluation rating than the teacher's performance warranted. On a recent survey of a random sample of principals nationwide, only 46 percent of all principals gave teachers in their school an overall rating of excellent.
- 5 Jonathan Rockkind and others, "Getting Started: A Survey of New Public School Teachers on their Training and First Months on the Job." In Carol A. Dwyer, ed., *America's Challenge: Effective Teachers for At-Risk Schools and Students* (Washington: National Comprehensive Center for Teacher Quality, 2007).
- 6 Noreen B. Garman and Helen M. Hazi, "Teachers Ask: Is There Life after Madeline Hunter?" *Phi Delta Kappa* 69 (9) (1988): 669–672.
- 7 National Commission on Excellence in Education, "A Nation at Risk: The Imperative for Educational Reform" (1983).
- 8 Chris Brandt and others, "Examining District Guidance to Schools on Teacher Evaluation Policies in the Midwest Region" (Washington: U.S. Department of Education, Institute of Education Sciences, 2007).
- 9 Most policies specify whether an evaluation is formative, primarily aimed at providing feedback for remediation or growth, or summative, yielding a final decision, for example, about continued employment. State law often governs the broad structure and consequences of evaluations by specifying evaluation frequency and timelines and the relationship between evaluation and tenure.
- 10 Brandt and others, "Examining District Guidance."
- 11 "2006-2009 United Teachers of Dade/Miami Dade County Public Schools Contract" (2006): 152.
- 12 In Los Angeles, for example, tenured teachers are evaluated once every two, three, four, or five years depending on their prior evaluation results and length of district service. See "2006-2009 Agreement Los Angeles Unified School District and United Teachers of Los Angeles," (2006): 61.
- 13 Brandt and others, "Examining District Guidance."
- 14 An increasing number of teachers' unions are making this clear to their members. For example, union leaders in San Francisco recently distributed a circular stating that "The contract does not allow for grieving the content of the evaluation, so paying close attention to correct procedure is very important. If the timelines are violated, let your Union Field Rep know immediately." See United Educators of San Francisco. "SFUSD TEACHERS: Important Advice from Your Union 2007-2008 School Year" (2007).
- 15 Teachers may respond in writing to the contents of their evaluation but it is quite difficult for teachers to legally challenge the substance of an unfavorable evaluation.
- 16 To assess one method's worth, scholars must employ another method. Since all methods yield proxies for instructional quality that are in essence imperfect, this sets up circular logic: if imperfect method A is correlated with imperfect method B, does this prove that method A is reasonably effective, method B is reasonably effective, or merely that they are correlated with each other? What can be concluded? As long as we cannot identify and measure true instructional quality and true student achievement, we will have to make do with proxies for both. This reminds us to interpret all findings cautiously.
- 17 For an in-depth review of the research on teacher evaluation methods, see Laura Goe, Courtney Bell, and Olivia Little, "Approaches to Evaluating Teacher Effectiveness: A Research Synthesis" (Washington: National Comprehensive Center for Teacher Quality, 2008).
- 18 Brandt and others, "Examining District Guidance."
- 19 Arkansas and Ohio have used a performance-based assessment, Praxis III, for licensure since the early 2000s. California uses performance-based assessment at the credential stage.
- 20 For an example, see Herbert Heneman and others, "Standards-based teacher Evaluation as a Foundation for Knowledge- and Skills-Based Pay" (Washington: Consortium for Policy Research in Education, 2006).
- 21 For a discussion of the ranges of purposes and designs of portfolios used to evaluate teachers, see Pamela D. Tucker and others, "The Efficacy of Portfolios for Teacher Evaluation and Professional Development: Do They Make a Difference?" *Educational Administration Quarterly* 39 (5) (2003): 572-602.
- 22 Helen M. Hazi and Daisy Arredondo Rucinski, "Teacher Evaluation as a Policy Target for Improved Student Learning: A Fifty-State Review of Statute and Regulatory Action since NCLB." *Education Policy Analysis Archives*, 17 (5) (2009).
- 23 Julia Crouse, "Points Would Decide Bonuses," *The Ledger* (2006), as cited in Cynthia D. Prince and others, *The Other 69 Percent: Fairly Rewarding the Performance of Teachers of Non-Tested Subjects and Grades* (Washington: Center for Educator Compensation Reform. U.S. Department of Education, 2008).
- 24 Brian A. Jacob and Lars Lefgren, "Can Principals Identify Effective Teachers? Evidence on Subjective Performance Evaluation in Education." *Journal of Labor Economics* 26 (1) (2008): 101-136.
- 25 Although current value-added estimates did a better job than principals' ratings in predicting future student achievement, the authors note that principal evaluations capture "a broader spectrum of educational outputs." See Jacob and Lefgren, "Can Principals Identify Effective Teachers?"
- 26 Anthony Milanowski, "The Relationship Between Teacher Performance Evaluation Scores and Student Achievement: Evidence From Cincinnati." *Peabody Journal of Education*, 79 (4) (2004): 33-53. Based on these results, he concluded that rigorous, performance-based assessment results do have criterion validity and could be used to make high-stakes judgments about teacher job security and pay, for example.
- 27 H. Alix Gallagher, "Vaughn Elementary's Innovative Teacher Evaluation System: Are Teacher Evaluation Scores Related to Growth in Student Achievement?" *Peabody Journal of Education* 79 (4) (2004): 79-107. Gallagher further finds the school's curriculum, professional development, and evaluation instrument were more aligned with each other and with the student tests in reading than in mathematics.

- 28 Subsequent research on these sites found similar overall relationships between evaluation scores and student achievement. Heneman and others, "Standards-based Teacher Evaluation."
- 29 Steven Cantrell and others, "National Board Certification and Teacher Effectiveness: Evidence from a Random Assignment Experiment" (Cambridge: Harvard University Graduate School of Education, 2008).
- 30 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 (1) (2007): 134-150.
- 31 Researchers have found similar effects at the high school level. See Charles T. Clotfelter, Helen F. Ladd, and Jacob L. Vigdor, "Teacher Credentials and Student Achievement in High School: A Cross-Subject Analysis with Student Fixed Effects." Working Paper 13617 (Washington: National Bureau of Economic Research, 2007)
- 32 Douglas N. Harris and Tim R. Sass, "The Effects of NBPTS-Certified Teachers on Student Achievement." Working Paper 4 (Washington: Center for Analysis of Longitudinal Data in Education Research, 2007). The researchers also found that the process of becoming nationally Board certified did not raise teachers' effectiveness either during or after certification. However, National-Board certified teachers who served in official mentor roles did raise their colleagues' effectiveness.
- 33 Kenneth D. Peterson, "Parent surveys for school teacher evaluation." *Journal of Personnel Evaluation in Education* 2 (3) (1989) 239-294, and Kenneth D. Peterson, Danielle Stevens, and Carol Mack, "Presenting Complex Teacher Evaluation Data: Advantages of Dossier Organizational Techniques over Portfolios." *Journal of Personnel Evaluation in Education*, 15 (2) (2001): 121-131.
- 34 Tucker and others, "The Efficacy of Portfolios."
- 35 Mark Wilson and others, "Using Test Scores as Evidence of External Validity for Indicators of Teacher Quality: Connecticut's Beginning Educator Support and Training Program" (Berkley: University of California, Berkley, Graduate School of Education, 2007).
- 36 Dan Goldhaber and Michael Hansen, "Assessing the Potential of Using Value-Added Estimates of Teacher Job Performance for Making Tenure Decisions" (Seattle: Center on Reinventing Public Education, 2008). Specifically, these researchers found that of teachers who were in the bottom quintile on average in math or reading in their first 2-3 yrs of teaching in North Carolina, 29 percent were in the two quintiles in reading and 24 percent were in the top two quintiles in math in years five and up (how many yrs?), on average.
- 37 John Papay, "Different Tests, Different Answers: The Stability of Teacher Value-Added Estimates across Outcome Measures" (Harvard Graduate School of Education, 2009).
- 38 Arthur Wise and others, "Teacher Evaluation: A Study of Effective Practices." *Elementary School Journal* 86 (1) (1985): 61-121, and personal communication from Lesley-Ann Gracey, Cincinnati Federation of Teachers, April 22, 2009.
- 39 Susan Moore Johnson, personal communication, May 16, 2009.
- 40 Kenneth D. Peterson and others, "Using More Data Sources to Evaluate Teachers." *Educational Leadership* 58 (5) (2001): 40-43.
- 41 Hazi and Rucinski, "Teacher Evaluation as a Policy Target."
- 42 The New Teacher Project, "Teacher Hiring, Assignment, and Transfer in Chicago Public Schools" (2007). 88 percent of Chicago public schools did not issue any unsatisfactory ratings between 2003-2006. Of 36,000 ratings issued by evaluators in that district between 2003-2006, only 0.3 percent were unsatisfactory.
- 43 Cadonna Peyton, "District Teachers Get Good Grades," *Riverside Press Enterprise*, July 26, 2005, p. A01.
- 44 Associated Press State and Local Wire, "Study: Weak Evaluations a Problem in Identifying Bad Teachers." December 9, 2005. 83 percent of Illinois districts did not rate any teachers unsatisfactory in this decade.
- 45 Betsy Hammond, "Teachers see little merit in this better performance, yes, but not tied to pay, as a ballot proposal would do," *The Sunday Oregonian*, October 5, 2008
- 46 For an example, see Hanushek and others, "The Market for Teacher Quality."
- 47 Principals *and* teachers believe that teachers are less effective than ratings would suggest. In Chicago, 56 percent of veteran principals said they had assigned a higher evaluation rating than the teacher's performance warranted. See The New Teacher Project, "Teacher Hiring, Assignment, and Transfer." On a recent survey of a random sample of principals nationwide, only 46 percent of all principals gave teachers in their school an overall rating of excellent. See Dana Markow and Suzanne Martin, "The MetLife Survey of the American Teacher, 2004-2005: Transitions and the Role of Supportive Relationships" (New York: MetLife, 2005).
- 48 As Miller and Chait discuss, not all states include incompetence or poor performance as cause for termination of tenured teachers. For an in-depth discussion of the effects of tenure policy on teacher dismissal, see Raegen Miller and Robin Chait, *Teacher Turnover, Tenure Policies, and the Distribution of Teacher Quality: Can High-Poverty Schools Catch a Break?* (Washington: Center for American Progress, 2008).
- 49 For an example, Julia E. Koppich and Connie Showalter, "Strategic Management of Human Capital: A Cross-Case Analysis of Five Districts" (Madison: Strategic Management of Human Capital, 2008).
- 50 The New Teacher Project, "Teacher Hiring, Assignment, and Transfer."
- 51 Brandt and others, "Examining District Guidance."
- 52 Koppich and Showalter, "Strategic Management of Human Capital". See also Brandt and others, "Examining District Guidance." They found a preponderance of what they called "vague terminology." Even when policies included specific terms, like "formative evaluation," the researchers assert that "it was often difficult to understand from the documents what the districts actually intended."
- 53 For example, to rate a teacher "satisfactory" who was previously rated "excellent," a principal must use a particular form and give this form to teacher 10 weeks prior to issuing the lower rating. See The New Teacher Project, "Teacher Hiring, Assignment, and Transfer."
- 54 Brandt and others, "Examining District Guidance."
- 55 On top of these responsibilities, 37.6 percent of principals in small towns teach classes. See Gregory A. Strizek and others, "Characteristics of Schools, Districts, Teachers, Principals, and School Libraries in the United States: 2003-04 Schools and Staffing Survey" (Washington: U.S. Department Education, 2007).
- 56 Richard R. Halverson and Matthew A. Clifford, "Evaluation in the Wild: A Distributed Cognition Perspective on Teacher Assessment." *Educational Administration Quarterly* 42 (4) (2006):578-619. This estimate was based on evaluations of 20 teachers each, six of whom received three evaluations and 14 of whom received one. It included time spent on writing evaluation reports for each teacher.
- 57 Brian Rowan, "Comparing Teachers' Work with Work in Other Occupations: Notes of the Professional Status of Teaching." *Educational Researcher*, 23 (6) (1994): 4-17, 21.
- 58 Strizek and others, "Characteristics of Schools, Districts, Teachers, Principals, and School Libraries in the United States."
- 59 For an example, see Stephen M. Kimball and Anthony Milanowski "Examining Teacher Evaluation Validity and Leadership Decision Making within a Standards-Based Evaluation System." *Educational Administration Quarterly* 45 (1) (2009): 34-70. These researchers found that administrators' evaluation training focused on managing the task rather than on evaluating accurately and providing good feedback.
- 60 For an example, see Kimball and Milanowski, "Examining Teacher Evaluation Validity and Leadership Decision Making within a Standards-Based Evaluation System."
- 61 Dan C. Lortie, *Schoolteacher: A Sociological Study* (Chicago: The University of Chicago Press, 1975). Also see Judith Warren Little, "The Persistence of Privacy: Autonomy and Initiative in Teachers' Professional Relations." *Teachers College Record* 91(4) (1990): 509-536, and Wise and others, "Teacher Evaluation."
- 62 Paul T. Hill, "The Costs of Collective Bargaining Agreements and Related District Policies" in *Jane Hannaway and Andrew J. Rotherham, eds. Collective Bargaining in Education* (Cambridge: Harvard Education Press, 2006).
- 63 Anthony T. Milanowski, "Split Roles in Performance Evaluation—A Field Study Involving New Teachers." *Journal of Personnel Evaluation* 18 (3) (2005): 153-169.

- 64 Shifts in the teaching force also call for evaluators to provide high-quality feedback through evaluation. 91 percent of principals surveyed nationwide had hired new teachers between 2000 and 2005. Markow and Martin. "The MetLife Survey."
- 65 Peyton, "District Teachers Get Good Grades."
- 66 In "Examining Teacher Evaluation," Kimball and Milanowski found that the written evaluations they reviewed contained little constructive criticism. Halverson et al. found that some principals spent almost half (48 percent) of post-conference time "checking in on personal or school activities/happenings" (p. 602).
- 67 Ann Duffett and others, "Waiting to Be Won Over Teachers Speak on the Profession, Unions, and Reform" (Washington: Education Sector, 2008). This study had a low response rate so findings should be considered with caution.
- 68 A recent MetLife survey of new teachers found that 89 percent of teachers surveyed strongly agreed with the statement "All children can learn." 92 percent of elementary principals and 82 percent of secondary principals strongly agreed with this statement. See Markow and Martin, "The MetLife Survey."
- 69 Kimball and Anthony Milanowski, "Examining Teacher Evaluation."
- 70 Thomas J. Kane and Douglas O. Staiger, "Using Imperfect Information to Identify Effective Teachers" (University of California Los Angeles School of Public Affairs, 2005).
- 71 Average number of teachers per school calculated based on figures in Strizek and others, "Characteristics of Schools, Districts, Teachers, Principals, and School Libraries in the United States."
- 72 Dale Mezzacappa and Connie Langland, "More Teachers Held Responsible for Students' Failures," *Passaic County Herald News*, March 24, 2005, p. B01, and Joe Smydo, "State School Boards Slow to Eliminate Bad Teaching," *Pittsburgh Post-Gazette*, February 14, 2007, p. ED-8.
- 73 Saoger and others, Identifying Effective Teachers Using Performance on the Job
- 74 Strizek and others, "Characteristics of Schools, Districts, Teachers, Principals, and School Libraries in the United States."
- 75 The New Teacher Project, "Teacher Hiring, Assignment, and Transfer,"
- 76 American Federation of Teachers and National Education Association, "Peer Assistance and Review: An AFT/NEA Handbook" (1998).
- 77 Duffett and others, "Waiting to Be Won Over," 16 percent say these teachers receive informal rewards, such as "better treatment," and 10 percent report that they receive a "token gift."
- 78 Susan Moore Johnson and others "Leading the Local: Union Presidents Chart Their Own Course." *Peabody Journal of Education*. (forthcoming).
- 79 William L. Boyd. "Public Education's Crisis of Performance and Legitimacy. in William L. Boyd and Debra Miretzsky, eds.: *American Educational Governance on Trial: Change and Challenges*. (Chicago: University of Chicago Press, 2003).
- 80 Koppich and Showalter. "Strategic Management of Human Capital." See also the other Strategic Management of Human Capital working papers available at <http://www.smhc-cpre.org/>.
- 81 Halverson and Clifford, "Evaluation in the Wild."
- 82 Susan Moore Johnson and others. "Leading the Local: Teachers Unions Presidents Speak on Change, Challenges" (Washington: Education Sector, 2007).
- 83 Ibid.
- 84 J. Goldstein, "Making Sense of Distributed Leadership: The Case of Peer Assistance and Review. *Educational Evaluation and Policy Analysis*, 25, 4, 397-421, and J. Goldstein, "Teachers at the Professional Threshold: Distributing Leadership Responsibility for Teacher Evaluation," unpublished dissertation, Stanford University, 2003. In the latter publication, Goldstein finds that teachers encountered substantial obstacles to evaluating accurately and moving to dismiss poorly performing teachers.
- 85 Ibid.
- 86 National Commission on Teaching and America's Future. "Learning Teams: Creating What's Next," (2009).
- 87 Rochkind and others, "Getting Started."
- 88 Alan Milanowski, "Performance Pay System Preferences of Students Preparing to Be Teachers." *Education Finance and Policy* 2 (2) (2007): 111-132.
- 89 Editors "Get bad teachers out of school; Cut 'em loose; our children deserve better." *Chicago Sun Times*, September 4, 2007, p. 27.
- 90 For example, Committee on Developments in the Science of Learning and Committee on Learning Research and Educational Practice, "How People Learn: Brain, Mind, Experience, and School, Expanded Edition" (Washington: National Academy Press, 2000); and M. Suzanne Donovan and James W. Pellegrino, eds. *Learning and Instruction: A SERP Research Agenda*. (Washington: The National Academies Press, 2003).
- 91 For a broad, initial outline of the concept of pedagogical content knowledge, which has been developed in different subject areas by subsequent work, see Lee Shulman, "Those who understand: Knowledge growth in teaching." *Educational Researcher*, 15 (2) (1986): 4-14, and Lee Shulman, "Knowledge and Teaching: Foundations of the New Reform." *Harvard Educational Review* 57 (1) (1987): 1-22.
- 92 David Kauffman and others, "Lost at Sea: New Teachers' Experiences with Curriculum and Assessment," *Teachers College Record* 104 (2) (2002), and Stacy Agee Szczesiul, "Initiatives and Initiative: Second-stage Teachers' Assessment of Autonomy," (Cambridge: Harvard University Graduate School of Education, 2007).
- 93 Based on a review of published documents describing TES and interviews with Lesleyann Gracey and Tim Kraus.
- 94 For a summary, see Heneman and others, "Standards-based teacher Evaluation."
- 95 On "step" years (16, 21, 26), teachers receive two annual evaluations (16, 21, 26),
- 96 Tucker and others, "The Efficacy of Portfolios."
- 97 Jacob and Lefgren found that principals rated male and non-tenured teachers' ability to raise achievement scores approximately .3 to .5 standard deviations lower than females and tenured teachers with same proficiency as measured by Value-added estimates. See Jacob and Lefgren, "Can Principals Identify Effective Teachers?"
- 98 For a discussion of performance-based principal compensation, which increasingly involves evaluating principals on the basis of student achievement see Dan Goldhaber. "Principal Compensation: More Research Needed on a Promising Reform" (Washington: Center for American Progress, 2007).
- 99 Herbert G. Heneman, III and Anthony T. Milanowski, "Assessing Human Resource Alignment: The Foundation for Building Total Teacher Quality Improvement." (Madison: Consortium for Policy Research in Education, 2007).

Acknowledgements

The Center for American Progress thanks The Bill & Melinda Gates Foundation for generously providing support for this paper.

The author would like to thank the following people for their insights on the issues discussed in this paper: Christopher Brandt, Kate Dickson, Benjamin Donaldson, Gordon Donaldson, Chad Ellis, H. Alix Gallagher, Lesley-Ann Gracey, Steven Kimball, Julia Koppich, Tim Krauss, Susan Moore Johnson, John Papay, Mark Simon, Gary Whiteley, and students in the Supervision for Instructional Improvement course at the University of Connecticut, Spring 2009. The author also thanks Raegen Miller and Robin Chait for excellent feedback on earlier drafts.

About the author

Morgaen L. Donaldson is an assistant professor of educational leadership at the University of Connecticut's Neag School of Education and a research affiliate at the university's Center for Education Policy Analysis. She is also a research associate at the Project on the Next Generation of Teachers at Harvard University. As a researcher she studies policies and practices related to teacher quality, teacher leadership, school and district human capital development, and teachers unions.

Donaldson, a former public high school teacher, was a founding faculty member of the Boston Arts Academy, Boston's public high school for the arts. She also served as a project manager in a Gates Foundation-funded initiative to replicate the practices of small schools that effectively served low-income and minority students.

Her recent publications include "Angling for Access, Bartering for Change: How Second-Stage Teachers Experience Differentiated Roles in Schools" in *Teachers College Record* (2008); "Leading the Local: Teachers Union Presidents Chart Their Own Course" in the *Peabody Journal of Education* (forthcoming, 2009); and "Into—and Out of—City Schools: The Retention of Teachers Prepared for Urban Settings" in *Equity and Excellence in Education* (forthcoming, 2009).

Donaldson received an Ed.D. in administration, planning, and social policy and an Ed.M. from the Harvard Graduate School of Education and an A.B. from Princeton University. She can be reached at: morgaen.donaldson@uconn.edu.

The Center for American Progress is a nonpartisan research and educational institute dedicated to promoting a strong, just and free America that ensures opportunity for all. We believe that Americans are bound together by a common commitment to these values and we aspire to ensure that our national policies reflect these values. We work to find progressive and pragmatic solutions to significant domestic and international problems and develop policy proposals that foster a government that is “of the people, by the people, and for the people.”

