reimagining accountability in K–12 education

Brian P. Gill, Jennifer S. Lerner, & Paul Meosky

abstract

Since the passage of the No Child Left Behind Act (NCLB) in 2002, American policymakers have relied primarily on outcome-based accountability in the form of high-stakes testing to improve public school performance. With NCLB supplanted in 2015 by the Every Student Succeeds Act—which gives states far greater discretion in the design of accountability systems—the time is ripe for policymakers to consider extensive behavioral science literature that shows outcome-based accountability is only one of multiple forms of accountability, each invoking distinct motivational mechanisms. We review rule-based, market-based, and professional accountability alongside outcome-based accountability, using evidence from the laboratory and the field to describe how each can produce favorable or unfavorable effects. We conclude that policymakers should (a) make greater use of professional accountability, which has historically been underutilized in education; (b) use transparency to promote professional accountability; and (c) use multiple, complementary forms of accountability, creating a complete system that encourages and supports the continuous improvement of educational practice.

Reimagining accountability in K–12 education

Brian P. Gill, Jennifer S. Lerner, & Paul Meosky

abstract. Since the passage of the No Child Left Behind Act (NCLB) in 2002, American policymakers have relied primarily on outcome-based accountability in the form of high-stakes testing to improve public school performance. With NCLB supplanted in 2015 by the Every Student Succeeds Act—which gives states far greater discretion in the design of accountability systems—the time is ripe for policymakers to consider extensive behavioral science literature that shows outcome-based accountability is only one of multiple forms of accountability, each invoking distinct motivational mechanisms. We review rule-based, market-based, and professional accountability alongside outcome-based accountability, using evidence from the laboratory and the field to describe how each can produce favorable or unfavorable effects. We conclude that policymakers should (a) make greater use of professional accountability, which has historically been underutilized in education; (b) use transparency to promote professional accountability; and (c) use multiple, complementary forms of accountability, creating a complete system that encourages and supports the continuous improvement of educational practice.

The Equity Project (TEP) Charter School is a public middle school in the Washington Heights neighborhood of New York City, where, since opening in 2009, it has served a student population that is almost exclusively low income and Hispanic or African American. TEP Charter School's founding principal, Zeke Vanderhoek a former teacher and entrepreneur who had read the growing body of research indicating that teachers are the most important school-based influence on students, decided to design a school that would focus virtually all of its resources on hiring and developing the best possible teachers. Vanderhoek ran the numbers and concluded that the standard per-pupil public funding available to New York City charter schools could be reallocated (for example, by eliminating administrative positions and increasing the size of a typical class from 27 students to 31) in a way that would allow him to pay each of his teachers $125,000 per year, plus bonuses based on school-wide student achievement. The plan worked. In its first 4 years of

operation, TEP Charter School produced substantial positive effects on student achievement: By the time they finished eighth grade, TEP Charter School's initial cohort of fifth graders had test scores showing an advantage equivalent to several months of additional instruction in English and science and more than a year and a half of additional instruction in math, relative to a similar comparison group.[1]

The Equity Project Charter School: A Case Study in Multi-Mode Accountability

Although media coverage of TEP Charter School focused on the eye-popping teacher salaries—which enabled TEP Charter School to hire, among other star educators, a physical education teacher who was previously a trainer for the Los Angeles Lakers—the school's approach involves much more than high salaries and performance-based bonuses. Those who seek to teach at TEP Charter School are rigorously screened in a process that requires applicants to spend a full day teaching TEP Charter School's students. Once hired, TEP Charter School teachers engage in 6 weeks of professional development annually. In addition, throughout the school year, they observe each other in the classroom at least twice a week, providing written feedback to their colleagues. TEP Charter School does not offer tenure, and teachers' contracts are renewed on the basis of their performance, which is evaluated primarily through classroom observation. In short, TEP Charter School staff are accountable for their students' achievement, but they are also accountable for their professional practice, which is observed not only by the principal but also by their colleagues, who are expected to help them improve their practice. In addition, as a charter school, TEP Charter School is implicitly accountable to the parents of its students, because its funding depends on persuading parents to enroll their children.

Accountability as High-Stakes Testing

Since the turn of the millennium, American policymakers have tended to understand accountability in education in narrow terms. The No Child Left Behind Act (NCLB) required reporting of every public school's level of student proficiency in math and reading and imposed sanctions for failure to achieve targets. Schools that fell short of state-determined proficiency goals faced mandatory interventions, which became increasingly aggressive after repeated failures. Chronically failing schools could be restaffed, taken over by the state, or shut down. *Accountability* came to be synonymous with high-stakes testing. The Obama administration went even further in promoting high-stakes testing, pushing states to include student achievement growth measures in evaluating teachers and principals.

High-stakes testing has come under attack on multiple fronts. Teachers unions objected to the use of test scores in teacher evaluations; high-profile cheating scandals ruined the careers of prominent educators; parents increasingly demanded the right to opt out of standardized testing of their children. Meanwhile, education scholars proposed replacing existing accountability regimes.[2–4] Recognizing the increasing unpopularity of the federal mandates for high-stakes testing, in December 2015, Congress passed the Every Student Succeeds Act, which replaces NCLB and gives states far more discretion to design their own accountability policies.

Opportunity to Craft Better Accountability Systems

With new latitude to refine and improve kindergarten through 12th grade (K–12) accountability policies, policymakers should heed the empirical evidence on accountability's effects—and not only the evidence from education itself. Extensive literature in experimental social psychology and behavioral economics identifies different types of accountability mechanisms and the conditions under which accountability improves outcomes.[5] Just as important, it also identifies conditions in which accountability may have no effect or even make matters worse. This behavioral science literature can provide valuable insights for the design of school accountability systems, but it has been largely overlooked in the education policy debate.[6] (The National Academies report cited here examined some of the behavioral literature related to the use of test-based incentives, but it did not examine other forms of accountability.)

> *Accountability comes in many forms, involving different mechanisms and different behavioral responses*

Jennifer S. Lerner of Harvard University and Philip Tetlock, of the University of Pennsylvania, conducted a comprehensive review of the behavioral science literature on accountability;[5] we use that review to provide a framework for our discussion here. Lerner and Tetlock's review made two broad points that motivate our discussion. First, they noted that accountability comes in many forms, involving different mechanisms and different behavioral responses. Outcome-based accountability is only one of these forms. In the context of schooling, this means that high-stakes testing is not the only tool available for improving educational performance. Second, Lerner and Tetlock emphasized that accountability can have positive or negative effects, depending on the accountability type, the decision context, and the nature of the task. This suggests that policymakers need to seriously consider the trade-offs associated with different types of accountability and explore ways to mitigate unintended effects.

Four Types of Accountability

In this article, we explore the behavioral science literature alongside the education literature on accountability. We discuss evidence related to four types of accountability applicable to education policy: **rule-based, market-based, outcome-based,** and **professional** accountability. We also propose how policymakers might use these different types of accountability to promote continuous improvement in schools. Schooling has multiple purposes with multiple constituencies, including parents, students, and the public. Any single form of accountability is unlikely to fully serve these multiple purposes and constituencies.

We reach several conclusions relevant to policy and practice. First, transparency alone can create accountability, even in the absence of explicit rewards and sanctions. Second, multiple forms of accountability can be used concurrently to take advantage of complementary strengths and weaknesses. Third, to promote continuous improvement in schools, a comprehensive accountability system should include mechanisms for the improvement of practice. Collectively, these three points suggest greater use of professional accountability than has historically been the norm in K–12 education, alongside rule-based, market-based, and outcome-based accountability.

Accountability Mechanisms and Types

Four Behavioral Mechanisms for Achieving Accountability

Lerner and Tetlock's review[5] identified four mechanisms that make people feel accountable: (a) the **mere presence of another**—simple awareness that someone else is watching, (b) **identifiability**—the expectation that an action or outcome will be attributable to oneself, (c) **reason giving**—the expectation that one will need to explain or rationalize one's actions, and (d) **evaluation**—the expectation that consequences will depend on an assessment of one's performance.

The outcome-based (high-stakes testing) accountability inaugurated by NCLB incorporates **identifiability and evaluation** but gives short shrift to **reason giving and the mere presence of another** as less aggressive ways to foster accountability. In particular, policies and practices that increase transparency, making the relevant activity more visible to others, may evoke any or all of the first three accountability mechanisms even without imposing formal consequences. TEP Charter School's expectation that teachers observe each other routinely is one example of this; later, we describe other examples.

Accountability as Applied in Other Professions

Other professions—from law to engineering to architecture to medicine—typically use multiple forms of accountability that collectively make use of all four of the behavioral accountability mechanisms identified by Lerner and Tetlock.[5]

Many professions rely on **rule-based accountability,** which sets rules that delineate mandated or forbidden activities. Rule-based accountability relies on the identifiability of actors and sometimes includes an evaluation component. Rule-based accountability is common in K–12 education: State education codes, regulations, and teacher contracts create rule-based accountability.

Market-based accountability is also pervasive in most fields. Dentists and engineers, for example, are

accountable to their customers, via their consumer choice, in a way that public school educators typically are not. Under market-based accountability, clients or customers can hold professionals responsible by choosing among providers. Market-based accountability applies the identifiability mechanism, and it encourages providers to describe and explain their services, thereby invoking reason giving as well. Market-based accountability is relevant in private schools; in public charter schools; and, in a limited way, in conventional public schools, to the extent that families have the ability to move to a desirable school district.

In many fields, **professional accountability** systems go beyond what markets and government regulations require and often evoke all four behavioral mechanisms of accountability. Professional organizations impose standards for entry, provide resources for continuing learning, and set standards of practice that may be enforced by direct observation—for example, medical residents who are closely supervised by attending physicians. In contrast, although K–12 education includes some forms of professional accountability—for example, certification requirements—this sector historically has held only modest requirements for entering the profession and minimal requirements for maintaining professional status.

The kind of **outcome-based accountability** that has been the primary focus of education policy over the last two decades has not generally been common in other fields, perhaps because market-based accountability serves the same function. Tort law, in which plaintiffs'

attorneys are paid only if they win, is a notable exception. Outcome-based accountability is increasingly being attempted in health care,[7] for example, with financial penalties for high readmission rates of patients after hospital discharge.

In sum, many forms of accountability are used in various fields, and the different forms of accountability evoke different psychological mechanisms. Table 1 provides an overview of the different mechanisms evoked by different accountability types, with examples (outside of K–12 education) in each relevant cell.

These examples from other fields highlight the fact that policymakers have more tools available than just outcome-based accountability (high-stakes testing) to enhance school performance. Indeed, outcome-based accountability systems ignore two of the four behavioral mechanisms that promote accountability: mere presence of another and reason giving. In the rest of this article, we consider applications of all of these accountability types, using research from the field and the laboratory to inform ways that K–12 accountability regimes might be designed to improve educational outcomes.

Outcome-Based Accountability

Twenty-five years ago, outcome-based accountability was almost unknown in K–12 schooling. The education standards movement that gained steam during the 1990s promoted clear performance expectations for each grade level and tests to measure students' proficiency. Beginning in 1994, federal law (the Improving

Table 1. Accountability types in policy and psychological accountability mechanisms

Psychological accountability mechanism	Accountability types in policy			
	Outcome based	Rule based	Market based	Professional
Mere presence of another				Surgical operating room with nurses in attendance
Identifiability	*Consumer Reports,* Zagat	Minimum certification requirements (various professions)	Branding	Membership in a professional organization
Reason giving			Annual report to company stockholders	Medical rounds with explanation of treatment
Evaluation	Contingent fees for attorneys	Driver licensing test		Bar exam

America's Schools Act—the predecessor of NCLB) required states to set proficiency standards, assess students in multiple grades, and report school-specific results. The primary behavioral mechanism used was identifiability, in the form of public reporting of results. Eight years later, NCLB added explicit sanctions to schools falling short of proficiency targets, which added evaluation as a behavioral mechanism. More recently, the federal government pushed states to extend outcome-based accountability from schools to individual educators by evaluating teachers and principals in part on the basis of their student's achievement growth; this growth, or lack thereof, is measured yearly for each student, and then averaged across all students served by a particular teacher or principal.

These initiatives were not informed by the behavioral science literature, which finds few positive effects for outcome-based accountability.[5] Because it does not constrain decisionmakers with rules, outcome-based accountability might be more effective than other forms of accountability at promoting innovation,[8] but this has not been extensively studied. The effectiveness of outcome-based accountability can be undermined by the sunk-cost bias, which makes decisionmakers more likely to pursue action because of prior investments, even when the odds of success are low.[9] In addition, tangible rewards sometimes undermine intrinsic motivation.[10]

Behavioral studies also find that outcome-based accountability may impair decisionmaking by eliciting stress and negative emotions, increasing a decision's difficulty.[11,12] Perhaps because stress burdens cognition,[13,14] outcome accountability sometimes fails to increase the use of strategies that require substantial effort.[5]

Additionally, accountability regimes can be counterproductive when they are viewed as illegitimate.[5] Many teachers are suspicious of *value-added models* (VAMs) that aim to measure their contributions to student achievement. Improving on cruder outcome-based accountability regimes that rely on student achievement levels, VAMs account statistically for factors outside of the teacher's control, including students' demographic characteristics and (most important) prior achievement. In essence, VAMs measure how much better or worse a teacher's students are doing relative to how well the same students would have done if taught by an average teacher. Although well-designed VAMs can produce

unbiased (fair) measures of teachers' contributions to student achievement growth,[15,16] suspicions on the part of practitioners could undermine their ability to promote performance improvements.

Moreover, even fair measures of educators' contributions to student achievement can be problematic in a high-stakes accountability system. Tests cannot capture all of the skills and knowledge that schools seek to impart. Some evidence suggests that instructional practices that raise test scores differ from those that promote students' effort and long-term goals.[17] High-stakes testing encourages "teaching to the test": Studies have found that many schools have narrowed their curriculums, focusing on reading and math to the exclusion of other subjects,[18] and spend a growing proportion of class time specifically preparing for the tests.[19,20] In extreme cases, educators have been caught cheating. Teacher-developed *student learning objectives,* increasingly used as outcome-based accountability measures, may be especially susceptible to inflation, because teachers grade themselves.[21] In short, test-based accountability may have the paradoxical effect of undermining the validity of the test itself.[22]

Even so, despite these unintended effects of outcome-based accountability, most existing field studies of the impact of high-stakes testing suggest positive effects in at least some schools, grades, and subjects.[23–27] (In the last cited reference, Deming et al. found mixed results in different schools.) The effects of performance-pay incentives for teachers on student achievement have been mixed, ranging from no effects to small positive effects.[28–30]

Rule-Based Accountability

Rule-based accountability relies on identifiability and sometimes on evaluation. Historically, rule-based accountability has been used to set constraints and conditions, such as in state lists of approved textbooks; contractual rules about working conditions, hours, and class size; and federal and state spending regulations. Through most of the past century, teachers had wide discretion about instruction,[31] which surely is one source of the wide variation in effectiveness in promoting student achievement.[32] Rules and protocols may have ensured minimum standards and reduced the most egregious inequities, but they may have also reduced opportunities for innovation. Charter schools were

created in part to allow innovation that would be less constrained by traditional rules, which may be particularly restrictive in the context of new instructional technologies that permit educators to organize schools and classrooms in novel ways.

Recently, some districts and school management organizations have become more directive about elements of instruction and school operations, pursuing a maximal version of rule-based accountability in which all teachers of the same courses may be expected to cover the same material at the same pace.[33] Principals have been asked to take on greater responsibilities as instructional leaders. Pacing guides are commonly used, and instruction is tied to state standards, with some lesson plans scripted to the minute.[34]

Limited evidence supports a maximal rule-based approach in schools. Scripted direct instruction has been found to promote student achievement in elementary grades.[35] But maximal rule-based accountability could become counterproductive, because psychological studies find that close monitoring often reduces intrinsic motivation.[36] The behavioral science literature also shows that intense monitoring can exacerbate the sunk-cost bias,[37] undermine innovation, and entrench suboptimal practices.[38] The perception of rules as illegitimate can produce a boomerang effect, leading people to react against the rules.[39] It is easy to imagine all of these effects operating in the classroom context, potentially undermining teaching and learning. Indeed, aggressive rule-based accountability may be especially unsuited to teaching, because it is an inherently complex task that requires daily adjustments and judgments. Highly scripted instructional programs might be useful in ensuring a minimal level of acceptable practice, but they are unlikely to promote excellence in teaching, particularly for deep and complex curricular material.

Market-Based Accountability

Market-based accountability is based on classical economic principles rather than newer behavioral approaches. It involves the identifiability and reason-giving mechanisms: Where choice is an option, schools chosen by families must be identifiable and attractive to parents. Historically, market-based accountability did not play a substantial role in U.S. public education. Operating alongside tuition-based private schools, the public system has been based on the *common school model,* which assumes that each community will educate its children together, with school districts maintaining local monopolies on publicly supported education.[40,41]

Policymakers have shown increasing interest in incorporating market-based accountability into education, reasoning that local monopolies controlled by school boards may not produce the best schools[42,43] and that giving families choice in schooling is inherently valuable.[44] *Vouchers*—scholarships for tuition at private schools—have been advocated by conservatives (and a few liberals)[44] ever since Milton Friedman proposed the idea over 60 years ago,[42] and publicly funded voucher programs have been established in several states. Charter schools—publicly funded schools of choice that are open to all students, do not charge tuition, and operate outside of the direct control of local school districts[45]—represent a newer market-based approach and a new kind of public school, and they have received support across the political spectrum (as well as opposition from teachers unions and supporters of conventional public schools). The first charter schools opened only a quarter of a century ago; today, over 6,000 operate in more than 40 states.

Empirically, the evidence on the effects of market-based schools on student test scores and longer term educational attainment, although not definitive, suggests that the concept holds promise. In some contexts and locations, charter schools are producing substantial positive effects,[46–49] but their performance varies widely.[50,51] A few studies of small-scale voucher programs have found positive educational impacts, particularly for African-American students.[52,53] Louisiana's statewide voucher program for students in low-performing public schools, in contrast, has been shown to have substantial negative effects on student achievement in its first years of operation.[54,55]

K–12 schooling differs from other services in ways that might make exclusive reliance on the market suboptimal. First, the classic principal–agent problem—aligning the interests of clients and agents (educators)—is complicated by the involvement of multiple clients (students, parents, and the public), whose interests may not be fully aligned. Second, children are not fully capable of assessing their own best interests. In addition, students' educational experiences are affected not

only by school quality but by externalities, including characteristics of other students.[56] As a consequence, an unfettered market may produce segregated schools, as parents with high levels of knowledge, wealth, or motivation seek out schools that educate children from families like their own.

Relatedly, skeptics worry that market-based schools will drain conventional public schools of funding and motivated families. Supporters argue that breaking the local monopoly (that is, giving families publicly funded options other than those offered by the local school district) produces healthy competitive pressure that will benefit all students. The research base provides very little evidence that reallocation of students and resources produces any harm for students who remain in conventional public schools. Only one study has found a negative effect of charter schools on student achievement in nearby district-operated schools,[57] several studies have found no effects on students in nearby schools,[58–60] and a few studies have found positive effects.[60–62]

Another externality relates to the original rationale for public education: Society benefits from the inculcation of the knowledge, skills, and attitudes necessary for effective citizenship. Historically, this key rationale for the common school model implicitly assumed that effective education of citizens required public operation of schools.[39,40] The fact that the education of citizens is a public good argues against relying entirely on market mechanisms.

In fact, existing school choice programs rarely rely exclusively on market accountability. Charter schools are exempt from some forms of regulation but, like conventional public schools, are subject to rule-based and outcome-based accountability. Charter schools cannot charge tuition, their students must take the same high-stakes tests as students in conventional public schools, and these schools (typically) must admit all applicants, as space allows. Moreover, charter schools operate under the supervision of publicly empowered authorizers. Even the private schools that participate in voucher programs typically must submit to some regulation to receive public funds. Milwaukee's program, the longest-operating publicly funded voucher program in the country, imposes requirements for instructional time, forbids tuition, requires state assessments, and does not allow selective admissions.

Professional Accountability

Prominent voices are calling for greater professional accountability that would give K–12 teachers support, opportunities for collaboration, and training while also setting higher expectations.[2,3] Professional accountability in education can take many forms, involving all four of the behavioral accountability mechanisms. For example, licensing and professional reviews involve evaluation; observations and assistance by supervisors, instructional coaches, peers, or mentors involve identifiability, reason giving, and sometimes evaluation; collaboration and coteaching involve the mere presence of another and reason giving.

Traditional and Novel Versions of Professional Accountability

States have long applied requirements for teaching licenses, including coursework, student teaching, and passing exams. Teacher contracts generally reward master's degrees and experience as proxies for professional skill. But master's degrees have little or no relation to improved student achievement,[63,64] and most studies find professional development has no effect on student achievement.[65,66] Tellingly, traditional teacher evaluations have typically concluded that 98% to 99% of teachers are satisfactory, with tiny percentages falling short of satisfactory and no one better than satisfactory, because there was typically no rating category available to identify exemplary teachers.[67] Meanwhile, state laws and teacher contracts that award permanent tenure insulate teachers from professional accountability (as well as from other forms of accountability).

More robust and ambitious forms of professional accountability may hold more promise. Licensing and professional requirements at a high enough level—such as the certification process of the National Board for Professional Teaching Standards—might help identify especially effective teachers.[68] Because teacher quality is the most influential school-controlled factor in student achievement growth,[32] several initiatives of the Obama administration have promoted increased rigor in teacher evaluation, encouraging, for example, the use of multiple performance measures and multiple rating categories that include not only the traditional categories of *satisfactory* and *unsatisfactory* but also categories that

recognize high-performing teachers. States and school districts are adopting extensive rubrics for the observation and rating of teaching practice. There is a risk that the new systems could deteriorate into compliance exercises that resemble traditional rule-based accountability mechanisms, but if they function as intended, they could substantially improve practice. Robust professional accountability systems—unlike outcome-based, rule-based, and market-based accountability—include tools and resources to help teachers improve their skills. If taken beyond screening and compensation reform, they have the unique advantage of coupling accountability with support for improvement.

Novel forms of professional accountability might include new job descriptions and training methods. Some school districts have recently created teacher residency programs modeled on medical residencies, in which aspiring teachers spend much more time in the classroom during their training. Other districts and schools are giving teachers leadership opportunities, such as serving as instructional coaches who help their colleagues develop teaching skills.[69,70]

In most professions, professional accountability includes being answerable to clients.[71] K–12 schooling, in contrast, traditionally involves little direct accountability of educators to students. Nonetheless, school districts such as the Pittsburgh Public Schools are now including student surveys in new teacher evaluation systems. Some are using teacher surveys in principal evaluation,[72] applying the business world's 360-degree feedback to the academic workplace.

Another professional accountability system is an intensive review of school quality conducted by independent, expert educators, as is common in British schools. A school quality review involves an extended visit by outside experts who observe instruction; interview teachers, students, and parents; and examine school performance data. The review concludes with clear recommendations for improvement. Scholars such as Marc Tucker of the National Center on Education and the Economy,[2] Linda Darling-Hammond of Stanford,[3] and Helen Ladd of Duke[73] have proposed that school quality reviews be included as part of reformed accountability systems in the United States.

Professional accountability would make teaching more transparent, potentially activating all four behavioral accountability mechanisms. Indeed, rich professional accountability systems emphatically reject allowing teachers complete discretion in the classroom, under the assumption that there are standards of practice to which teachers should be held. Using transparency to promote professional accountability is a significant departure from an older, rule-based approach that values teacher autonomy over accountability. The old approach is regrettably evident in the rules of the Chicago Public Schools, which explicitly prohibit the use of classroom video recordings for evaluating teachers[74]—in sharp contrast to the expectations and transparent culture of the TEP Charter School where teachers observe each other every week. Schools like the Kauffman Charter School in Kansas City have gone one step further, literally making teaching transparent by giving classrooms interior windows that make them visible to adults in the hallways.

Greater teaching transparency is common in some countries that consistently outperform the United States in international comparisons of student achievement. A recent international study of educational practice found that although responding American teachers were more likely than their international peers to receive feedback from principals, only 11% received feedback from mentors, versus 39% in Japan, 38% in Singapore, and 24% in Australia.[75] Further, only 27% of responding American teachers received feedback from colleagues, versus 84% in Korea, 57% in the Netherlands, and 43% in Finland. All of those countries outperformed the United States in math, reading, and science in the most recent study of the Program for International Student Assessment.[76] American teachers were also far less likely than their counterparts in other countries to receive feedback from student surveys and less likely to report that the feedback they received led to public recognition, career advancement, or increased compensation.

Indeed, one study in an American urban school district found that improvements in student achievement were associated with teams of teachers who had strong mutual professional ties and with individual teachers who had strong ties with their principals.[77] Professional accountability could promote ties among teachers, potentially developing the social capital and trust that have been found to be markers of effective schools.[78]

Behavioral Evidence on Professional Accountability

Many studies in psychology demonstrate the favorable effects of requiring people to justify their decisions to

others, a common expectation of professional accountability systems. One study found that requiring subjects to justify their decisions encourages high-effort strategies that are sensitive to evidence that can inform a decision, alleviating mistakes and inconsistencies.[79] Similarly, another study found that the need to justify decisions stimulated systematic thinking and attention to evidence.[80] In a third study, a justification requirement reduced reliance on stereotypic thinking.[81] Subjects who had to justify their judgments have also been found to be less likely to overattribute responsibility to individuals rather than situations.[82]

Other studies suggest that an increased sense of control—which might be promoted by a professional accountability environment that promotes initiative—may improve performance on attention-demanding tasks, promote more considerate decisionmaking, and assist memory formation.[83,84] More generally, professional accountability may best encourage the systematic, effortful, and self-critical thinking associated with even-handed, accurate reasoning.[85]

Professional accountability is also compatible with the behavioral nudges that are increasingly being adopted in various areas of public policy.[86] Field trials have demonstrated, for example, that appealing to social norms (using a wording like, "most people like you do X," following the work of Robert Cialdini)[87] powerfully influences behavior in contexts ranging from collecting taxes[88] to motivating parents to keep their children in school.[89] This suggests the possibility that providing relevant, appropriate evaluation feedback to teachers could lead to improvement even in the absence of explicit consequences.

Even though professional accountability is compatible with low-cost behavioral nudges, many forms of professional accountability are expensive or make substantial demands on educators. Teachers require time to observe each other and provide feedback. Instructional coaches need to be hired. School quality reviews must be staffed. More research is needed to assess whether some forms of professional accountability are more cost-effective than others.

Field Evidence on Professional Accountability

A few studies have examined new forms of professional accountability for educators. Teacher residencies are showing promise in producing high-performing teachers and keeping them in the classroom,[90] and early evidence on the effects of instructional coaching on student achievement is encouraging.[91–93] Several recently developed rubrics for observing and evaluating instructional practice have produced evaluation ratings that are correlated with teachers' contributions to student achievement,[94–96] suggesting that careful observation can produce feedback that could improve student outcomes.

In higher education, student evaluations of teachers have had positive effects on teaching.[97] Recent studies examining student surveys in K–12 schools have found the results are (modestly) correlated with measures of teachers' contributions to student achievement,[94,95] suggesting that they hold promise.

An intensive, peer-based teacher evaluation system used in Cincinnati offers encouraging evidence on formal, job-embedded professional accountability. Participating teachers substantially increased their effectiveness in raising student achievement during and after the year they were evaluated by peers—even though the evaluation criteria were based entirely on professional practice, not on test results.[98] Thomas Dee of Stanford University and James Wyckoff of the University of Virginia found that Washington, DC's ambitious teacher evaluation system—which uses multiple classroom observations by multiple observers, as well as measures of teachers' impacts on their students' achievement growth—produced positive effects on student achievement.[99] And one experimental study found that nudging school principals with information about teachers' effectiveness (their prior contributions to student achievement) raised test scores and increased the attrition rate of low-performing teachers, even without incorporating the information in a formal, high-stakes evaluation measure.[100]

Increasing Professional Accountability and Transparency Using a Multimode Approach

The outcome-based accountability that has been the focus of policymakers' attention has produced some positive results, but relying on it exclusively is unlikely to produce large, sustained improvements and can lead to unintended and undesirable side effects. The evidence from behavioral science laboratories and from the field makes clear that other accountability approaches can also produce favorable results, suggesting that a

narrow focus on outcome-based accountability leaves important tools unused.

Below is a modified version of the table from the beginning of the article. As in the original table, in Table 2 we provide examples illustrating how different accountability types can evoke different psychological accountability mechanisms. Here the examples are specific to K–12 education, confirming that policy-makers and educators have a wide range of account-ability tools to use, just as they do in other professions and fields.

After reviewing the evidence from behavioral science and the field literature, we propose that policy-makers designing K–12 accountability systems should consider three key points. *First, policymakers should make greater use of professional accountability, which has historically been underutilized in education.* As Table 2 indicates, different professional approaches can invoke all four of the motivational mechanisms inherent in different types of accountability. In the past, K–12 schooling has used professional accountability in limited ways, largely related to establishing minimum standards for entry or promotion. Forms of professional accountability that are more robust and ambitious could set higher expectations for professional practice and simultaneously create feedback mechanisms that help educators improve their practice.

Second, transparency alone can create professional accountability. All of the examples of professional accountability included in Table 2 involve making educational practice more transparent to other educators. Transparency of practice activates several behavioral accountability mechanisms (mere presence of another, identifiability, and reason giving) that power-fully influence behavior. Transparency also provides an opportunity to offer feedback to improve performance. Professional accountability can involve trans-parency in various ways, including not only the literal transparency of glass-walled classrooms but also peer observation and evaluation, instructional coaching, and 360-type feedback.

Third, multiple forms of accountability can be used in complementary ways, creating a complete system that encourages and supports the continuous improvement of educational practice and outcomes. Using multiple approaches can play to the advantages of each type while minimizing disadvantages—much as TEP Charter School uses outcome-based accountability (in the form of bonuses for school-wide performance) and market-based accountability (it must attract students to survive) alongside a rich professional accountability system. For example, Ken Frank of Michigan State University has proposed a form of school governance that would utilize multiple modes of accountability, making the principal accountable to a community board and giving the principal greater authority to remove the lowest performing teachers while at the same time giving teachers more say in school operations, including the evaluation of the principal.[101] Shefali Patil of the Univer-sity of Texas, Ferdinand Vieider of Reading University (UK), and Philip Tetlock of the University of Pennsylvania recently noted that outcome-based accountability

Table 2. Accountability types and psychological accountability mechanisms with applications in kindergarten through 12th grade schooling

Psychological accountability mechanisms	Accountability types			
	Outcome based	Rule based	Market based	Professional
Mere presence of another				Classroom windows
Identifiability	Public reporting of school-wide test results	Minimum certification requirements	School choice	Peer observation
Reason giving			Charter-school enrollment fairs	Instructional coaching
Evaluation	Value-added model incorporated in teacher evaluations	Formal observation by principal		Peer review; advanced certification

may better promote innovation, but process-based accountability (including forms of professional accountability) may better promote the use of best practices.[38] They suggested that the disadvantages of both types might be counteracted by systems that promote the empowerment of decisionmakers to rethink ineffective practices, encourage focus on outcomes, and facilitate organizational learning.[102] This kind of empowerment is implicit in professional accountability, and it can also be incorporated into an outcome-based accountability system that communicates a desire to achieve shared objectives. Similarly, according to a National Academies report on high-stakes testing,[6] external rewards are most likely to be effective when they are closely aligned with educators' intrinsic aims, promoting "autonomous motivation."

Moreover, organizational learning requires feedback for improvement.[102] Outcome-based accountability and market-based accountability can create incentives for improved performance, but they provide no information or resources to help professionals actually improve their performance. In a more integrated approach, transparency of practice creates opportunities for educators to improve, rich data on student outcomes help diagnose students' needs, and rewards for success encourage educators to innovate in productive ways.

Much of the backlash against high-stakes testing has come in the context of teacher evaluations that include student achievement growth as a formal component. There are good reasons to avoid relying exclusively on test scores for accountability. But the last decade has seen an enormous amount of work on other elements of the evaluation system that are related to professional accountability, including higher-quality classroom observations, an increase in the number of rating categories to differentiate teachers at the high end of the performance distribution as well as the low end, and the inclusion of student feedback. Efforts to roll back the use of test scores in educator evaluation risk undermining these fledgling efforts to promote increased professional accountability. In Los Angeles, for example, the district and teachers union recently agreed to new evaluation policies that eliminated the highest rating category, with the result that it is impossible for a teacher to exceed expectations.[103] Refusing to recognize exemplary performance among teachers is not only contrary to the evidence about variation in teacher effectiveness but also incompatible with the goal of creating a system of continuous improvement in professional practice.

As policymakers and educators take advantage of the Every Student Succeeds Act's new flexibility, we hope they recognize that reducing a near-exclusive reliance on outcome-based accountability does not have to mean reducing accountability as a whole; that a wide range of tools are available for creating a richer accountability system that can promote continuous improvement; and that professional accountability should play an important role in that system, raising expectations for teachers and schools while providing better opportunities to meet those raised expectations.

author affiliation

Gill, Mathematica Policy Research; Lerner, Harvard Kennedy School, Harvard University; Meosky, Harvard Kennedy School, Harvard University. Corresponding author's e-mail address: bgill@mathematica-mpr.com

References

1. Furgeson, J., McCullough, M., Wolfendale, C., & Gill, B. (2014). *The Equity Project Charter School: Impacts on student achievement*. Cambridge, MA: Mathematica Policy Research.
2. Tucker, M. S. (2014). *Fixing our national accountability system*. Washington, DC: National Center on Education and the Economy.
3. Darling-Hammond, L., Wilhoit, G., & Pittenger, L. (2014). Accountability for college and career readiness: Developing a new paradigm. *Education Policy Analysis Archives, 22*(86).
4. Center on Reinventing Public Education. (2014). *Designing the next generation of state education accountability systems: Results of a working meeting*. Retrieved from http://www.crpe.org/sites/default/files/CRPE_designing-next-gen-state-ed-accountability.pdf
5. Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin, 125,* 255–275.
6. Hout, M., & Elliott, S. W. (Eds.). (2011). *Incentives and test-based accountability in education*. Washington, DC: National Academies Press.
7. Gold, M. (2010). *Accountable care organizations: Will they deliver?* Princeton, NJ: Mathematica Policy Research.

8. Patil, S. V., & Tetlock, P. E. (2014). Punctuated incongruity: A new approach to managing trade-offs between conformity and deviation. *Research in Organizational Behavior, 34,* 155–171.

9. Simonson, I., & Staw, B. M. (1992). De-escalation strategies: A comparison of techniques for reducing commitment to losing courses of action. *Journal of Applied Psychology, 77,* 419–426.

10. Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin, 125,* 627–688.

11. Zhang, Y., & Mittal, V. (2005). Decision difficulty: Effects of procedural and outcome accountability. *Journal of Consumer Research, 32,* 465–472.

12. Siegel-Jacobs, K., & Yates, J. F. (1996). Effects of procedural and outcome accountability on judgment quality. *Organizational Behavior and Human Decision Processes, 1,* 1–17.

13. Mendl, M. (1999). Performing under pressure: Stress and cognitive function. *Applied Animal Behaviour Science, 65,* 221–244.

14. Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping.* New York, NY: Springer.

15. Kane, T. J., McCaffrey, D. F., Miller, T., & Staiger, D. O. (2013). *Have we identified effective teachers? Validating measures of effective teaching using random assignment.* Seattle, WA: Bill & Melinda Gates Foundation.

16. Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review, 104,* 2593–2632.

17. Ferguson, R. F. (with Danielson, C.). (2014). How Framework for Teaching and Tripod 7Cs evidence distinguish key components of effective teaching. In T. J. Kane, K. A. Kerr, & R. C. Pianta (Eds.), *Designing teacher evaluation systems: New guidance from the Measures of Effective Teaching project* (pp. 98–143). Hoboken, NJ: Jossey-Bass.

18. Dee, T. S., Jacob, B., & Schwartz, N. L. (2013). The effect of NCLB on school resources and practices. *Educational Evaluation and Policy Analysis, 35,* 252–279.

19. Hamilton, L. S., Stecher, B. M., & Yuan, K. (2012). Standards-based accountability in the United States: Lessons learned and future directions. *Education Inquiry, 3,* 149–170.

20. Jennings, J., & Sohn, H. (2014). Measure for measure: How proficiency-based accountability systems affect inequality in academic achievement. *Sociology of Education, 87,* 125–141.

21. Gill, B., English, B., Furgeson, J., & McCullough, M. (2014). *Alternative student growth measures for teacher evaluation: Profiles of early-adopting districts.* Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic.

22. Campbell, D. T. (1976). *Assessing the impact of planned social change* (Occasional Paper No. 8). Hanover, NH: Dartmouth College Public Affairs Center.

23. Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management, 30,* 418–446.

24. Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis, 24,* 305–331.

25. Ahn, T., & Vigdor, J. (2014). *The impact of No Child Left Behind's accountability sanctions on school performance: Regression discontinuity evidence from North Carolina* (NBER Working Paper No. 20511). Cambridge, MA: National Bureau of Economic Research.

26. Chiang, H. (2009). How accountability pressure on failing schools affects student achievement. *Journal of Public Economics, 93,* 1045–1057.

27. Deming, D. J., Cohodes, S., Jennings, J., & Jencks, C. (2013). *School accountability, postsecondary attainment, and earnings* (NBER Working Paper No. 19444). Cambridge, MA: National Bureau of Economic Research.

28. Chiang, H., Wellington, A., Hallgren, K., Speroni, C., Herrmann, M., Glazerman, S., & Constantine, J. (2015). *Evaluation of the Teacher Incentive Fund: Implementation and impacts of pay-for-performance after two years.* Washington, DC: U.S. Department of Education, Institute of Education Sciences.

29. Springer, M. G., Pane, J. F., Le, V., McCaffrey, D. F., Burns, S. F., Hamilton, L. S., & Stecher, B. (2012). Team pay for performance: Experimental evidence from the Round Rock Pilot Project on Team Incentives. *Educational Evaluation and Policy Analysis, 34,* 367–390.

30. Glazerman, S., & Seifullah, A. (2012). *An evaluation of the Chicago Teacher Advancement Program (Chicago TAP) after four years.* Washington, DC: Mathematica Policy Research.

31. Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia: A century of public school reform.* Cambridge, MA: Harvard University Press.

32. Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica, 73,* 417–458.

33. Snipes, J., Doolittle, F., & Herlihy, C. (2002). *Foundations for success: Case studies of how urban school systems improve student achievement.* New York, NY: MDRC.

34. Beatty, B. (2011). The dilemma of scripted instruction: Comparing teacher autonomy, fidelity, and resistance in the Froebelian kindergarten, Montessori, Direct Instruction, and Success for All. *Teachers College Record, 113,* 395–430.

35. Borman, G. D., Hewes, G. M., Overman, L. T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research, 73,* 125–230.

36. Enzle, M. E., & Anderson, S. C. (1993). Surveillant intentions and intrinsic motivation. *Journal of Personality and Social Psychology, 64,* 257–266.

37. Ross, J., & Staw, B. M. (1993). Organizational escalation and exit: Lessons from the Shoreham nuclear power plant. *Academy of Management Journal, 36,* 701–732.

38. Patil, S., Vieider, F., & Tetlock, P. E. (2012). Process versus outcome accountability. In M. Bovens, R. E. Goodin, & T. Schillemans (Eds.), *Oxford handbook of public accountability* (pp. 69–89). New York, NY: Oxford University Press.

39. Baer, R., Hinkle, S., Smith, K., & Fenton, M. (1980). Reactance as a function of actual versus projected autonomy. *Journal of Personality and Social Psychology, 35,* 416–422.

40. Tyack, D., & Hansot, E. (1982). *Managers of virtue: Public school leadership in America, 1820–1980.* New York, NY: Basic Books.

41. Glenn, C. L., Jr. (1988). *The myth of the common school.* Amherst: University of Massachusetts Press.

42. Friedman, M. (1955). The role of government in education. In R. A. Solo (Ed.), *Economics and the public interest.* Piscataway, NJ: Rutgers University Press.

43. Chubb, J., & Moe, T. (1990). *Politics, markets, and America's schools.* Washington, DC: Brookings Institution Press.

44. Coons, J. E. (1992). School choice as simple justice. *First Things, 22,* 193–200.

45. Gill, B., Timpane, P. M., Ross, K. E., Brewer, D. J., & Booker, K. (2007). *Rhetoric versus reality: What we know and what we need to know about vouchers and charter schools.* Santa Monica, CA: RAND.

46. Abdulkadiroglu, A., Angrist, J. D., Dynarski, S. M., Kane, T. J., & Pathak, P. A. (2011). Accountability and flexibility in public schools: Evidence from Boston's charters and pilots. *Quarterly Journal of Economics, 126,* 699–748.

47. Booker, K., Sass, T. R., Gill, B., & Zimmer, R. (2011). The effects of charter high schools on educational attainment. *Journal of Labor Economics, 29,* 377–415.

48. Dobbie, W., & Fryer, R. G. (2015). The medium-term impacts of high-achieving charter schools. *Journal of Political Economy, 123,* 985–1037.

49. Gleason, P. M., Tuttle, C. C., Gill, B., Nichols-Barrer, I., & Teh, B. (2014). Do KIPP schools boost student achievement? *Education Finance and Policy, 9,* 36–58.

50. Gill, B., & Nichols-Barrer, I. (2014). Charter schools. In D. Brewer & L. Picus (Eds.), *Encyclopedia of education economics and finance* Thousand Oaks, CA: Sage.

51. Center for Research on Education Outcomes. (2013). *National charter school study.* Stanford, CA: Stanford University.

52. Wolf, P., Kisida, B., Gutmann, B., Puma, M., Eissa, N., & Rizzo, L. (2013). School vouchers and student outcomes: Experimental evidence from Washington, DC. *Journal of Policy Analysis and Management, 32,* 246–270.

53. Chingos, M. M., & Peterson, P. E. (2015). Experimentally estimated impacts of school vouchers on college enrollment and degree attainment. *Journal of Public Economics, 122,* 1–12.

54. Mills, J. N., & Wolf, P. J. (2016). *The effects of the Louisiana Scholarship Program on student achievement after two years* (Louisiana Scholarship Program Evaluation Report 1). Fayetteville: University of Arkansas, School Choice Demonstration Project.

55. Abdulkadiroglu, A., Pathak, P. A., & Walters, C. R. (2015). School vouchers and student achievement: Evidence from the Louisiana Scholarship Program (NBER Working Paper No. 21839). Cambridge, MA: National Bureau of Economic Research.

56. Zimmer, R. W., & Toma, E. F. (2000). Peer effects in public and private schools across countries. *Journal of Policy Analysis and Management, 19,* 75–92.

57. Imberman, S. (2011). The effect of charter schools on achievement and behavior of public school students. *Journal of Public Economics, 95,* 850–863.

58. Bifulco, R., & Ladd, H. (2006). The impacts of charter schools on student achievement: Evidence from North Carolina. *Education Finance and Policy, 1,* 50–90.

59. Bettinger, E. P. (2005). The effect of charter schools on charter students and public schools. *Economics of Education Review, 24,* 133–147.

60. Zimmer, R., Gill, B., Booker, T. K., Lavertu, S., Sass, T. R., & Witte, J. (2009). *Charter schools in eight states: Effects on achievement, attainment, integration, and competition.* Santa Monica, CA: RAND.

61. Winters, M. A. (2012). Measuring the competitive effect of charter schools on public school student achievement in an urban environment: Evidence from New York City. *Economics of Education Review, 31,* 293–301.

62. Jinnai, Y. (2014). Direct and indirect impact of charter schools' entry on traditional public schools: New evidence from North Carolina. *Economics Letters, 124,* 452–456.

63. Chingos, M. M., & Peterson, P. E. (2010). It's easier to pick a good teacher than to train one: Familiar and new results on the correlates of teacher effectiveness. *Economics of Education Review, 30,* 449–465.

64. Clotfelter, C. T., Ladd, H. F., & Vigdor, J. L. (2007). Teacher credentials and student achievement: Longitudinal analysis with student fixed effects. *Economics of Education Review, 26,* 673–682.

65. Garet, M. S., Porter, A. C., Desimone, L., Birman, B. F., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal, 38,* 915–945.

66. Hawley, W., & Valli, L. (1999). The essentials of effective professional development: A new consensus. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 151–180). San Francisco, CA: Jossey-Bass.

67. Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness.* New York, NY: New Teacher Project.

68. Cowan, J., & Goldhaber, D. (2015). National Board certification and teacher effectiveness: Evidence from Washington (CEDR Working Paper 2015-3). Seattle: University of Washington Bothell, Center for Education Data and Research.

69. Barnwell, P. (2015, February 18). Why schools need more 'hybrid' teaching roles. *Education Week.* Retrieved from http://www.edweek.org/tm/articles/2015/02/18/why-schools-need-more-hybrid-teaching-roles.html

70. Gawande, A. (2011, October 3). Personal best. *The New Yorker.* Retrieved from http://www.newyorker.com/magazine/2011/10/03/personal-best

71. Newton, L. H., Hodges, L., & Keith, S. (2013). Accountability in the professions: Accountability in journalism. *Journal of Mass Media Ethics, 19,* 166–190.

72. Porter, A. C., Polikoff, M. S., Goldring, E. B., Murphy, J., Elliott, S. N., & May, H. (2010). Investigating the validity and reliability of the Vanderbilt Assessment of Leadership in Education. *The Elementary School Journal, 111,* 282–313.

73. Ladd, H. F. (2016, May 26). Now is the time to experiment with school inspections for accountability [Blog post]. Retrieved from https://www.brookings.edu/2016/05/26/now-is-the-time-to-experiment-with-inspections-for-school-accountability/

74. Chicago Public Schools. (2014). *REACH Students educator evaluation handbook 2014–15.* Chicago: Author.

75. Organisation for Economic Co-operation and Development. (2014). *TALIS 2013 results: An international perspective on teaching and learning.* Paris, France: Author.

76. Organisation for Economic Co-operation and Development. (2014). *PISA 2012 results in focus.* Paris, France: Author.

77. Pil, F. K., & Leana, C. (2009). Applying organizational research to school reform: The effects of human and social capital on student performance. *Academy of Management Journal, 52,* 1101–1124.

78. Bryk, A. S., & Schneider, B. (2004). *Trust in schools: A core resource for improvement.* New York, NY: Russell Sage Foundation.

79. Ashton, R. H. (1992). Effects of justification and a mechanical aid on judgment performance. *Organizational Behavior and Human Decision Processes, 52,* 292–306.

80. Lerner, J. S., Goldberg, J. H., & Tetlock, P. E. (1998). Sober second thought: The effects of accountability, anger, and authoritarianism on attributions of responsibility. *Personality and Social Psychology Bulletin, 24,* 563–574.

81. Bodenhausen, G. V., Kramer, G. P., & Susser, K. (1994). Happiness and stereotypic thinking in social judgment. *Journal of Personality and Social Psychology, 66,* 621–632.

82. Tetlock, P. E. (1985). Accountability: A social check on the fundamental attribution error. *Social Psychology Quarterly, 48,* 227–236.

83. Sherman, G. D., Lee, J. J., Cuddy, A. J. C., Renshon, J., Oveis, C., Gross, J. J., & Lerner, J. S. (2012). Leadership is associated with lower levels of stress. *Proceedings of the National Academy of Sciences, USA, 109,* 17903–17907.

84. Hancock, P. A., & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31,* 519–537.

85. Kahneman, D. (2011). *Thinking, fast and slow.* New York, NY: Farrar, Straus and Giroux.

86. Thaler, R. H., & Sunstein, C. B. (2008). *Nudge: Improving decisions about health, wealth, and happiness.* New Haven, CT: Yale University Press.

87. Cialdini, R. B. (2007). Descriptive social norms as underappreciated sources of social control. *Psychometrika, 72,* 263–268.

88. Hallsworth M., List, J., Metcalfe, R., & Vlaev, I. (2014). *The behavioralist as tax collector: Increasing tax compliance through natural field experiments* (NBER Working Paper No. 20007). Cambridge, MA: National Bureau of Economic Research.

89. Kraft, M., & Rogers, T. (2015). The underutilized potential of teacher-to-parent communication: Evidence from a field experiment. *Economics of Education Review, 47,* 49–63.

90. Hallberg, K., & Green, G. (2015, March 11). How can we hire and keep high quality teachers in struggling schools? [Blog post]. Retrieved from http://educationpolicy.air.org/blog/how-can-we-hire-and-keep-high-quality-teachers-struggling-schools

91. Furgeson, J., Gill, B., Haimson, J., Killewald, A., McCullough, M., Nichols-Barrer, I., & Lake, R. (2012). *Charter-school management organizations: Diverse strategies and diverse student impacts.* Cambridge, MA: Mathematica Policy Research.

92. Marsh, J. A., McCombs, J. S., & Martorell, F. (2010). How instructional coaches support data-driven decision making. *Educational Policy, 24,* 872–907.

93. Blazar, D., & Kraft, M. A. (2015). Exploring mechanisms of effective teacher coaching: A tale of two cohorts from a randomized experiment. *Educational Evaluation and Policy Analysis, 37,* 542–566.

94. Kane, T. J. (2012). Capturing the dimensions of effective teaching. *Education Next, 12*(4), 35–41.

95. Chaplin, D., Gill, B., Thompkins, A., & Miller, H. (2014). *Professional practice, student surveys, and value added: Multiple measures of teacher effectiveness in the Pittsburgh Public Schools* (REL 2014-024). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Mid-Atlantic.

96. Walsh, E., & Lipscomb, S. (2013). *Classroom observations from Phase 2 of the Pennsylvania Teacher Evaluation Pilot: Assessing internal consistency, score variation, and relationships with value added.* Cambridge, MA: Mathematica Policy Research.

97. L'Hommedieu, R., Menges, R. J., & Brinko, K. T. (1990). Methodological explanations for the modest effects of feedback from student ratings. *Journal of Educational Psychology, 82,* 232–241.

98. Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *American Economic Review, 102,* 3628–3651.

99. Dee, T. S., & Wyckoff, J. (2015). Incentives, selection, and teacher performance: Evidence from IMPACT. *Journal of Policy Analysis and Management, 34,* 267–297.

100. Rockoff, J. E., Staiger, D. O., Kane, T. J., & Taylor, E. S. (2012). Information and employee evaluation: Evidence from a randomized intervention in public schools. *American Economic Review, 102,* 3184–3213.

101. Frank, K. (2012, February 24). Constitution for effective school governance [ID No. 16715]. *Teachers College Record.* Available from http://www.tcrecord.org

102. Schillemans, T., & Smulders, R. (2015). Learning from accountability? Whether, what, and when. *Public Performance & Management Review, 39,* 248–271.

103. Blume, H. (2016, June 13). Less test-iness over L.A. teacher evaluations. *Los Angeles Times.* Retrieved from http://www.latimes.com/