Reimagining Accountability in K-12 Education: A Behavioral Science Perspective

Brian P. Gill
Mathematica Policy Research

Jennifer S. Lerner
Harvard University

Paul Meosky
Harvard University

DRAFT, 2015.10.06

Contact: Brian P. Gill, bgill@mathematica-mpr.com
Abstract

Over the last two decades, the primary lever American policymakers have used to improve K-12 school performance has been “accountability” in the particular form of high-stakes testing. The policy debate has raged largely without reference to the behavioral science literature on accountability, despite the fact that decades of experiments illuminate conditions under which specific forms of accountability can improve outcomes, have no effect, or make matters worse. We seek to bridge these literatures, extract policy implications from behavioral science and offer tools for policy makers seeking to improve educational performance. We conclude that an effective accountability regime requires multiple forms of accountability, multiple measures of educational practice and outcomes, and feedback mechanisms to promote improved practice. Professional accountability, which has historically been underutilized in schools and which is compatible with behavioral “nudges,” particularly merits a larger role.
The No Child Left Behind Act of 2002 (NCLB) established a regime of public reporting of every public school’s level of student proficiency in math and reading, along with automatic sanctions for failure to achieve test-score targets. Schools that fell short of state-determined test-score goals faced mandatory interventions: chronically failing schools could be re-staffed, taken over by the state, or shut down. In education policy, “accountability” came to be synonymous with the use of high-stakes testing and the primary lever used by American policymakers to improve school performance.

More recently, the Obama Administration doubled down on high-stakes testing, offering states flexibility to waive some of the NCLB’s most onerous aspects if they adopted several favored policies, notably the incorporation of student achievement growth as an explicit component in the evaluation of teachers and principals. Over 40 states are beginning to include student test scores in educator evaluations.

As Congress now considers revamping NCLB, the efficacy of high-stakes testing is hotly contested. Critics have called for the reduction or abandonment of high-stakes testing. An “opt out” movement has enlisted parents who refuse to have their children take the standardized tests. Scholars have proposed refining or replacing the current accountability regime (Tucker, 2014; Darling-Hammond, Wilhoit, and Pittenger, 2014; Center on Reinventing Public Education and Fordham Institute, 2014).

An extensive literature in experimental psychology and behavioral economics identifies the conditions under which specialized forms of accountability can improve outcomes, have no effect, or make matters worse. The education policy debate has gone on largely without reference to this literature, which could provide valuable insights for the design of school accountability
regimes.¹ Lerner and Tetlock’s (1999) comprehensive review of the behavioral literature makes two broad points that motivate our discussion. First, accountability comes in many forms, involving different mechanisms and different behavioral responses. Outcome-based accountability (such as high-stakes testing) is only one of these forms. Second, accountability can have positive and/or negative effects on judgments and decision making, depending on the accountability type, the decision context, and the nature of the task.

We apply insights from the behavioral science literature to accountability regimes in education, and propose how policymakers might use a range of accountability tools to promote school performance. Complicating the classic principal-agent problem that accountability policies seek to address, schooling has multiple public and private purposes, with multiple constituencies, including parents, students and the public. The K-12 accountability system must attend to these purposes and constituencies.

We conclude that continuous improvement in schools would be best served by (1) employing multiple forms of accountability that (2) use multiple measures of student outcomes and professional practice and (3) incorporate feedback mechanisms to promote better practice. More specifically, policymakers should consider increasingly using robust forms of professional accountability, which historically has been underutilized in schools.

Defining Accountability

Lerner and Tetlock’s review identifies four mechanisms that make people feel accountable. Ranging from minimal to maximal, these include: (1) the mere presence of another, or simple awareness that someone else is watching; (2) identifiability, or the expectation

---

¹ A National Academies report (Hout and Elliot, 2011) examined some of the behavioral literature related to the use of test-based incentives, but did not examine other forms of accountability.
that an action or outcome will be personally attributable to oneself; (3) *reason-giving*, or the expectation that one will need to explain actions to another; and (4) *evaluation*, or the assessment of one’s performance by another with particular rules and consequences (1999).

The high-stakes testing regime inaugurated by NCLB and expanded under President Obama incorporates the identifiability and evaluation components described in the behavioral literature, but the generality of the four mechanisms suggests many other ways to employ accountability in schooling. We describe four kinds of accountability systems that operate in various fields and professions and are potentially relevant in schools.

- **Outcome-based accountability** holds decision makers responsible for achieving particular results, with positive or negative consequences depending on their success. Public officials form the constituency and define the outcomes. In education, high-stakes testing is the prominent example of outcome-based accountability.

- **Rule-based accountability** requires decision makers to act in accordance with rules or regulations that delineate mandated or forbidden activities. Public officials form the constituency and define the rules. State education codes, regulations, and teacher contracts create rule-based accountability in schools.

- Under **market-based accountability**, clients or customers are the primary constituency, and they implicitly hold decision makers responsible by choosing among different providers. Private schools, voucher schools, and charter schools are subject to market-based accountability.

- **Professional accountability** gives resources to decision makers resources and holds them responsible for following professional standards of practice. Professional peers form the
constituency and define the rules. Programs in which teachers observe and provide feedback on colleagues’ professional practice are examples in schooling.

Most of these accountability mechanisms are used in other professions, which establish rules and standards for minimal performance, operate in markets that allow their clients a choice of providers, and expect practitioners to use state-of-the-art practices. Professional accountability mechanisms in other fields also typically include substantial educational expectations for entry into the profession and resources for practitioners’ continuing learning. Outcome-based accountability is not prominent in most professions (with the notable exception of tort law, in which plaintiffs’ attorneys are typically paid only if they win), perhaps because market-based accountability serves the same function. Outcome-based accountability is increasingly being attempted in health care, however (Gold, 2010), which resembles K-12 schooling in that consumers have limited market power.

Table 1 suggests how the accountability types described above employ the mechanisms identified in the behavioral literature.

Table 1. Accountability types in schooling and psychological accountability mechanisms

<table>
<thead>
<tr>
<th>Psychological accountability mechanisms</th>
<th>Outcome-based</th>
<th>Rule-based</th>
<th>Market-based</th>
<th>Professional</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mere presence of another</td>
<td>✔</td>
<td></td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Identifiability</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Reason-giving</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Evaluation</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

Having culled from both education practice and the behavioral science literature a matrix of possible forms accountability can take and the mechanisms through which it works, we now describe applications of these accountability types in schools. When the literature allows, we
delineate which findings arise from laboratory-based research as well as which arise from field-based research.

Outcome-based Accountability

Prior to the 1990s, outcome-based accountability was almost unknown in K-12 schooling. The education standards movement that gained steam during the 1990s promoted “proficiency” in reading and math, which required clear expectations at each grade level and measurement of students’ success at meeting them.

The 1994 re-authorization of ESEA required states to set proficiency standards, assess their students in multiple grades, and publicly report school-by-school proficiency levels. Thus, the primary accountability mechanism was identifiability: student outcomes were publicly reported, but typically no explicit sanctions were applied to schools. Eight years later, NCLB added explicit sanctions to schools falling short of state-defined proficiency targets, thereby incorporating the evaluation mechanism.

Behavioral Evidence on Outcome-based Accountability. The behavioral literature has found few instances of positive effects of outcome-based accountability in randomized experiments. Because it does not constrain decision makers with rules, outcome accountability might be more effective than other forms of accountability at promoting innovation (Patil, Vieider, and Tetlock, 2012), but the existing empirical literature has not deeply tested this possibility. The effectiveness of outcome-based accountability can be undermined by the sunk-cost bias, which makes decision makers more likely to pursue courses of action simply because of prior investments, even with low odds of success (Simonson & Staw 1992). In addition, in many contexts, using tangible rewards as performance incentives has been shown to undermine intrinsic motivation (Deci, Koestner, and Ryan, 1999).
By increasing a decision’s difficulty, outcome-based accountability may impair decision making by eliciting stress and negative emotions (Zhang & Mittal 2005; Siegel-Jacobs & Yates 1996). Perhaps because stress burdens one’s cognitive load (Mendl, 1999; Lazarus & Folkman, 1984), outcome accountability sometimes fails to increase the use of strategies that require substantial effort (Lerner & Tetlock 1999).

In addition, accountability regimes viewed as illegitimate can be counterproductive (Lerner & Tetlock, 1999). Many teachers are suspicious of “value-added” models (VAMs) that aim to quantify their contributions to student achievement. Improving on cruder outcome-based accountability regimes that rely on student achievement levels, VAMs account statistically for factors outside the teacher’s control. Although there is evidence that well-designed VAMs can produce unbiased measures of teachers’ contributions to student achievement growth (Kane et al., 2013; Chetty, Friedman, and Rockoff, 2013), suspicions on the part of practitioners could undermine their ability to promote performance improvements.

Moreover, even unbiased measures of educators’ contributions to student achievement can be problematic in a high-stakes accountability regime. Achievement tests never fully capture all of the skills and knowledge that society expects schools to impart. Some evidence suggests that instructional practices that raise test scores differ from those that promote students’ effort and long-term goals (Ferguson and Danielson, 2014). Using tests for accountability encourages schools to focus on tested elements of learning to the detriment of untested elements, as demonstrated by the psychology literature on conformity and the availability heuristic (Tetlock, Skitka, & Boettger, 1989), and may undermine the validity of the test itself (Campbell, 1976).

**Field Evidence on Outcome-based Accountability.** The effects of NCLB in particular and high-stakes testing in general are a matter of heated debate. Most existing studies suggest
positive effects in at least some grades and subjects (Dee & Jacob, 2011; Carnoy & Loeb, 2002; Hanushek & Raymond, 2005; Ahn & Vigdor, 2014; Deming et al., 2013 finds mixed results). Dee and Wyckoff (2015) find evidence of favorable effects for low-rated and high-rated teachers on the teacher evaluation system now used by Washington, D.C. public schools. The effects of performance-pay incentives for teachers on student achievement have been mixed (Springer et al., 2012; Glazerman & Seifullah 2012).

Meanwhile, since the passage of NCLB, many schools have narrowed the curriculum (Dee, Jacob, & Schwartz, 2013) and spent increasing time preparing for state assessments (Hamilton, Stecher, & Yuan, 2012; Jennings and Sohn, 2014). Scores on low-stakes assessments have improved more slowly than scores on high-stakes assessments, and sometimes not at all (Jacob, 2005; Jennings and Sohn, 2014; Holcombe, Jennings, and Koretz, 2013). In extreme cases, educators have been caught cheating on high-stakes assessments. Teacher-developed “student learning objectives” (SLOs), increasingly used as outcome-based accountability measures for teachers of students in non-tested grades (Gill et al., 2014), may be especially susceptible to inflation, because they ask teachers to grade themselves. Not surprisingly, Balch and Springer (2015) found no correlation between teachers’ success in achieving SLOs and their value added as measured by standardized assessments.

Rule-based Accountability

Rule-based accountability regimes implicitly rely on the mechanism of identifiability, and sometimes the presence of another. Historically, schools have relied on rule-based accountability to define structural conditions of schooling and set minimum teaching standards (Tyack, 1974). Rule-based accountability has included traditional teacher evaluations based on principal observations, which typically deem 98-99% of teachers satisfactory (Weisberg et al.
2009); state-level decisions about textbooks; contractual rules about working conditions, hours, and class size; and federal and state spending regulations. However, teachers typically had wide discretion about instruction throughout the 20th century (Tyack and Cuban, 1995), a practice that allowed enormous variation in school and teacher quality. Over time, rules and protocols may have raised minimum standards and reduced the most egregious inequities, but they have reduced opportunities for innovation and may have sacrificed efficiency for regularity (see Tyack, 1974; Chubb & Moe, 1990; Wilson, 1989).

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 (see Snipes et al., 2002). Principals have been asked to facilitate curriculum management by taking 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 (Beatty, 2011).

Limited evidence supports a maximal rule-based approach in schools. Scripted “direct instruction” has been found to promote student achievement in elementary grades (Borman et al., 2003). But maximal rule-based accountability could become counterproductive, because close monitoring often reduces intrinsic motivation due to perceived loss of autonomy (Enzle & Anderson, 1993). Intense procedural scrutiny can also exacerbate the sunk-cost bias (Ross & Staw, 1993), undermine innovation, and entrench suboptimal practices (Patil, Vieider, & Tetlock, 2012). The perception of rules as illegitimate can produce a boomerang effect (Baer, Hinkle, Smith, & Fenton, 1980; Brehm, 1966; Heilman & Toffler, 1976). Indeed, aggressive
rule-based accountability may be especially unsuited to teaching, an inherently complex task that requires daily adjustments and judgments (Duffy et al., 1986; O’Keefe & Johnston, 1989).

**Market-based Accountability**

Market-based accountability is based on principles from traditional neoclassical economics rather than newer behavioral approaches. Implicitly, it involves the identifiability and reason-giving mechanisms: 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 policies, unlike in most other economic sectors. 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 the provision of publicly supported education (Tyack & Hansot, 1982; Glenn, 1988).

Policymakers have shown increasing interest in incorporating market-based accountability into education policies, motivated by the belief that (1) local monopolies controlled by school boards may not be the most effective way to provide schooling (Friedman, 1955; Chubb & Moe, 1990) and (2) giving families choices in schooling is inherently valuable (Coons, 1992).

The most prominent examples of market-based K-12 education reform are charter schools—publicly funded schools of choice that operate outside the direct control of local school districts (Gill et al., 2007). The first charter schools opened only two decades ago; today, over 6,000 operate in more than 40 states, representing about 6% of all schools across the country (NAPCS) and much higher percentages in some cities. Meanwhile, a smaller number of states
have adopted voucher programs that permit some students to attend private schools at public expense.

Empirically, the evidence on the effects of market-based schools on students’ test scores and longer-term educational attainment suggests promise but is not definitive. In some contexts and locations, charter schools are producing substantial positive effects on their students (Abdulkadiroglu et al., 2011; Furgeson et al., 2012; Booker et al., 2011; Tuttle et al., 2013), but their performance varies widely (Gill & Nichols-Barrer, 2014; Center for Research on Educational Outcomes, 2013). A few studies of small-scale voucher programs have found positive educational impacts, particularly for African-American students (Wolf et al., 2013; Chingos & Peterson, 2015).

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 agents (educators) and clients—is complicated by the involvement of multiple clients (students, parents, and the public), whose interests may not be fully aligned. Moreover, underage students are presumed to be not fully capable of knowing their own interests. In addition, students’ educational experiences are affected not only by school quality, but by externalities, including characteristics of other students (Betts & Morell, 1999; Zimmer & Toma, 2000; Lavy, Paserman, & Schlosser, 2011). As a consequence, an unfettered market may produce segregated schools, as parents with high levels of knowledge, wealth, or motivation seek out schools with other families like their own (Bifulco & Ladd, 2006; Booker et al., 2008).

Relatedly, skeptics worry that market-based schools will drain conventional public schools of funding and motivated families. Supporters argue that breaking the local monopoly produces healthy competitive pressure that will benefit all students. Although several studies
have found neither positive or negative effects of charter schools on achievement in nearby conventional schools (Bettinger, 2005; Bifulco & Ladd, 2006; Zimmer et al., 2009), others have found evidence of small positive effects (Jinnai, 2013; Zimmer et al., 2009; Winters, 2012); only one has found a negative effect (Imberman, 2011).

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 in a democracy. Historically, this key rationale for the common-school model implicitly assumed that effective education for citizenship required public operation of schools (Tyack & Hansot, 1982; Glenn, 1988). The fact that the education of citizens is a public good argues against relying entirely on market mechanisms.

Indeed, 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 a substantial rule-based and outcome-based accountability: They cannot charge tuition, they must submit their students to the same high-stakes tests taken in conventional public schools, and they are (typically) required to admit all applicants for whom they have space. Moreover, charter schools operate under the supervision of a publicly empowered authorizer. 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 admissions criteria.

Professional Accountability

Prominent voices are calling for an increased reliance on professional accountability which would give teachers more support, collaboration, and training, and set higher expectations
(Tucker, 2014; Darling-Hammond et al., 2014). Professional accountability can take many forms, involving all four of the accountability mechanisms identified by Lerner and Tetlock. Licensing, standards, and reviews by organizations representing the teaching profession involve evaluation; observations and assistance by supervisors, instructional coaches, peers, or mentors involve identifiability, reason-giving, and sometimes evaluation; collaboration and co-teaching involve the presence of another and reason-giving.

Professional accountability is not synonymous with professionalism. In many fields, professionalism implies an ethic of meeting standards even in the absence of observers and consequences. Thus, we do not consider professionalism per se to be a form of accountability, which by definition involves an external observer.

**Traditional and Novel Versions of Professional Accountability in Education**

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 relationship to improved student achievement (Chingos & Peterson, 2011; Clotfelter, Ladd, & Vigdor, 2007), and most studies find no effect of professional development on student achievement (e.g., Garet, Porter, Desimone, Birman, & Yoon, 2001; Hawley & Valli, 1999).

More robust and ambitious forms of professional accountability may hold more promise. Licensing and professional requirements at a high enough level—such as those completing the rigorous certification process of the National Board for Professional Teaching Standards—might help identify especially effective teachers (Cowan & Goldhaber, 2015; Clotfelter, Ladd, & Vigdor, 2007). Teacher quality is the most influential school-controlled factor in student achievement growth (Rivkin et al., 2005), a fact that has prompted policymakers’ aggressive
promotion of more rigorous teacher evaluations. States and school districts are adopting extensive rubrics for the observation and rating of teaching practice. 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. If taken beyond screening and compensation reform, they have the unique advantage of coupling accountability with support.

Novel forms of professional accountability might include restructuring traditional 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 than is typical (Hallberg and Green, 2015). Other districts and schools are giving teachers leadership opportunities to develop the skills of their colleagues (e.g., Barnwell, 2015), as through instructional coaching (Gawande, 2011).

In most professions, professional accountability includes answerability to clients (Newton, Hodges, and Keith, 2013). Unlike in higher education, K-12 schooling traditionally involves little direct accountability of educators to students. School districts such as the Pittsburgh Public Schools are now including student surveys in new teacher evaluation systems (Chaplin, Gill, Thompkins, and Miller, 2014). And some are using teacher surveys in principal evaluation, applying the business world’s 360-degree feedback (Porter et al., 2010).

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 school visit by outside experts who observe instruction; interview
teachers, students, and parents; and examine school performance data (see, e.g., Tucker, 2014; Darling-Hammond et al., 2014). The review concludes with clear recommendations for improvement.

Professional accountability approaches would make teaching more transparent, activating all four accountability mechanisms. Indeed, such rich professional accountability systems emphatically reject the notion that professionalism means allowing teachers complete discretion to practice as they choose. Rules such as those found in the Chicago Public Schools (2014) that explicitly prohibit the use of classroom video recordings for evaluation are anathema to this vision of professional accountability. Meanwhile, schools like the Kauffman Charter School in Kansas City are 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 (OECD, 2014) found that although responding American teachers were more likely than average to receive feedback from principals, only 11% received feedback from mentors, versus 39% in Japan, 38% in Singapore, and 24% in Australia. And only 27% of responding American teachers received feedback from colleagues, versus 84% in Korea, 57% in the Netherlands, and 43% in Finland. Those countries significantly outperformed the U.S. in math, reading, and science in the most recent study of the Program for International Assessment. American teachers were also far less likely than their counterparts in other countries to receive feedback from student surveys and they were less likely to report that the feedback they received led to public recognition, career advancement, or increased compensation.
Behavioral Evidence on Professional Accountability’s Merit. Many studies demonstrate the value of requiring people to justify their decisions to others, a common expectation of professional accountability regimes. Ashton (1992) found that requiring subjects to justify their decisions encouraged high-effort strategies that were sensitive to evidence, alleviating mistakes and inconsistencies. Similarly, Lerner et al. (1998) found that the need to justify decisions stimulated systematic thinking and attention to evidence. Bodenhausen, Kramer, and Susser (1994) demonstrated that justification reduced reliance on stereotypic thinking. Tetlock (1985) found that subjects who had to justify their judgments were less likely to be led astray by the fundamental attribution error (i.e., the tendency to over-attribute responsibility to individuals rather than situations). Process accountability for groups has been found to increase demand for information, induce information sharing, and produce better outcomes (Scholten, van Knippenberg, Nijstad, and De Dreu, 2007).

Other studies suggest that by encouraging people to take initiative, professional accountability may increase their sense of control and consequently improve their performance on attention-demanding tasks, promote more considerate decision-making, and assist memory formation (Sherman et al., 2012; Nabi et al., 2013; Hancock, 1989). More generally, professional accountability may best encourage the systematic, effortful, and self-critical thinking associated with even-handed, accurate reasoning using systematic (rather than heuristic) processing (see Kahneman, 2011). Professional accountability might offer the flexibility needed for innovation while disallowing the adoption of the cognitively lazy but easily defensible decision (Tetlock, 1991).

Professional accountability also finds support from political science. In discussing accountability in political representation, Mansbridge (2009) argues that effective selection of
agents is superior to a sanctions-based approach to the principal-agent problem. Although Mansbridge is writing about representative democracy, the point is also relevant in the schooling context, where educators’ primary aims are consonant with the aims of policymakers, parents, and students. When goals are aligned, the selection model allows greater freedom for educators, reducing the need for detailed rules and extrinsic motivators. Similar arguments have been made about accountability in business (Pratt and Zeckhauser, 1985). In the education context, the selection model suggests the need for transparent professional standards, rigorous evaluation, and a complementary element of market-based accountability.

Field Evidence on the Effects of Professional Accountability’s Merit. 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 (Hallberg and Green, 2015), and early evidence on the effects of instructional coaching on student achievement is encouraging (Furgeson et al., 2012; Marsh, McCombs, & Martorell, 2010). Several recently developed rubrics for observing and evaluating instructional practice have produced scores significantly correlated with teachers’ value added (Kane, 2012; Chaplin et al., 2014; Walsh & Lipscomb, 2013), suggesting that careful observation can produce feedback that could improve student outcomes.

In higher education, student evaluations of teachers have had some positive effects on teaching (L’Hommedieu, Menges, & Brinko, 1990). Recent studies examining student surveys in K-12 schools have found the results are (modestly) correlated with measures of teachers’ contributions to student achievement (Kane, 2012; Chaplin et al., 2014), suggesting promise.

An intensive, peer-based teacher evaluation system used in Cincinnati offers promising 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 (Taylor & Tyler, 2012).

Professional accountability is also compatible with the behavioral “nudges” (Thaler & Sunstein that are increasingly being adopted in various areas of public policy. Field trials have demonstrated, for example, that appealing to social norms (“most people like you do X”) can powerfully influence behavior in contexts ranging from tax collection (Hallsworth et al., 2014) to motivating parents to keep their children in school (Kraft & Rogers, 2015). This suggests the possibility that evaluation feedback to teachers, for example, could lead to improvement even in the absence of high-stakes outcome accountability, if relevant information is provided appropriately.

Conclusion: Increasing Professional Accountability in a Multi-mode Approach

“Accountability,” as it is used in K-12 education policy, reflects an unnecessarily narrow understanding of accountability mechanisms. The outcome-based accountability that has been the focus of policymakers’ attention can produce positive results, but relying on it exclusively is likely to lead to a variety of unintended and undesirable consequences.

Rule-based accountability, long included in American education policy, has helped set minimum standards and expectations, but has not been designed to promote high performance. And the effectiveness of more aggressive rule-based approaches is likely to decline with the decision maker’s increasing distance from the classroom. Detailed mandates from distant officials are especially susceptible to being perceived by educators as illegitimate.

Market-based accountability likewise can play a productive role in improving school performance, but even most supporters of school choice do not advocate an unconstrained
educational market. Given the effects of student sorting and the public purposes of schooling, rule-based modifications to market-based accountability make good sense.

Professional accountability is the most underused form of accountability in K-12 schooling in America. Barriers to entering the profession are low. Evaluation standards have historically been low. Compensation and career advancement have not been based on any measure of performance. And there has been little expectation that teachers will continually improve their practice, or even that their practice will be regularly observed by peers. This is changing, with initiatives related to rigorous educator evaluation and job-embedded professional development (coaching) now underway in schools across the country.

After reviewing the evidence from behavioral science and the field literature, we propose that accountability systems to improve the performance of K-12 schools should include three key elements. First, the diversity of advantages and disadvantages of different forms of accountability suggests that multiple forms of accountability might be usefully employed in complementary ways. Patil, Vieider, and Tetlock (2012) recently noted that outcome-based accountability may better promote innovation, while process-based accountability (including forms of professional accountability) may better promote the use of identified best practices. They suggest that the disadvantages of both types might be counteracted by systems that promote the empowerment of decision makers to reduce conformity to deficient standard practices, encourage focus on outcomes, and facilitate organizational learning. Empowerment is implicit in professional accountability but can also be incorporated in an outcome accountability regime that communicates a desire to achieve shared objectives, they argue. Similarly, according to a National Academies report (Hout and Elliott, 2011), external rewards are most likely to be effective when they are well aligned with educators’ intrinsic aims, promoting “autonomous
motivation.” Creative use of behavioral nudges permits professional discretion while encouraging desirable practices.

The Equity Project (TEP) Charter School in New York City provides a compelling example of the advantages of using multiple accountability approaches (Furgeson, McCullough, Wolfendale, and Gill, 2014). The school pays its teachers $125,000 plus bonuses based on school-wide improvements in student achievement, but its approach goes beyond compensation. TEP’s teachers are rigorously screened; have contracts renewed based on performance; engage in six weeks of professional development annually; and observe each other in the classroom at least twice a week, providing written feedback to their colleagues. In short, TEP’s model includes professional accountability alongside market accountability and substantial outcome accountability. In its first years of operation, TEP has produced substantial positive effects on student achievement (Furgeson et al., 2014).

Second, accountability regimes should incorporate multiple measures of educational practice and educational outcomes. One weakness of outcome-based accountability is that standardized assessments do not fully measure all of the skills, knowledge, and attitudes that students should be learning. Incorporating multiple measures should mitigate the incentive to narrow the curriculum and the possible corruption of the measures. The U.S. Department of Education’s guidelines for granting flexibility from elements of NCLB allow states to use a broader range of measures for school accountability. California’s CORE districts, for example, include in their accountability regime not only test scores, but also attainment measures (such as high school graduation) and social-emotional measures (such as absenteeism, suspensions, and measures of non-cognitive skills).²

Similarly, multiple measures of professional practice should be incorporated. As Ferguson and Danielson (2014) have found, instructional practices associated with growth in test scores are not identical to instructional practices associated with some favorable social-emotional outcomes. Moreover, multiple measures of practice permit the accountability regime to respond to multiple constituencies. Student and parent perspectives on teachers provide additional accountability that can complement peer and supervisor perspectives. Similarly, evaluations of school principals are incomplete without systematic input from teachers like that provided by the Vanderbilt Assessment of Leadership in Education (Porter et al., 2010).

Third, employing multiple forms of accountability and multiple measures of practice and outcomes helps to create a complete system that provides mechanisms to promote the improvement of practice. Transparency of practice creates opportunities for educators to improve, rich data on student outcomes help diagnose students’ needs, and rewards for success allow educators to innovate in productive ways.

In sum, improving educational effectiveness will likely require multiple accountability approaches in a package that empowers educators in the service of achieving positive student outcomes. And that will require shifting the balance toward a larger role for professional accountability.
References


Kane, T. J. (2012, Fall). Capturing the dimensions of effective teaching. Education Next, 12(4).


