

# Educational Researcher

<http://er.aera.net>

---

## Teacher Evaluation Policy and Conflicting Theories of Motivation

William A. Firestone

*EDUCATIONAL RESEARCHER* published online 27 January 2014

DOI: 10.3102/0013189X14521864

The online version of this article can be found at:

<http://edr.sagepub.com/content/early/2014/01/27/0013189X14521864>

---

Published on behalf of



American Educational Research Association

and



<http://www.sagepublications.com>

Additional services and information for *Educational Researcher* can be found at:

**Email Alerts:** <http://er.aera.net/alerts>

**Subscriptions:** <http://er.aera.net/subscriptions>

**Reprints:** <http://www.aera.net/reprints>

**Permissions:** <http://www.aera.net/permissions>

>> [OnlineFirst Version of Record](#) - Jan 27, 2014

[What is This?](#)

# Teacher Evaluation Policy and Conflicting Theories of Motivation

William A. Firestone<sup>1</sup>

Current interest in teacher evaluation focuses disproportionately on measurement issues and performance-based pay without an overarching theory of how evaluation works. To develop such a theory, I contrast two motivation theories often used to guide thinking about teacher evaluation. External motivation theory relies on economics and extrinsic incentives. Internal motivation uses psychology and intrinsic incentives. These theories and available evidence raise doubts about performance-based pay, but not the use of other extrinsic incentives. These theories also suggest that to maintain effective intrinsic incentives, policies to remove ineffective teachers should not reduce autonomy or trust among effective teachers and that evaluations should provide teachers with useful feedback and policy makers with information on the conditions that facilitate good teaching.

**Keywords:** motivation; performance assessment; policy; policy analysis; teacher assessment; teacher context

Interest has again turned to teacher evaluation, driven partly by research that confirms that good teachers enhance student learning (e.g., Hanushek & Rivkin, 2010; Rowan, Correnti, & Miller, 2006) and partly by federal policies such as the Teacher Incentive Fund (Heyburn, Lewis, & Ritter, 2010) and Race to the Top (U.S. Department of Education, 2009). The hope is that new data-based approaches will provide a firmer basis for offering formative feedback to teachers, motivating them to improve their practice, and removing the ineffective ones.

The attention now given to teacher evaluation raises many questions: how to assess teacher quality; how to design teacher evaluation policies; how to implement such policies; how these policies affect teachers and students; how teachers, students, and the public make sense of teacher evaluation; and how larger social and cultural forces shape the debates about teacher evaluation. Most research to date examines how to measure teacher quality (Baker et al., 2010; Bill and Melinda Gates Foundation, 2013; McCaffrey, Lockwood, Koretz, & Hamilton, 2003; Sanders & Horn, 1994), with a smaller body of work assessing one kind of policy, performance incentives (e.g., Springer, 2009). While research on the implementation of teacher evaluation policy is becoming available, most other aspects of teacher evaluation—including the design of teacher evaluation policy—are underrepresented in the literature.

This article draws attention to the challenges of designing teacher evaluation policy. One fundamental question is how much emphasis to give to using teacher evaluation data to reward or sanction teachers and how much to use those same data to

leverage teachers' learning opportunities to improve instruction (Bell, 2012). These options are informed by different theories of motivation. The first theory, grounded in economics, relies heavily on extrinsic incentives to motivate educators, emphasizes removing bad teachers, and advocates using differential rewards to improve teaching. The second, based on psychology, stresses intrinsic incentives and focuses on improving current teachers through capacity building using training and professional development. The idea of this second theory is to motivate teachers by giving them tools to succeed. However, research suggests that these theories are difficult to reconcile because the incentives that come with the threat of losing one's job and the promise of extra pay for high performance can undermine intrinsic incentives (Ryan & Deci, 2006). Finding a compromise will be challenging.

In this article, I present these two theories—the economics-based theory and the psychology-based theory—as ideal types, synthesizing a variety of writings to highlight the differences between them. I contrast the theories as a means to clarify their competing assumptions and recommendations and to identify more clearly their implications. I then identify the policies and practices each theory supports before describing the challenges both face. I hope this exercise will shift some attention from measurement to the design of teacher evaluation policy, ground that discussion better in social science theory, and suggest new questions about which approaches are most constructive.

<sup>1</sup>Rutgers University, New Brunswick, NJ

## Theories of Evaluation

Two streams of social science theory support different ideas about teacher evaluation. The economics-based theory, which is apparent in *Race to the Top* requirements (U.S. Department of Education, 2009) and state policies (Commissioner's Task Force on Quality Teaching in New Jersey, 2005), recommends using conventional quantitative data to distribute rewards and punishments through continued tenure and financial incentives, including rewards for working in hard-to-staff areas as well as for measured high performance (e.g., Podgursky & Springer, 2007). The focus is on extrinsic incentives. In contrast, the psychology-based theory relies on intrinsic incentives through professional development and job design that could be triggered by teacher assessment. For example, New Jersey justified its pilot teacher evaluation program in part by stating that its goals included using data on teachers and the instruments to measure teaching to clarify what constituted effective teaching, provide a shared language to describe good teaching and provide feedback to help teachers improve (New Jersey Department of Education, 2011). Both theories rely on use of incentives as a means of motivation.

### *Incentives*

Extrinsic motivation theory assumes that people respond to extrinsic incentives, including money. This theory focuses on what people should receive money for. The intrinsic motivation theory assumes that people reward themselves in response to the feedback they receive from their work. They feel good when they do certain things (Deci & Ryan, 1996; Hackman & Oldham, 1980).

*Extrinsic motivation.* Some of the challenges of distributing extrinsic incentives are explicated through principal agent theory, which covers situations in which a principal or authorizer<sup>1</sup> has the authority to demand an agent's compliance but cannot adequately monitor the agent's work. The authorizer can monitor the outcome of the work but not the agent's action. Moreover, the authorizer and the agent have different preferences. Because of this difference in preferences and inability to monitor how work is done, the authorizer cannot trust the agent to comply with the authorizer's wishes. Still, the authorizer has the authority to set the terms of employment. To gain influence, the authorizer builds incentives into the contract. What makes these incentives extrinsic is that they are conditional on the agent generating some measurable outcome. The authorizer's challenge is to define a contract in which the promised incentives overcome the agent's information advantage and yet still induce the agent to keep working within the contract (Miller, 2005). The theory does not examine why extrinsic incentives should motivate the agent. However, it offers extensive guidance on how to design incentives for various kinds of performance measured in different ways under many conditions and has been used to address a variety of problems in economics (Haubrich, 1994), organizational studies (Eisenhardt, 1989), and political science (Miller).

*Intrinsic motivation.* Several theories must be combined to develop a complete understanding of intrinsic motivation, but generally internally motivated individuals experience both

autonomy and self-efficacy. Autonomously motivated people find the activity itself so interesting that no additional incentive is needed. The opposite of the fully autonomous individual is the person performing an activity under duress. In between are interim states where people have internalized reasons for taking action (Deci & Ryan, 1996). These individuals work for personal interest without oversight or coercion, and they can also be motivated by internalizing others' goals. When these goals come from a legitimate authority (e.g., the school principal), then specific, achievable, yet challenging goals generate great effort (Locke & Latham, 2002). In addition to interest, the sense of an autonomous (or at least reasoned) choice of the goals is important for motivation. An earlier review of the working conditions that contribute to high teacher commitment, a condition similar to strong positive teacher motivation, found 10 studies showing a positive association between teacher autonomy and commitment and three that did not confirm that relationship (Firestone & Pennell, 1993).

Efficacy also helps to motivate teachers. Research on self-efficacy (Bandura, 1997) and teacher efficacy (Tschannen-Moran, Woolfolk Hoy, & Hoy, 1998) suggest that two factors combine to motivate teachers. The first is competence, the belief that the individual has the capacity to successfully carry out certain tasks. The second is expectancy, the estimate of the probability that carrying out the task will lead to the intended outcome (Vroom, 1964). When teachers develop a lesson on fractions, they have some belief that students will learn what is taught. The stronger that belief is, the higher is the teachers' expectancy. Motivation is strongest when individuals feel competent to carry out their assigned tasks and expect that doing so will have the intended effect (Bandura). Competence motivation is quite strong and has effects on outcomes as diverse as expending effort and adopting new practices (Tschannen-Moran et al.). However, efficacy assessments depend on teachers' specific assignments (Ball & Bass, 2000); the same teacher will not feel equally competent to teach ninth-grade history and first grade. Efficacy assessment will depend partly on conditions that teachers do not control (Tschannen-Moran et al.). For instance, the quality of teaching is a joint product of contributions of teachers and students interacting with materials (Bell, 2012; Cohen, Raudenbush, & Ball, 2003). Thus, according to this theory, although the teacher's competence strongly influences the expectation and actual production of success, it is not the only factor that matters.

### *Allocation*

Extrinsic and intrinsic incentive theories offer different guidance about how to allocate incentives and design evaluations.

*Extrinsic incentives.* Principal agent theory offers a theory of how different rewards offered under specific conditions will meet the goals of an authorizer and how the agent might take advantage of authorizer's limited information in spite of those rewards. Some typologies of teacher salary practices speak directly to this issue by identifying performances to be remunerated (Springer, 2009). Others identify forms of compensation (Kolbe & Strunk, 2012).

Teachers are rewarded for their knowledge, for doing extra work, for working in hard-to-staff schools or fields, and for achieving measurable objectives. The oldest form of knowledge-based pay might be the single-salary schedule, which replaced earlier, more discriminatory policies. The single-salary schedule uses credits, degrees, and time on the job as proxies for knowledge (Springer, 2009). Recent efforts have refined this approach by assessing knowledge more directly. Some use assessments such as certification by the National Board for Professional Teaching Standards. Others develop criteria such as those defined by the Interstate New Teacher Assessment and Support Consortium (InTASC). A district will then develop evaluation measures with some mix of portfolios and classroom observation to determine whether those criteria have been met. When the criteria are met, the teacher gets a bonus or a permanent salary increase (Odden & Kelley, 2002). When this knowledge increases teachers' competence—especially in newer forms of knowledge-based pay—extrinsic rewards reinforce intrinsic ones.

Another performance incentive is the career ladder, which rewards specific kinds of additional work. In some experimentation with career ladders in the 1980s, teachers received extra pay for tasks such as leading professional development, coaching beginning teachers, or curriculum development (Firestone, 1991). In addition to increasing the compensation of some teachers, this approach increased the collective knowledge shared among teachers. It was relatively easy to evaluate. Principals could monitor the contracts they negotiated because the work that received extra pay had a clear, understandable product. Here, too, extrinsic rewards reinforced intrinsic ones.

A third approach is to use incentives to recruit and retain teachers (Kolbe & Strunk, 2012). This approach can be used to get teachers to come to a district or to work in hard-to-staff schools or hard-to-staff fields. One version is to reward working in a school serving low-achieving, poor, or minority students. Another version is to pay extra to people with scarce skills—such as knowledge of mathematics, science, or special education—to work in the K–12 sector rather than elsewhere. Once one assumes that the skills in question are important, the compensated performances, such as having majored in a field or teaching in an urban school, are easy to evaluate (Springer, 2009).

The most recent variation is to compensate teachers for achieving measurable ends. This is what is usually called performance-based or merit pay. The challenge is to find the best way to measure the teacher's contribution to student learning, and two possibilities are open. The first is to use student achievement growth measures. The second is to use measures of teaching practices that, some have argued (Bill and Melinda Gates Foundation, 2013; Hill, Kaptula, & Umland, 2011; Weisberg, Sexton, Mulhern, & Keeling, 2009), offer additional useful information.

The compensation that follows from reaching these measured contributions can take many forms. Kolbe and Strunk (2012) listed adjustments to the salary schedule, ongoing compensation (pay or some form of tax waiver) above the salary schedule, bonuses, education and training incentives, in-kind incentives such as housing allotments, and retirement incentives.

Another source of variation is whether the target for the incentive is the individual teacher or a group. The theory most aligned with conventional economic theory and American ideas

about responsibility is to offer the incentives to individuals, but some have argued that such incentives would undermine the collegial relationships that improve teaching. Thus, some districts have experimented with providing incentives to whole departments, schools, or teams (Marsh et al., 2011).

The most powerful incentive may not be adjustments to compensation but access to employment itself. Removing incompetent teachers is a central tenet of the extrinsic approach. The core argument for this approach is not really about motivation. However, staying employed is a strong incentive. In this view, research shows that good teachers contribute substantially to student learning. However, research cannot predict which individuals will teach well based on individual characteristics or preservice preparation (Goldhaber & Hansen, 2010). Therefore, policy makers and administrators should worry less about selecting good teachers but use available data to remove poor teachers after teachers have had a chance to prove themselves (Weisberg et al., 2009).

In the short run, performance-based pay is expected to motivate individuals to put more effort into rewarded activities because of the reward. It should also have a long-term selection effect. As current and future teachers understand what the system recognizes, those who can engage in the rewarded behaviors should either go into teaching or stay in the field whereas those who cannot should seek employment elsewhere (Springer, 2009).

Although most attention recently has been given to using financial incentives to improve performance, they also influence retention. For that, the base salary is important. Teachers leave the field when they do not make competitive salaries or enough to maintain a decent living (Ingersoll & May, 2012; Johnson & Birkeland, 2003). There are many tensions between the goals of paying for performance and using pay to support retention. These complexities have been recognized for some time in the business literature on designing salary schemes (Lawler, 1990), but they do not appear to be central to the current discussion linking teacher evaluation to compensation.

*Intrinsic incentives.* With intrinsic incentives, individuals reward themselves; the challenge is to create the conditions to maximize those rewards for effective teachers. Evaluation is important insofar as it helps individuals reward themselves or it contributes to rewarding conditions.

Autonomy is crucial not only as a psychological state but also as a working condition. Historically, a considerable body of research suggests that teachers have had great autonomy (Corwin & Borman, 1988). In fact, at one time, some argued that teachers wanted more guidance on their goals (Lortie, 1969). Even today with the rise of extensive centralized accountability policies, teachers still report that their work is less influenced by those policies than by such traditional sources of guidance as informal feedback from their students and feedback from fellow teachers (Firestone, Nordin, Shcherbakov, Blitz, & Kirova, in press). Clear, challenging goals can provide needed clarity for teachers and typically come from two sources. Locally, setting challenging goals is a task for transformational leadership, usually from the principal or district level (Leithwood & Jantzi, 2005). At the policy level, the accountability movement has been trying to set challenging, but accomplishable, goals

for teachers through state tests and standardized curricula (Hamilton, Stecher, & Yuan, 2008).

Competence also promotes intrinsic motivation. Professional development and capacity building can help build teachers' knowledge, but how can policy makers ensure that professional development promotes improved competence? Available research suggests that effective professional development

- challenges teachers intellectually, while giving them powerful images of teaching and learning and building their pedagogical content knowledge;
- actively engages teachers in collaborative settings;
- reinforces learning through congruent learning activities that permit practice and refinement; and
- offers teachers opportunities to solve their own real instructional problems (Borko, 2004; Garet, Porter, Desimone, Birman, & Yoon, 2001; Knapp, 2003).

Competence enhances intrinsic motivation most when the individual gets feedback on performance. Ideally, this feedback is direct, clear information coming from the work itself (Hackman & Oldham, 1980). Teachers' feedback has historically come from students. When teachers get positive feedback from students, they feel motivated; without it, they feel frustrated (Hart & Murphy, 1990; Johnson, 1990). Feedback strategies have become more systematic and diverse. They include teacher-designed formative assessments (Black & Wiliam, 2009). Formal teacher evaluation can contribute through both measures of student performance and the structured administrative observation tools that were initially intended to provide nonevaluative feedback for teachers (Danielson & McGreal, 2000). From an intrinsic incentives perspective, feedback can help teachers clearly recognize their accomplishments or offers guidance so they can enhance their instructional competence as long as it does not constrain their autonomy.

Other resources can increase teachers' expectancy of success by creating the conditions that allow them to demonstrate their competence (Firestone & Pennell, 1993). One such condition is an orderly environment. Teachers are more motivated in schools that are orderly and not overly punitive (Firestone & Rosenblum, 1988; Garet et al., 2001; Kushman, 1992). Poor school discipline is a major reason for teacher attrition (Ingersoll & May, 2012; Johnson & Birkeland, 2003). Four other conditions help teachers experience success:

- *administrative support* in the form of a consistent environment where roles are clear, rules are regularly enforced, and fairness is assured;
- *adequate physical facilities* such as ceilings that do not leak, doors that lock, and sufficient functional student desks;
- *adequate instructional materials* such as books, supplies, reading kits, and computers; and
- *workloads* that allow teachers to prepare lessons and monitor student work (Firestone & Pennell, 1993).

These are influenced by class size, number of preparations, non-teaching assignments, and available funding.

None of these conditions that allow teachers to demonstrate their competence can be assumed in every school in the country. Without them, even competent teachers cannot show what they can do, but with them, only competent teachers will experience success.

## Challenges

Using these theories to guide practice is not straightforward. The challenges to implementing extrinsic incentives are currently being clarified through research on teacher evaluation, performance incentives, and other policies that rely heavily on those incentives. Although intrinsic incentives have received less explicit attention, this approach also has important problems. Finally, because of the tensions between extrinsic and intrinsic incentives, it is sometimes difficult to design policies to integrate both kinds of incentives.

### *Challenges to Extrinsic Incentives*

Because extrinsic incentives require authorizers to distribute rewards and sanctions, extrinsic incentives create substantial measurement challenges in designing systems to operationalize those distribution rules. The strictly methodological issues related to value-added assessment are being analyzed and debated by researchers (e.g., Baker et al., 2010; Hill et al., 2011). Researchers have given less attention to the probability that most teachers will not be assessed using value-added data because their students rarely take assessments amenable to generating value-added scores in fields like art or even social studies. In New Jersey, Department of Education staff have reported that as few as 20% of teachers will be assessed using student growth data from state tests. Approaches that combine student growth and teaching practice data are getting more attention, and their strengths and weaknesses are just being clarified (Bill and Melinda Gates Foundation, 2013).

Other problems relate to mismatches between what the authorizer wants, what the incentive system rewards, and what the agent tries to accomplish. One problem is multitasking where the authorizer has more goals than are actually rewarded (Burgess & Ratto, 2003). In education, test-based rewards usually focus on mathematics and language arts, but the public expects students to master several other subjects as well as capacities that are more difficult to test such as problem solving and responsibility. Even before the recent interest in teacher evaluation, observers documented that untested skills were being deemphasized (Hamilton et al., 2008). A related problem occurs when teachers obtain the incentives without necessarily achieving the policy makers' ends (Burgess & Ratto). Responses as diverse as teaching to the test and teacher cheating have all been documented even before performance-based pay became more prevalent (Booher-Jennings, 2005; Brown & Clift, 2010; Hamilton, Berends, & Stecher, 2005). These dysfunctional aspects of extrinsic incentives may become more common as the incentives get larger.

The evidence on performance incentives in the public sector raises significant questions about their utility for motivating teachers. Reviews of the literature in the United States and the

United Kingdom suggest that public sector organizations must typically cope with many goals, not all of which can be measured; that most middle- and upper-level civil servants (including teachers) have so much discretion and so many goals to achieve that it is hard to design effective incentive systems; that systems work better when agents believe they can trust the authorizers operating the system; that typically agents learn how to “game” the system, which requires constant adjusting of incentives in challenging ways; and that across all public sectors, performance-based incentives work best among lower-level employees and in medicine (Burgess & Ratto, 2003; Heinrich & Marschke, 2010; Perry, Engbers, & Jun, 2009). Though not impossible to design, effective public sector performance incentive systems are rare and unstable (Burgess & Ratto).

Recently, three studies from the National Center on Performance Incentives used randomized assignment to assess the effects of performance-based pay (i.e., pay linked to test scores) on student achievement in schools in Nashville (Springer et al., 2011), New York City (Marsh et al., 2011), and Round Rock Independent School District in Texas (Springer et al., 2012). None of these studies found significant effects of performance-based pay on student learning. None of the programs either motivated teachers as extrinsic motivation theory would predict or undermined motivation as the critics of such policies expected (Yuan et al., 2012). Moreover, a recent National Research Council report (2011) concluded that test-based incentive programs are not making American students more competitive and that we need to know much more about such programs before investing substantially in them. Although these negative results may stem from specific design features (e.g., the size of the reward, the mix of individual and group rewards, or other factors), this pattern of findings is not encouraging.

### *Challenges to Intrinsic Incentives*

Less attention has been paid to how facilitators of intrinsic incentives are distributed. Professional development, leadership, books, and supplies are the mundane regular stuff of schooling. Most would not seem to require much new knowledge. Yet, the conditions that promote teacher commitment are often hard to find. For instance, good professional development is extremely difficult to scale up. Borko's (2004) findings about effective professional development were based on existence proofs in single sites. When she looked for high-quality professional development projects implemented with integrity across multiple sites, she found none. A survey of New Jersey teachers found that most experienced one-shot professional development episodes that offered little content knowledge and did not address the problems they faced (Firestone & Hirsch, 2006).

Teachers still leave the field at debilitating rates because of inadequate supplies and poor support in matters of discipline (Ingersoll & May, 2012). The current debate about teacher evaluation has not addressed the problem of teacher retention. Teacher evaluation could help address these “opportunity to teach” conditions, but that has not been part of the discussion.

Ultimately, intrinsic incentive theory has two limitations. One is when teachers' preferences do not align with authorizers' preferences. The theoretical emphasis on autonomy suggests that

the agent decides what is rewarding. The authorizer's influence comes through persuasive goal setting or hiring those who share the school's goals, but neither approach is that reliable (Locke & Latham, 2002; Pendergast, 2008). The organizational literature suggests that diffuse goals are common in schools, and building goal consensus to the level where authorizers and agents agree is challenging (Weick, 1976). A more difficult problem may be that although individual teacher competence is rewarding, incompetence alone may not reduce incentives enough to induce people to leave teaching. Intrinsic incentive theory and the more grounded approaches to developing teachers' capacity provide useful guidance about how to help teachers—including those whose students have not been learning the curriculum—develop instructional competence. However, in some cases, even after attempting remediation, a few teachers do not improve sufficiently, and their children suffer as a result. Research on intrinsic incentives, professional development, and capacity development more generally offers no guidance for how to handle these cases.

### *Challenges to Combining Incentives*

Available research suggests three reasons why combining extrinsic and intrinsic rewards in one policy may be especially difficult. First, psychological research suggests that extrinsic rewards sometimes drive out intrinsic ones. The research on the interaction of extrinsic and intrinsic rewards is mixed (Cameron & Pierce, 1994; Ryan & Deci, 2006). Still in situations where rewards are tangible and predictable—as they are in merit pay programs—they undermine the autonomy necessary to support intrinsic rewards (Lepper & Henderlong, 2000). Such contradictions substantially increase the complexity of the reward design challenge.

Second, the state accountability tests used in many extrinsic incentive programs are not optimal tools to give teachers feedback that enhances their sense of competence. Teachers' use of such data is most constructive when the data are safe (i.e., information will not be used to reward or punish), when it is delivered quickly, and when it is fine grained enough to help teachers understand the learning challenges their students face, conditions rarely met by state achievement tests (Jennings, 2012; Supovitz, 2012). It is hard to design central assessments that monitor the system, distribute extrinsic incentives, and create intrinsic ones (Weiss, 2012).

Finally, the time required to collect the information to reliably allocate extrinsic incentives competes with the time administrators need to create the working conditions for teacher efficacy. Administrators cope with the increased demands of teacher observation for evaluation in several ways. For instance, they reduce the time they spend supporting teachers who need more support, they provide less well documented feedback, and they offload tasks to others (Curtis, 2012; Firestone et al., 2013; Milanowski & Kimball, 2003).

Some challenges of integrating intrinsic and extrinsic incentives are illustrated by the Teacher Advancement Program (TAP). A schoolwide reform with more than a decade's history, TAP includes a career ladder component with mentor and master teachers, enhanced opportunities for professional growth and learning, accountability for teaching practice and student

growth, and performance-based incentives, that is, a mix of both intrinsic (enhanced work variety, increased competence) and extrinsic (pay for performance) incentives (Teacher Advancement Program Foundation, n.d.). Yet, the more rigorous evaluations find small to no improvement in student achievement in TAP schools even after as much as 4 years of implementation and modest but inconsistent effects on teacher retention (Glazerman & Seifullah, 2012; Springer, Ballou, & Peng, 2008). On the other hand, there are indications that some of the career ladders implemented in the 1980s to pay teachers for extra work while offering opportunities to improve their competence and increase expectancy for success did enhance teachers' motivation (Firestone, 1991). These findings also suggest that the specifics of program design may make a great difference.

## Conclusion

Although a great deal more research is needed in several areas of teacher evaluation, some policy and practice issues need to be addressed. Foremost, available evidence is discouraging about the efficacy of incentive programs using performance-based pay such as bonuses or salary increases for meeting measured targets. At best, constant tinkering is required to make these programs work with complex, public sector jobs. Although performance-based pay may help in rare instances where management and labor can agree on a plan so that teacher autonomy is less threatened, it seems as likely to undermine the intrinsic incentives available for teachers as to increase the extrinsic ones. Even if one agrees with Hess (2004) that a major realignment of incentives is needed to reform education, performance-based pay will generate more problems than possibilities.

Other extrinsic incentives, however, can be helpful. These incentives include recruitment bonuses for hiring people in hard-to-staff fields or schools. Market-based incentives are much more common in the private than the public sector (Lawler, 1990). The full array of knowledge-based pay and pay for additional work have not received due attention and may provide promising ways to offer extrinsic incentives (Firestone, 1991; Odden & Kelley, 2002).

Removing incompetent teachers—another extrinsic incentive—is more complex. There is quite a bit of evidence that a few teachers are holding back American students, especially in the least advantaged schools (Weisberg et al., 2009), and the issue is on the agenda for every state with Race to the Top funds. The literature on measuring teacher quality is growing substantially (Baker et al., 2010; Bell, 2012; Bill and Melinda Gates Foundation, 2013; Curtis, 2012; Hill et al., 2011). However, these evaluation programs will only improve teaching where the programs are introduced with a great deal of trust between agents and authorizers and where programs designed to remove bad teachers do not undermine the supports and autonomy good teachers have to demonstrate their competence.

Feedback—the main source of intrinsic incentives coming from teacher evaluation—has been an underanalyzed aspect of teacher evaluation. Midlevel state officials, some district leaders, and some measurement experts continue to hope that state tests can be used formatively. The hope for the extensive effort some

states are putting into administrator observation is similar. Although new tests aligned with the Common Core State Standards offer some promise, state testing will always have policy makers and the public as its main client. Teachers need quick, fine-grained feedback that does not encourage gaming the system, and such feedback seems unlikely to come from the state. It remains to be seen how helpful mass observation data will be. Are there enough skilled observers with enough time to provide useful feedback and still do the other things teachers need done to teach well? Will any teacher evaluation system provide data on opportunities to teach provided by a school, district, or state, as well as teachers' and students' strengths and weaknesses? Providing teachers with extensive formative data is a promising idea, but will the benefits outweigh the opportunity costs, and will data-based professional development be any better than the many approaches tried before?

Finally, the whole discussion of teacher evaluation needs a more encompassing framework. Analysis needs to map backwards from a conception of healthy teacher evaluation that signals desired educational ends, removes poorly performing teachers, provides all teachers with information on how to improve their practice, and provides administrators and policy makers with data on how to improve conditions for teaching. Because teacher evaluation contributes to the selective retention and removal of teachers, it is fundamental to human capital management in education. Ultimately, an adequate approach to teacher evaluation needs effective measurement, but it also requires a clear understanding of how such policies allocate the incentives that motivate teachers.

## NOTES

I want to thank Melinda Mangin, Peter Youngs, and the editors and reviewers of *Educational Researcher* for their helpful comments on this paper.

<sup>1</sup>I use the term *authorizer* to refer to what economists and political scientists call a *principal* so as not to confuse this more abstract individual with a school principal. The language of principal-agent theory is very unfortunate in K–12 education.

## REFERENCES

- Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., & Shepard, L. A. (2010). *Problems with the use of student test scores to evaluate teachers*. Washington, DC: Economic Policy Institute.
- Ball, D. L., & Bass, H. (2000). Interweaving content and pedagogy in teaching and learning to teach: Knowing and using mathematics. In J. Boaler (Ed.), *Multiple perspectives on the teaching and learning of mathematics* (pp. 83–104). Westport, CT: Ablex.
- Bandura, A. (1997). *Self-efficacy in changing societies*. New York, NY: Cambridge University Press.
- Bell, C. A. (2012, September). *Validation of professional practice components of teacher evaluation systems*. Paper presented at the 14th annual Reidy Interactive Lecture Series, Boston, MA.
- Bill and Melinda Gates Foundation. (2013). *Ensuring fair and reliable measures of effective teaching: Culminating findings from the MET Project's three-year study*. Seattle, WA: Author.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation, and Accountability*, 21, 5–31.

- Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas accountability system. *American Educational Research Journal*, 42, 231–268.
- Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, 33(3), 3–15.
- Brown, A. B., & Clift, J. W. (2010). The unequal effect of adequate yearly progress: Evidence from school visits. *American Educational Research Journal*, 47, 774–798.
- Burgess, S., & Ratto, M. (2003). The role of incentives in the public sector: Issues and evidence. *Oxford Review of Public Policy*, 19, 285–300.
- Cameron, J., & Pierce, W. D. (1994). Reinforcement, reward, and intrinsic motivation: A meta-analysis. *Review of Educational Research*, 64, 363–423.
- Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis*, 25, 119–142.
- Commissioner's Task Force on Quality Teaching in New Jersey. (2005). *Quality teaching in New Jersey*. Trenton, NJ: New Jersey Department of Education.
- Corwin, R. G., & Borman, K. M. (1988). School as workplace: Structural constraints on administration. In N. J. Boyan (Ed.), *Handbook of research on educational administration* (pp. 209–238). New York, NY: Longman.
- Curtis, R. (2012). *Building it together: The design and implementation of Hillsborough County public schools' teacher evaluation system*. Washington, DC: Aspen Institute.
- Danielson, C., & McGreal, T. G. (2000). *Teacher evaluation to enhance professional practice*. Alexandria, VA: ASCD.
- Deci, E. L., & Ryan, R. M. (1996). Need satisfaction and the self-regulation of learning. *Learning & Individual Differences*, 8, 165–184.
- Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *Academy of Management Review*, 14, 57–74.
- Firestone, W. A. (1991). Merit pay and job enlargement as reforms: Incentives, implementation, and teacher response. *Educational Evaluation and Policy Analysis*, 13, 269–288.
- Firestone, W. A., Blitz, C. L., Gitomer, D. H., Gradinarova-Kirova, D., Shcherbakov, A., & Nordin, T. L. (2013). *Year 1 report: New Jersey teacher evaluation pilot program*. New Brunswick, NJ: Rutgers Graduate School of Education.
- Firestone, W. A., & Hirsch, L. S. (2006). *A formative evaluation of New Jersey's professional development requirements for teachers: Year 5*. New Brunswick, NJ: Center for Educational Policy Analysis.
- Firestone, W. A., Nordin, T. L., Shcherbakov, A., Blitz, C. L., & Kirova, D. (in press). *New Jersey teacher evaluation: Rutgers Graduate School of Education, Year 2 final report*. New Brunswick, NJ: Center for Effective School Practices.
- Firestone, W. A., & Pennell, J. R. (1993). Teacher commitment, working conditions, and differential incentive policies. *Review of Educational Research*, 63, 489–529.
- Firestone, W. A., & Rosenblum, S. (1988). Building commitment in urban high schools. *Educational Evaluation and Policy Analysis*, 23, 285–300.
- 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, 915–946.
- Glazerman, S., & Seifullah, A. (2012). *An evaluation of the Chicago teacher advancement program (Chicago TAP) after four years*. Washington, DC: Mathematica Policy Research.
- Goldhaber, D., & Hansen, M. (2010). Implicit measurement of teacher quality: Using performance on the job to inform teacher tenure decisions. *American Economic Review*, 100, 250–255.
- Hackman, J. R., & Oldham, G. R. (1980). *Work redesign*. Reading, MA: Addison-Wesley.
- Hamilton, L., Berends, M., & Stecher, B. (2005). *Teachers' responses to standards-based accountability*. Santa Monica, CA: RAND.
- Hamilton, L. S., Stecher, B. M., & Yuan, K. (2008). *Standards-based reform in the United States: History, research, and future directions*. Santa Monica, CA: Center on Education Policy, RAND.
- Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *American Economic Review*, 100, 267–271.
- Hart, A. W., & Murphy, M. J. (1990). New teachers react to redesigned teacher work. *American Journal of Education*, 98, 224–250.
- Haubrich, J. G. (1994). Risk aversion, performance pay, and the principal-agent problem. *Journal of Political Economy*, 102, 258–276.
- Heinrich, C. J., & Marschke, G. (2010). Incentives and their dynamics in public sector performance management systems. *Journal of Policy Analysis and Management*, 29, 183–208.
- Hess, F. M. (2004). Common sense school reform. *American Experiment Quarterly*, 7, 16–44.
- Heyburn, S., Lewis, J., & Ritter, G. (2010). *Teacher incentive fund grantees*. Nashville, TN: National Center on Performance Incentives, Vanderbilt University.
- Hill, H. C., Kaptula, L., & Umland, K. (2011). A validity argument approach to evaluating teacher value-added scores. *American Educational Research Journal*, 48, 794–831.
- Ingersoll, R. M., & May, H. (2012). The magnitude, destinations, and determinants of mathematics and science teacher turnover. *Educational Evaluation and Policy Analysis*, 34, 435–464.
- Jennings, J. (2012). The effects of accountability system design on teachers' use of test score data. *Teachers College Record*, 114(11), 1–23.
- Johnson, S. M. (1990). *Teachers at work: Achieving success in our schools*. New York, NY: Basic Books.
- Johnson, S. M., & Birkeland, S. E. (2003). Pursuing a "sense of success": New teachers explain their career decisions. *American Educational Research Journal*, 40, 581–617.
- Knapp, M. S. (2003). Professional development as a policy pathway. In R. E. Floden (Ed.), *Review of research in education* (pp. 109–157). Washington, DC: AERA.
- Kolbe, T., & Strunk, K. O. (2012). Economic incentives as a strategy for responding to teacher staffing problems: A typology of policies and practices. *Educational Administration Quarterly*, 48, 779–813.
- Kushman, J. W. (1992). The organizational dynamics of teacher workplace commitment: A study of urban elementary and middle schools. *Educational Administration Quarterly*, 28, 5–42.
- Lawler, E. E. (1990). *Strategic pay*. San Francisco, CA: Jossey-Bass.
- Leithwood, K., & Jantzi, D. (2005). A review of transformational school leadership research 1996–2005. *Leadership and Policy in Schools*, 4, 177–199.
- Lepper, M. R., & Henderlong, J. (2000). Turning "play" into "work" and "work" into "play": 25 years of research on intrinsic versus extrinsic motivation. In C. Sansone & J. Harackiewicz (Eds.), *Intrinsic and extrinsic motivation: The search for optimal motivation and performance* (pp. 257–307). San Diego, CA: Academic Press.
- Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, 57, 705–717.
- Lortie, D. C. (1969). The balance of control and autonomy in elementary school teaching. In A. Etzioni (Ed.), *The semi-professions and their organization*. New York, NY: Free Press.
- Marsh, J. A., Springer, M. G., McCaffrey, D. F., Yuan, K., Epstein, S., Koppich, J., . . . Peng, A. (2011). *A big apple for educators: New York City's experiment with schoolwide performance bonuses: Final evaluation*. Santa Monica, CA: RAND.

- McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability*. Santa Monica, CA: RAND.
- Milanowski, A., & Kimball, S. M. (2003). *The framework-based teacher performance assessment systems in Cincinnati and Washoe* (No. TC-03-07). Madison, WI: Consortium for Policy Research in Education.
- Miller, G. J. (2005). The political evolution of principal-agent models. *Annual Review of Political Science*, 8, 203–225. doi:10.1146/annurev.polisci.8.082103.104840
- National Research Council. (2011). *Incentives and test-based accountability in education*. Washington, DC: National Academies Press.
- New Jersey Department of Education. (2011). *Notice of grant opportunity: Excellent Educators for New Jersey (EE4NJ) pilot program teacher effectiveness program*. Trenton, NJ: Author.
- Odden, A., & Kelley, C. (2002). *Paying teachers for what they know and do: New and smarter compensation strategies to improve schools* (2nd ed.). Thousand Oaks, CA: Corwin.
- Pendergast, C. (2008). Intrinsic motivation and incentives. *American Economic Review*, 98, 201–205.
- Perry, J. L., Engbers, T. A., & Jun, S. Y. (2009). Back to the future? Performance-related pay, empirical research, and the perils of persistence. *Public Administration Review*, 69, 39–51.
- Podgursky, M. J., & Springer, M. G. (2007). Teacher performance pay: A review. *Journal of Policy Analysis and Management*, 26, 909–950. doi:10.1002/pam.20292
- Rowan, B., Correnti, R., & Miller, R. J. (2006). What large-scale survey research tells us about teacher effects on student achievement: Insights from the *prospects* study of elementary schools. *Teachers College Record*, 104, 1525–1567.
- Ryan, R. M., & Deci, E. L. (2006). Self-regulation and the problem of human autonomy: Does psychology need choice, self-determination, and will? *Journal of Personality*, 74, 1557–1586. doi:10.1111/j.1467-6494.2006.00420.x
- Sanders, W. L., & Horn, S. (1994). The Tennessee value-added assessment system (TVAAS): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education*, 8, 299–311.
- Springer, M. G. (2009). Rethinking teacher compensation policies: Why now, why again? In M. G. Springer (Ed.), *Performance incentives: Their growing impact on American K–12 education* (pp. 1–21). Washington, DC: Brookings Institution.
- Springer, M. G., Ballou, D., Hamilton, L. S., Vi-Nhuan, L., Lockwood, J. R., McCaffrey, D. F., . . . Stecher, B. (2011, March). *Teacher pay for performance: Experimental evidence from the project on incentives in teaching (POINT)* [Abstract]. Paper presented at the Spring 2011 Society for Research on Educational Effectiveness (SREE) Conference, Washington, DC. Retrieved from <https://www.sree.org/conferences/2011/program/downloads/abstracts/136.pdf>
- Springer, M. G., Ballou, D., & Peng, A. (2008). *Impact of the teacher advancement program on student test score gains: An independent appraisal*. Nashville, TN: National Center on Performance Incentives.
- Springer, M. G., Pane, J., Vi-Nhuan, L., McCaffrey, D. F., Burns, S. F., Hamilton, L. S., & Stecher, B. M. (2012). Team pay for performance: Experimental evidence from the round rock pilot project on team incentives. *Educational Evaluation and Policy Analysis*, 34, 367–390.
- Supovitz, J. A. (2012). Getting at student understanding—The key to teachers' use of test data. *Teachers College Record*, 114(11), 1–29. Teacher Advancement Program Foundation. (n.d.). *Understanding the teacher advancement program*. Santa Monica, CA: Author.
- Tschannen-Moran, M., Woolfolk Hoy, A., & Hoy, W. K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research*, 68, 202–248.
- U.S. Department of Education. (2009). *Race to the top program: Executive summary*. Washington, DC: Author.
- Vroom, V. H. (1964). *Work and motivation*. New York, NY: John Wiley.
- Weick, K. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly*, 21, 1–19.
- Weisberg, D., Sexton, S., Mulhern, J., & Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. Brooklyn, NY: The New Teacher Project.
- Weiss, J. A. (2012). Data for improvement, data for accountability. *Teachers College Record*, 114(11).
- Yuan, K., Vi-Nhuan, L., McCaffrey, D. F., Marsh, J. A., Hamilton, L. S., Stecher, B. M., & Springer, M. G. (2012). Incentive pay programs do not affect teacher motivation or reported practices: Results from three randomized studies. *Educational Evaluation & Policy Analysis*, 35, 3–22. doi:10.3102/0162373712462625

#### AUTHOR

**WILLIAM A. FIRESTONE** is Distinguished Professor of Educational Policy and Leadership at the Rutgers Graduate School of Education, 10 Seminary Place, New Brunswick, NJ 08901; [william.firestone@gses.rutgers.edu](mailto:william.firestone@gses.rutgers.edu). His interests include educational leadership, policy implementation, and the effects of a variety of policies on teaching practice and teacher motivation.

Manuscript received January 23, 2013  
Revisions received April 22, 2013,  
and September 25, 2013  
Accepted January 9, 2014