Educational Evaluation and Policy Analysis
March 2015, Vol. 37, No. 1, pp. 29–49
DOI: 10.3102/0162373714524622

Time to Pay Up: Analyzing the Motivational Potential of Financial Awards in a TIF Program

Jennifer King Rice
Betty Malen
Cara Jackson
Kathleen Mulvaney Hoyer
University of Maryland

The effectiveness of educator incentive programs rests on the assumption that the potential rewards for participants will motivate them to behave in certain ways (e.g., choose certain jobs, expend greater effort, engage in capacity-building professional development). Some researchers have examined the impact of financial incentives on teacher behaviors and work conditions, but few look inside schools to examine how current teachers interpret their rewards, how payouts affect teachers' willingness to participate in these programs, and how incentives might be structured to motivate teachers. This mixed-methods study of one Teacher Incentive Fund–supported program addresses that gap. We draw on expectancy and goal setting theories to analyze teachers' reactions to financial rewards and how their reactions may shape the motivational potency of the incentives.

Keywords: *educator incentives, teacher policy, teacher effectiveness, expectancy theory*

Fueled in part by the Teacher Incentive Fund (TIF), a federal grant program that promotes the implementation of incentive pay in local contexts, compensation reforms have become prominent strategies for improving human capital in schools. Many compensation reforms assume that financial incentives matter to current and prospective teachers and motivate them to behave in certain ways (e.g., choose certain jobs, expend greater effort, engage in capacity-building professional development). Their underlying theory is that incentives will increase human capital in schools (a) by motivating current teachers to take particular assignments and improve their performance (U.S. Department of Education, 2012; Yuan et al., 2013) and (b) by attracting stronger teachers into the profession and encouraging ineffective teachers to leave (Barnett & Ritter, 2008; Hanushek & Lindseth, 2009). This article focuses on the potential of voluntary financial incentive programs to motivate current teachers.

Scholars have examined the impact of financial incentives on teacher behaviors (e.g., instructional practices, level of effort, job decisions) and on work conditions (e.g., job stress, collegiality; C. Clotfelter, Glennie, Ladd, & Vigdor, 2008; C. Kelley, Heneman, & Milanowski, 2002; J. A. Marsh et al., 2011; U.S. Department of Education, 2012; Yuan et al., 2013). In efforts to understand the mixed findings, some have underscored the importance of unpacking how teachers view the reward structure (e.g., J. A. Marsh et al., 2011), but few look inside schools to examine how current teachers interpret their rewards, to gauge how their payouts affect their willingness to participate in these programs, and to explain the conflicting evidence about the motivational power of the incentives. We address that gap by analyzing teachers' views of financial rewards and how their views may shape the motivational potency of the incentives.

Scholars have recognized that expectancy theory, at times augmented by components of goal setting theory, provides an instructive conceptual framework for analyzing teacher reactions to incentive programs (C. Kelley et al., 2002; Yuan et al., 2013). Consequently, this article draws on an adaption of expectancy theory that integrates concepts from goal setting theory to analyze data from a mixed-methods study of the implementation of one TIF-supported financial incentive program, the Financial Incentive Rewards for Supervisors and Teachers (FIRST) program in the Prince George's County Public School System (PGCPS). We examine payout distribution patterns, describe how current teachers view their payouts and why they view them that way, and document how payouts may influence ongoing participation in these programs. While the theory of action underlying incentive programs posits that the rewards will prompt (presumably effective) teachers who receive large awards to remain in the profession and (presumably ineffective) teachers who receive low rewards to improve or depart, our study does not address that issue because the voluntary nature of FIRST made it possible for teachers to "opt out" of the incentive program without leaving the profession, the district, or even their schools. However, like several recent studies, our analysis provides evidence and generates insights about the power of the financial rewards offered to teachers and the design features that influence their motivational potency (J. A. Marsh et al., 2011; Yuan et al., 2013). Following a discussion of the theoretical and empirical literature on factors shaping the motivational potency of financial awards, we describe the FIRST program and the data sources and methods used for this analysis. We present our findings and discuss their implications for theory, policy, and future research.

Educator Incentives: Theoretical Perspectives and Empirical Evidence

Expectancy theory helps explain individuals' motivation to engage in certain behaviors, particularly those that they perceive will lead to rewards that they value (Heneman, 1998; Vroom, 1964). It models motivation as a function of three factors: (a) *valence*, that is, the intrinsic attractiveness of the expected rewards to the individual; (b) *expectancy*, that is, the perceived likelihood that effort will lead to performance; and (c) *instrumentality*, that is, the perceived likelihood that performance will lead to desired rewards. Taken together, these elements provide a general framework for examining the potential of an incentive program to motivate workers.

Applications of this theory to education have often incorporated insights from goal setting theory, which suggests that the assignment of specific and reasonable goals may increase individuals' motivation and enhance their performance (Bryan & Locke, 1967; Locke & Latham, 2006). Goal setting theory recognizes program fairness as well as goal attainability as important mediating factors. Taken together, studies drawing on these two theories suggest that (a) awards must have high valence, meaning they must be salient and sizable enough to appeal to teachers; (b) awards must address expectancy and instrumentality (i.e., attainability) by demonstrating understandable and credible connections between work, performance, and reward; and (c) goals, measures, and awards must be perceived as fair (Hatry, Greiner, & Ashford, 1994; Heneman, 1998; Odden, Kelley, Heneman, & Milanowski, 2001; Springer et al., 2008; U.S. Department of Education, 2012; Yuan et al., 2013). Drawing on these theoretical and empirical foundations, we crafted a conceptual framework (Figure 1) to analyze teachers' responses to financial incentives in the FIRST program. In the following paragraphs, we synthesize research on the factors included in this framework.

Valence of Financial Awards

Empirical studies of educator incentive pay programs indicate that the salience and size of the awards may shape how educators view their valence and mediate the motivational power of financial initiatives (J. A. Marsh et al., 2011; U.S. Department of Education, 2012; Yuan et al., 2013). Existing evidence on salience of financial rewards for educators suggests that money matters. It is one of multiple salient factors that influence teachers' decisions about where to work, whether to remain in the profession, and what kinds of continuing education to pursue (Goldhaber & Brewer, 1997; Gritz & Theobald, 1996; Guarino, Santibañez, & Daley, 2006;

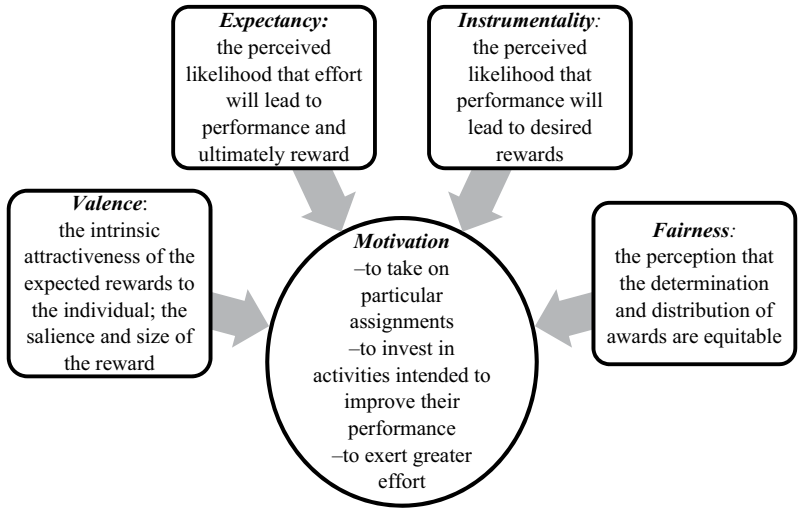

FIGURE 1. *Expectancy theory applied to financial incentives for educators.*
Source. Adapted from Vroom (1964).

Ladd, 2011; Murnane & Olsen, 1990; Rice, 2003; Snyder & Dillow, 2011). However, the mixed evidence on the impact of financial incentives on teacher performance suggests that money alone may not be sufficient to motivate teachers to alter performance. For example, research has documented that teachers may be motivated more by helping students learn than by monetary incentives (Heneman, 1998; Malen et al., 1987). Perhaps that is one reason scholars have had difficulty establishing a dependable link between financial incentives and improved performance (J. A. Marsh et al., 2011; U.S. Department of Education, 2012; Yuan et al., 2013).

The valence of financial incentives also may be related to the size of the rewards. A recent evaluation of multiple TIF programs found that educators in programs with higher payouts were more likely to report being motivated by the financial awards (U.S. Department of Education, 2012). Research on a pay-for-performance program in Texas found that teachers receiving no bonus or a relatively small award were significantly more likely to leave the school, while those receiving a large award were less likely to depart (Springer et al., 2008). Drawing on work in a variety of organizational settings, Lawler (1990) argued that "in order to be motivating, changes in pay ought to be 10 to 15 percent [of base salary] rather than the often seen 2 to 6 percent" (p. 17). Research in education suggests that awards less than 2% of base salary are not sufficient to motivate teachers, but those that exceed 4% may motivate the desired behaviors (U.S. Department of Education, 2012). Research cautions, however, that large performance rewards can have perverse effects. For example, C. T. Clotfelter and Ladd (1996) found evidence of educators gaming the system to obtain incentives and concluded that a major challenge in designing incentive programs is to provide a reward large enough to induce the desired behaviors but not so large that it encourages cheating.

Expectancy and Instrumentality: Connections Between Work, Performance, and Reward

According to expectancy theory, the motivational power of an incentive depends on the degree to which an individual perceives that increased effort will lead to higher performance and that higher performance will lead to the desired reward (Heneman, 1998). Research suggests that when teachers perceive rewards as attainable and linked to individual effort, the motivational potency of the incentives increases (C. J. Kelley & Finnigan, 2003; C. Kelley et al., 2002).

Efforts to enhance the attainability of the reward, however, may lead to depressed standards. In the past, to secure stakeholder support and bolster the longevity of incentive pay programs, policymakers tended to design pay-for-performance plans so most participants could attain awards (J. Marsh, 2012; Murnane & Cohen, 1986). However, the strategy may undercut key program goals, that is, to reward high performance and to prompt educators to take on challenging assignments and engage in performance-enhancing activities. In the current context, many advocates of incentive pay programs privilege ambitious goals, rigorous performance, and selective rewards over broadly attainable rewards. Others emphasize both rigorous standards and broad distribution by incorporating strategies to enhance teacher capacity to earn rewards (Heneman, 1998; U.S. Department of Education, 2012). If such efforts are successful, designers may be able to establish challenging goals that motivate educators and still maintain the stakeholder support often required to implement and sustain these programs.

Fairness

Research documents teachers' reluctance to support differentiated salary systems largely because they are not convinced awards will be fair (C. Kelley et al., 2002; Lewis & Springer, 2008; J. A. Marsh et al., 2011; U.S. Department of Education, 2012; Yuan et al., 2013). In fact, C. J. Kelley and Finnigan (2003) found that the perceived fairness of bonus programs was the strongest predictor of teachers' willingness to invest effort in the incentivized behaviors. Concerns about procedural (i.e., appropriate management of the program), substantive (i.e., valid measures of performance), and distributional (i.e., equitable allocation of awards) fairness are pronounced in the literature (C. J. Kelley & Finnigan, 2003; C. Kelley, Odden, Milanowski, & Heneman, 2000; U.S. Department of Education, 2012).

Various design features may affect the perceived fairness of educator incentive systems. For example, programs that are relatively inclusive in their eligibility tend to be seen as more fair and to generate more support than programs that are more selective (MacAllum, Wells, & Ristow, 2011; Malen et al., 1987; J. Marsh, 2012; Murnane & Cohen, 1986). Efforts to incorporate credible evaluation systems and guard against gaming the system also may enhance the perceived fairness of programs and garner support for them (C. T. Clotfelter & Ladd, 1996). Thus, the onus is on creators of performance pay initiatives to design programs that attend to the multiple dimensions of procedural, substantive, and distributional fairness.

The Study of FIRST

Our mixed-methods study of FIRST allowed us to build on the existing theoretical and empirical literature. It permitted us to examine payout distribution patterns, teacher reactions to their payouts, and the various reasons for their often mixed reactions. In this section, we describe the key features of the FIRST program, the district context, and our research methods.

The FIRST Program and Its Context

Supported in part by the federal TIF, FIRST was a voluntary district-developed pay-for-performance program aimed at recruiting, retaining, and improving the performance of educators in 42 high-need schools. Upon receipt of the TIF grant in the spring of 2007, the district dedicated a year (2007–2008) to planning for the graduated implementation of the program and a year (2008–2009) to piloting the initiative in 12 Cohort 1 schools (2 of which closed at the end of the pilot year). In 2009–2010, the district expanded the program to include an additional 10 schools in Cohort 2; in 2010–2011, the district added another 12 schools in Cohort 3. At this point, 32 schools, 65 building administrators, and approximately 600 teachers were enrolled in the FIRST program. In the fall of 2011, the district added the last cohort of 10 schools to "the FIRST family." Because this analysis focuses on payouts, we include data from the cohorts (1 and 2) that received payouts during the course of our study.

Only teachers who were "highly qualified" under the No Child Left Behind (NCLB) law and teachers or coteachers of record in a classroom were eligible for FIRST. Teacher participation varied across school sites. During the pilot year, school-level participation rates ranged from 18%

TABLE 1
FIRST Award Components

Category	Teacher	Eligibility
Student performance		
School growth over time	Up to US$2,500	All participants
Classroom value added	Up to US$2,500	Must teach in tested subject and grade and have pre- and posttest data for a classroom of students
Professional growth and contribution		
Professional development	Up to US$1,000	All participants
Leadership project	US$1,000	All participants
Evaluation		
Observations using Danielson's *Framework for Teaching*	Up to US$1,500	All participants
Hard-to-staff subject certification	US$1,500	Must be certified in and teach a hard-to-staff subject
Total award	Up to US$10,000	US$6,000–US$10,000 maximum award, depending on component eligibility

Source. Malen et al. (2009).
Note. FIRST = Financial Incentive Rewards for Supervisors and Teachers.

to 81%; in Year 2, participation rates ranged from 6% to 80% (Malen et al., 2011).

FIRST had multiple award components which the district communicated to teachers through required professional development sessions. As Table 1 illustrates, these award components enabled teachers to earn up to US$10,000. They could earn a substantial amount for realizing student achievement goals at the school and classroom levels. The school-wide awards were based on growth over time models (GOTM) that used locally established targets adjusted for schools' prior achievement levels.[1] The classroom awards were based on value-added models (VAM) that measured changes in student proficiency at the elementary and middle school levels and the relationship between actual and expected performance at the high school level.[2] Teachers could also earn awards for engaging in professional development,[3] completing and documenting leadership projects, undergoing evaluations based on the Danielson (2007) *Framework for Teaching* (FFT),[4] and teaching in a hard-to-staff subject. Only teachers who taught in tested subjects and grades and had pre- and posttest data for a classroom of students were eligible to receive a classroom value-added payout; only teachers who were certified in and taught a hard-to-staff subject were eligible to receive a payout for this component.

Because FIRST mirrored TIF's list of optional and mandatory program components, many of its central features were comparable with other early TIF programs. Like nearly all Cohorts 1 and 2 TIF grantees, FIRST offered teachers a sizable award. While some TIF projects mandated teacher participation, FIRST was voluntary. The bases for FIRST awards were similar to those found in many Cohorts 1 and 2 programs: Approximately half included observation-based evaluations and hard-to-staff assignments; approximately a third included multiple test-based measures of teacher performance; many also included professional development and leadership opportunities for participating teachers (U.S. Department of Education, 2012).

FIRST operated in an "initiative-rich," resource-strapped school district. The 18th largest district in the United States, PGCPS, at the time of our study, served approximately 130,000 students from diverse backgrounds, faced challenges in recruiting and retaining high-quality educators, and struggled to maintain high student performance across all its schools. Although the district had some exemplar schools and had recorded student achievement gains in select sites, the state had designated 58 schools (29% of the district's schools) as "in need of improvement" under NCLB. In response to these chronic

staffing and student performance challenges, the district had launched a vast array of reforms aimed at creating environments where all children experience academic success.

After launching the FIRST program, PGCPS experienced sharp reductions in revenue. For example, the district's FY 2011 operating budget of US$1.63 billion represented a net decrease of more than US$77 million from the previous year. None of the district's teachers had received salary increases for several years; in 2010–2011, all teachers were furloughed for 4 days and all administrators were furloughed for 9 days. Subsequent budgets represented deep cuts, even in the face of cost-saving steps such as school closures; reductions in central office staff, classroom positions, and support staff; elimination of select instructional programs and extracurricular activities; and increases in class size for all but the youngest students. Thus, though administrators and educators in PGCPS were working toward improving student achievement, they were doing so with shrinking resources (Malen et al., 2011).

Data Sources and Methods

This analysis focuses on one subset of findings from a larger study of the FIRST program in PGCPS. This larger study spanned two phases (August 2008–June 2009 and August 2010–July 2011) and involved the collection and analysis of multiple sources of qualitative and quantitative data that shed light on the planning, implementation, and scale-up of the program; the payouts; and the perceived effects of the program (Malen et al., 2009; Malen et al., 2011). This article draws on the payout records, a teacher survey, and case studies.

Payout Records. District officials provided a payout binder, the official record of payouts to program participants. We analyzed the "payout binder data" by using descriptive statistics that characterize the amount, range, and distribution of FIRST payouts issued after the pilot year and the second year of program implementation. We also used these data to identify the percent of the potential award individuals received and to determine award averages by school level (elementary, middle, and high schools), program component, and award eligibility.

Teacher Survey. We surveyed teachers who began participating in the FIRST program during the pilot year. The survey focused on their views of their pilot-year and second-year payouts and enabled us to depict (a) the degree to which teachers reported their payouts held valence, were commensurate with the effort they invested in the program, were linked to their performance, and were fair; and (b) how, if at all, the payouts influenced their ongoing participation in FIRST. We pretested the survey and piloted the online versions with members of our research team.

We used district email addresses, websites, and Internet searches to secure addresses for all but 27 teachers (less than 6% of the population of interest). We released the online survey in late October 2010 to 500 teachers. We sent reminders and incorporated a modest incentive (eligibility for a drawing for a US$50 gift card) to encourage completion. In the end, we obtained a response rate of approximately 56% from teachers. Data included responses from at least a third of the pilot-year participants in each of the pilot schools. Survey respondents were representative of the population of FIRST teachers in terms of school level (elementary, middle, and high school) and eligibility for program components (VAM and hard to staff). Self-reported payouts of survey respondents were higher than the average FIRST payout (US$3,732 compared with US$3,083), but a number of survey respondents chose not to report payout data.

We conducted a series of bivariate analyses to investigate more fully the relationships between educators' characteristics and their responses to the awards received. Specifically, we ran a series of chi-square tests to ascertain whether teachers' perceptions of payouts varied by school level (elementary, middle, and high school), by teachers' years of experience, and by the self-reported award amount received.

Case Studies. We conducted case studies of FIRST implementation in three schools (two elementary and one middle school) that were more, rather than less, likely to implement FIRST with reasonable fidelity due, at least in part, to reportedly effective principals, reputedly strong principal–staff relationships, and relatively high rates of teacher participation in the program. These case studies included semistructured individual

and focus group interviews with 76 FIRST teachers (78% of the participating teachers) in these sites who received pilot year and/or second-year payouts. We conducted these interviews during the 2010–2011 school year and dedicated a portion of them to an explicit discussion of the pilot- and second-year payouts. We used open-ended questions to elicit teachers' reactions to the size and salience of their pilot-year and/or second-year payouts and follow-up probes to secure their assessments of whether the pilot-year and/or second-year payouts reflected their effort, their performance, and their sense of fairness. In addition, we asked directly whether and how their payouts affected their willingness to participate in the program.

We incorporated recommended procedures to check for bias and error in the collection and analysis of interview data (Murphy, 1980; Patton, 2002). For example, we provided all study participants written assurances of confidentiality and created interview protocols to increase the likelihood that we would secure candid and comparable data across sites. We audio recorded nearly all interviews; prepared interview transcripts or detailed interview logs; identified emergent themes; developed analytic memos that enabled us to display and weigh confirming and disconfirming evidence within and across sites; corroborated and triangulated information secured from interviews, documents, observations, and the survey; asked district officials and study site principals for their candid assessment of our findings; and sought what Murphy (1980) terms "the fresh eye of a neutral colleague" as an additional check for weaknesses in our analysis that people more directly involved with the program and the research might not see.

Mixed-Methods Considerations and Limitations

The mixed-methods approach we used provided a substantial amount of survey and field data to analyze the bases and consistency of teachers' reactions to their payouts. We used survey data to document how FIRST teachers across the pilot schools throughout the district viewed their payouts in terms of their valence; the degree to which they were commensurate with the effort they invested in the program, linked to their performance, and fair; and their influence on their ongoing participation in FIRST. We used the case study data to unpack the multiple reasons for those general reactions to the payouts. Because these mixed methods secured information from teachers at different points in time, we sorted teacher comments on payouts by year (pilot and second) to ensure appropriate comparisons, and we reported our data accordingly. That is, we specify whether the comments refer to the pilot-year or the second-year payouts.

We acknowledge several limitations. First, because our analysis is confined to incumbent teachers' responses to the first two rounds of payouts in a voluntary program, it does not address teachers' more long-term responses or the potential of payouts to serve as a mechanism for altering the composition of the workforce. Second, our analysis is constrained by the low survey response rate. Third, because we selected case study schools on the basis of indicators that tilt toward favorable implementation sites rather than typical cases, the detailed explanations of teachers' reactions to the payouts may not hold for other sets of teachers. Finally, we grant that people's perceptions are not always accurately, fully, or candidly reported.

Findings

In this section, we describe findings on payout distribution patterns, teacher reactions to the payouts, and the influence of payouts on teachers' willingness to participate in the program.

The Distribution of Payouts

Perhaps due to the multicomponent award structure, FIRST teachers received a wide range of payouts across both years of the program. As shown in Table 2, their pilot-year payouts ranged from US$0 to just more than US$6,000; their second-year payouts ranged from US$0 to almost US$7,450. No FIRST teacher received the full US$10,000 in either of these years. Except for elementary school teachers, average payouts were lower in the second year than in the pilot year.

To refine our understanding of the distribution of the payouts, we examined payouts for the various components of FIRST. Figure 2 presents the average teacher payouts for each component (except professional development, which we do

TABLE 2
Range and Average FIRST Payouts for Teachers, by School Level

	Range (in US$)		Average (in US$)	
Level	2008–2009	2009–2010	2008–2009	2009–2010
Elementary	0–6,050	1,100–7,450	2,893	3,339
Middle	500–5,750	0–5,800	3,340	2,848
High school	750–4,750	0–5,800	2,949	2,241
All	0–6,050	0–7,450	3,012	2,847

Note. Award amounts do not include pay for professional development. FIRST = Financial Incentive Rewards for Supervisors and Teachers.

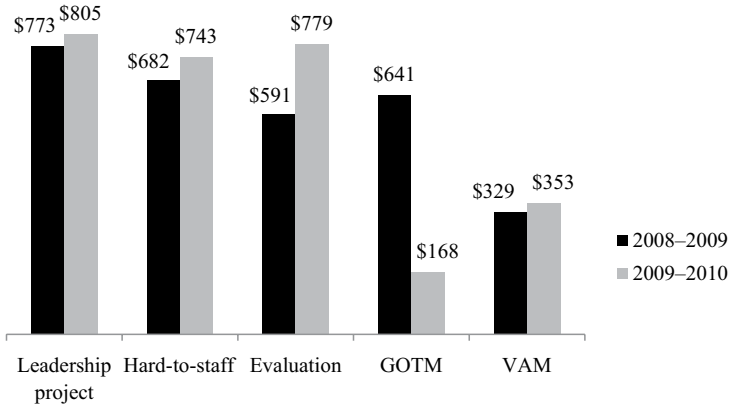

FIGURE 2. *Teachers' average payout by component.*
Note. GOTM = growth over time models; VAM = value-added models.

not include because participants received these awards throughout the year rather than in a lump sum with the rest of the FIRST payout). On average, the leadership project yielded the highest award amounts. The hard-to-staff component also generated relatively high awards; while many teachers did not qualify for this component of the award, it appears to be a powerful award-driver for those who did. Teachers' lowest average payouts were generally associated with components based on student achievement.

FIRST teachers' total payout amounts were largely a function of two factors: (a) eligibility for components and (b) achievement of student performance goals. Fewer than half of the teachers had the opportunity to earn the full US$10,000 because they simply were not eligible for all program components. As earlier noted, only teachers who met the NCLB Act's highly qualified teacher stipulation and served as the teacher of record in hard-to-staff subject areas were eligible to receive the US$1,500 associated with hard-to-staff. Likewise, only teachers in tested subjects and grades were eligible to receive the US$2,500 for the classroom value-added component of the FIRST award; a small subset of those who taught in tested subjects and grades were not eligible to receive classroom value-added awards because they were linked to fewer than five "valid students" (i.e., students for whom the district had both pretest and posttest scores). Not surprisingly, educators who were eligible for more components tended to receive higher awards. See Table 3 for comparisons of the

TABLE 3
Average Payout by Eligibility Category

Eligibility Category	Maximum Award (in US$)	Approximate Averages (in US$)	
		2008–2009	2009–2010
Eligible for all	10,000	4,225	4,643
Not hard-to-staff	8,500	3,356	3,829
Not VAM	7,500	3,491	3,200
Not hard-to-staff or VAM	6,000	2,122	1,648

Note. Averages do not include pay for professional development. VAM = value-added models.

average teacher payouts by eligibility category across the 2 years. Because student achievement measures drove a major portion of teachers' total award, student achievement was a key determinant of teachers' payouts. Because schools' success in attaining adequate yearly progress (AYP) and realizing growth targets fluctuated from year to year, school-wide student achievement measures largely accounted for the decrease in award amounts across the 2 years of payouts.

The payout distribution patterns presented here should be interpreted in light of two caveats. First, the initial determination of payout amounts was a complicated, labor- and time-intensive process that required the district and its union partners to develop a "credible process" that would hold teachers harmless for recordkeeping shortcomings, missteps in the implementation of the teacher evaluation system, and unclear expectations. As a result, pilot-year payouts often compensated for implementation errors as well as rewarded performance. Second, the actual payout amounts received by teachers were considerably smaller than reward amounts presented in this section because the payouts were taxed at a 40% rate. Most teachers did not understand that their payouts would be classified as emoluments, a designation that made teachers solely responsible for all federal, state, and local taxes on this portion of their income.

Teachers' Reactions to the Payouts

Generally speaking, teachers who participated in our study expressed mixed views about the degree to which their payouts held valence, were perceived as commensurate with the effort they invested in the program, linked to their performance, and fair.

Valence. Our case study data indicate that teachers' perceptions of the valence of the payouts varied across individuals and over time. In terms of size, nearly all teachers reported that they felt "shocked," "disappointed," and "shafted" when they received their pilot-year payouts because they were much lower than expected. A major reason for the lower-than-anticipated payments was the high tax rate that came as a surprise to almost all FIRST teachers. Only a handful of teachers reported they were "satisfied" with or "excited" by their pilot-year payouts. Teachers' reactions to their second-year payouts tended to be less negative, even though average second-year payouts were lower, because their experience with the pilot-year payouts prompted them to lower their expectations.

In terms of salience, virtually all teachers were appreciative of the opportunity to earn additional money, even if they were disappointed by the size of their payouts. With rare exception, FIRST teachers were "thankful" for their pilot-year and their second-year payouts. The words of one capture a common theme: "It's nice to have that extra little bit of money, however little it might be." Many added that economic circumstances in the district (e.g., no salary increases, furloughs) and the timing of the checks (distributed shortly before the winter holidays) enhanced the salience of the award. The prevailing view was "some money is better than no money," especially in tight economic times.

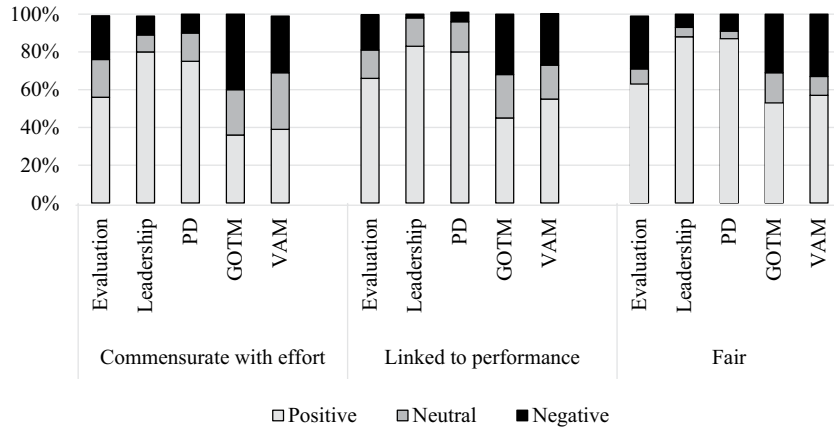

FIGURE 3. *Teachers' perceptions of pilot-year payouts, by component.*
Note. Survey questions: Please rate to what degree your payout was (a) commensurate with the effort you invested in each component of FIRST, (b) linked to your performance of each component, and (c) fair. Response options (combined into three categories): not at all, minimally, somewhat, moderately, and highly. PD = professional development; GOTM = growth over time models; VAM = value-added models; FIRST = Financial Incentive Rewards for Supervisors and Teachers.

Commensurate With Effort. In response to a survey question regarding whether FIRST prompted teachers to exert additional effort, more than three quarters of the respondents said they put in moderate to considerable additional effort into all five components of the program. As Figure 3 shows, in response to a survey question regarding whether payouts were commensurate with the additional effort invested in each component of the program, three-quarters or more of respondents rated their payouts associated with leadership projects and professional development as commensurate with effort; more than half indicated that the evaluation payouts were commensurate with effort. However, less than 40% viewed payouts based on student achievement (GOTM and VAM) as commensurate with effort.

Our bivariate analyses show that the size of the award and the level of schooling were related to teachers' perceptions of the degree to which payouts were commensurate with effort. Among survey respondents who earned US$3,500 or more, 64% viewed these payouts as commensurate with effort, but of those who earned less than US$3,500, 40% viewed the payouts based on teacher evaluations as commensurate with effort.[5] Teachers' perceptions of the evaluation component of the payout also were related to school level; elementary school teachers were more likely than middle or high school teachers to view payouts resulting from teacher evaluations as commensurate with effort, and middle school teachers were more likely than high school teachers to describe teacher evaluation payouts as commensurate with effort. Perceptions of professional development, leadership project, GOTM, and VAM awards as commensurate with effort did not vary by school level or by teachers' years of experience.

Our case study data reveal, as one might predict, that teacher judgments about the degree to which their pilot-year and second-year payouts were commensurate with their level of effort depended on the amount of additional effort they reportedly invested. In the following, we present data regarding teachers' sense of the amount of effort FIRST required and their perceptions of the extent to which specific portions of the payout reflected the effort exerted.

"It's just like free money for things that you already do" (comment on second-year payout). Despite survey responses indicating that many teachers thought they invested a considerable amount of additional effort in FIRST, our interview data suggest that a sizable group saw the program as payment for relatively minor amounts of additional effort, or for the substantial extra effort they were already investing in the school. These individuals tended to

view both the pilot-year and the second-year payouts as welcome recognition of their effort. In discussing pilot-year payouts, teachers frequently noted that "there [are] a lot of teachers who do those kinds of things and don't get paid, so it was nice to say, 'Okay, you're putting in this extra effort, we'll give you X amount of dollars.' In conversations about second-year payouts, teachers echoed that sentiment. They typically said the payout was "nice . . . we're getting a little extra carrot for some of the things we would normally do anyway."

"There's a lot of work to do and the payout is not that much" (comment on pilot-year payout). Other teachers did view FIRST as requiring considerable extra time and effort. Teachers who held this view had varying opinions regarding whether the payouts were adequate to compensate for this effort. A few teachers characterized the amount of their pilot- and second-year payouts as sufficient compensation for their effort in statements like the following:

> In each of the areas . . . we really did put in a lot of our time and energy and we exposed ourselves to our administrators and peers for a lot of peer review, critique. . . . If I was looking into [the pilot year] payout, I'd say, "Okay, I got paid this amount. That would be equivalent to $300 a month." That's okay.

More often, however, teachers who reported investing substantial additional effort indicated that across both payout years, the amount of their award was *not* commensurate with that effort. For example, in conversations about their pilot-year payouts, teachers indicated that "If you put in your effort toward this goal and [the payout] is not even half [of the maximum amount], even a child would cry." In comments about their second-year payouts, teachers reported that their payouts "did not come close" to compensating them for the extra effort they invested in their work and suggested that the payout "would be more commensurate if we got upwards near $10,000."

The degree to which teachers perceived payouts to be commensurate with effort varied by component. For instance, teachers pointed to the new FFT evaluation system as requiring considerable effort particularly during the pilot year. Most teachers who held this view indicated that their pilot-year payout for this component did not capture their effort. The words of one capture a common sentiment.

> I felt like, especially the first year. . . . [The payout for the *FFT* evaluation] was substantially less than I was expecting. I felt like I had put a lot of energy . . . [into] my observations.

During the course of our study, teachers talked about the difficulty of making AYP and meeting growth targets. While teachers did not comment directly on the level of effort needed to realize those targets, their references to the "difficult" task imply that a high level of effort would be required.

> It's very difficult to make AYP when you have so many students that have been identified [for special education services].

> To try to make that growth with them . . . in nine months' time was, you know, quite difficult.

None of our case study sites met all growth targets in either year; teachers in these schools received little compensation for their efforts to improve school-wide student achievement.

Linked to Performance. A majority of teachers viewed the various components of their payouts as linked to their performance; the only exception involved the payouts based on school-wide growth over time. As seen in Figure 3, about 80% of teachers responding to the survey rated the pilot-year payouts associated with leadership projects and professional development as linked to performance, and about two thirds indicated that the evaluation payouts were linked to performance. Regarding school-wide growth over time payouts, only 45% of our survey respondents viewed these payouts as linked to performance; just more than half said that classroom value-added payouts were linked to performance.

Perceptions of the extent to which professional development and teacher evaluation payouts were linked to performance varied across school level. Elementary school teachers were more likely than middle or high school teachers to view these payouts as linked to performance; middle school teachers rated these payouts as

linked to performance more frequently than high school teachers. Teachers' perceptions of the professional development and evaluation-based payouts did not vary by their level of experience or their award amount. Perceptions of leadership project, GOTM, and VAM awards as linked to performance did not vary by award amount, school level, or teachers' years of experience.

Our case study data indicate that the extent to which teachers viewed payouts as linked to performance was related to teachers' perceptions of whether the measures used to gauge their performance accurately captured performance. For example, while some teachers applauded the new evaluation system as a more robust, albeit imperfect, tool for assessing teacher performance, many questioned the validity of measures based on student achievement.

"For the most part it [payout] probably does [link to performance]" (comment on second-year payout). Various study participants said their pilot- and second-year payouts "reflected" their performance. Those who held this view tended to credit the FFT-based evaluation system for providing the link between payouts and performance. As seen in Figure 3, two thirds of the teachers held positive perceptions of the link between evaluation-based payouts and their performance. Our interview data corroborate that finding in statements like the following:

> Yes. It was [linked to performance]. I mean . . . our old evaluation system was basically a check list. It [*FFT*-based evaluation] was much more helpful for the teacher and that does link it directly to your teaching and how the students are doing.

> . . . especially since the pay was directly from the observation . . . it's directly from what I was doing for my students.

While teachers tended to have positive reactions to the validity of the teacher evaluation system, a subset of those teachers expressed concerns about their colleagues "staging" their performance and "gaming" the evaluation system by going "baseline" to show more growth. Those concerns intensified during the second year. To the extent that these sorts of practices occurred, the link between payout and performance would be compromised.

"Who knows what made a difference [in student performance]" (comment on second-year payout). While some teachers perceived payouts as linked to performance, a sizable minority (almost one third of survey respondents) were skeptical of the extent to which payouts related to student achievement were linked to performance. Some teachers were not convinced that the VAM could isolate the impact of one teacher on a student's performance, or capture the progress that different types of students made. Consequently, they had concerns about whether VAM payouts were linked to the performance of individual teachers. The following quotations represent prevalent responses across study years to the measurement issues embedded in calculating payouts and linking them to the performance of teachers.

> The kids have been tested on what the teacher last year taught them from March into June and then what I taught them from August to March, so you're still assessing me and basing it on what somebody else did before the summer.

> The growth that we see in special education does not necessarily translate to the MSA [Maryland State Assessment].

Some teachers who described discrepancies between their payouts and their performance noted that both the pilot and the second-year payouts were more a function of eligibility for specific components than for performance. The following quotations substantiate these views:

> You also put in a lot of work but just because of your job description [the grade and subject taught], you get less.

> My job is just as complicated and just as involved with testing, so then when you have teachers that receive money for their individual class even though you work in their classroom, and you're a part of their test scores . . . you're not tied to any classroom. So if you're not classroom based, you don't receive [the payout].

Finally, some teachers reasoned that payouts were not linked to individual teacher performance

because a large portion of the payout depends on whole-school student achievement. As one teacher stated,

> If I were to put my . . . students' scores up against any . . . [other] scores in the county and in the state, we would have made AYP . . . [but] because we didn't make [AYP] as a whole, we all had to suffer.

Fairness. As shown in Figure 3, most teachers viewed payouts as fair. Teachers were most positive about the fairness of the payouts associated with the leadership and professional development components; nearly 90% of the teachers responding to the survey rated the payouts associated with professional development and leadership project components of FIRST as fair or very fair. About two thirds of survey respondents indicated that the evaluation payouts were fair, and just more than half described the payouts based on student achievement as fair. That said, almost a third of teachers indicated that the evaluation and student achievement payouts were unfair.

Again, we ran a series of bivariate analyses to ascertain whether perceptions of fairness varied by school level, teachers' years of experience, and the size of the teachers' awards. Perceptions of leadership project and VAM awards did not vary along any of these dimensions. Teachers held similar perceptions of professional development, teacher evaluation, and GOTM payouts regardless of their level of experience, but perceptions of these payouts varied by school level; elementary teachers were more likely than middle or high school teachers to view these payouts as fair. In addition, teachers receiving higher awards reported that the evaluation and GOTM payouts were fair more often than did teachers receiving lower payouts.

While our survey data indicate the extent to which teachers throughout the district perceived pilot-year payouts as fair, our case study data reveal how teachers conceptualized fairness and how the payouts measured up to their criteria. Teachers gauged fairness in terms of the degree to which payouts were consistent with one or more of the following criteria: (a) their expectations, (b) the principle of equal pay for equal investment in and comparable quality of work, and (c) their view of the accuracy of performance measures and the attainability of standards. Below, we describe findings on teachers' perceptions of the payouts based on their broad conceptions of fairness.

"Fair because it was explained . . . " (comment on pilot-year payout). Teachers who gauged fairness in terms of the alignment between payouts and expectations talked about the degree to which payouts were perceived as consistent with what the district promised, the dependability of the award and the transparency of award calculations. On these dimensions, their reactions to both their pilot- and second-year payouts were evenly divided; some viewed the payouts as fair and others did not. For example, some teachers reported that the program and the payouts associated with it were fair because the district "explained" the program; "nothing was hidden." As one summarized this perspective,

> They outlined everything in a way that you understand exactly what goes into each component and how you get that particular amount, so I already know, "Oh I automatically won't get certain things."

Others used the same principle to render their pilot- and second-year payouts unfair. These teachers maintained that the FIRST program "was initially sold as $10,000 . . . [teachers] were told one thing and when the people got the checks they were quite upset because it wasn't what they thought it was going to be." Because some payouts, particularly pilot-year payouts, were substantially less than anticipated for reasons including but not limited to the amount of taxes withheld, teachers reported that "most of us felt kind of shafted."

Some teachers viewed fairness in terms of dependability, an orientation that may help explain the high percentages of survey respondents who reported the pilot and second-year payouts for leadership projects and professional development were fair. As the following comments illustrate, these teachers saw these components as a way to earn "easy" money and appreciated the straightforward, predictable nature of the payouts attached to these components:

> As long as you follow everything they said to do and turn it all in, you get it.

> I think from people that I've spoken to it was just one flat reward for that. So I guess, on that level it was fair

> in that way 'cause it didn't matter what you did as long as you completed it, you got a payout for it.

Teachers differed in the extent to which they viewed awards as transparent. Those who were concerned about the transparency of the award cited the lack of clarity regarding how pilot and second-year awards were calculated and the absence of warning about the high tax rates. In their words,

> You don't really know what they're using to evaluate, I mean, where the data [are] coming from. Which test, or is it a combination of tests? What kids? So it's hard to say.

> No, I didn't think it was fair—I had an issue with and I never quite understood part of the payout, but I had an issue with comparing two different sets of children [in a school that had considerable turnover in the student body].

> They offered me this money, but then you're going to take most of it away [for taxes] . . . I didn't believe it was fair.

Even teachers in their third year of participation reported that the calculation of certain awards, particularly the GOTM and VAM awards, was not clear despite district efforts to explain these calculations in required professional development sessions. This finding may help to explain some of the patterns in Figure 3. However, teachers' reactions of "shock" in response to the tax rate were less pervasive in response to second-year payouts because previous experience with the payout process reduced the element of surprise.

"Their payout is significantly more . . . even though they're both doing . . . the same work" (comment on second-year payout). Many teachers questioned the fairness of the awards because educators who had comparable responsibilities and often worked with the same children were not eligible for the program or components of it; hence, they did not have access to the full array of payouts. The following comments on pilot and second-year awards illustrate these concerns:

> I truly feel that the paraprofessionals should have the same ability to participate in this and receive the compensation for the amount of work they do, which is just as much as I am [doing].

> [In] inclusionary classes . . . you've got teacher A who's a regular ed teacher and teacher B who's a special ed teacher. They're actually teaching the same kids, yet one person, because they have the title of being a special ed teacher, their payout is significantly more than the regular ed teacher even though they're both doing . . . the same work.

> It just doesn't seem fair if you're not teaching a tested grade why they're telling you you're out a whole lump of money because they couldn't figure out a way [to measure student progress].

Other teachers expressed concerns about the uniformity of the pilot-year payouts for leadership projects in the face of variations in the effort invested and the quality of work accomplished. These teachers recognized that some teachers were rewarded for what they saw as mediocre projects; consequently, they viewed this portion of the pilot-year payout as unfair. Select comments illustrate,

> Somebody had a so-so project last year and [s/he] got this [payout]. I'm thinking, "You're going to be paid $1,000 for something and you did this kind of project?" That wasn't fair.

> We put hard work into these leadership projects . . . and other teachers would do something small for fifteen minutes and they would get their thousand [dollars].

These concerns appear to have subsided during the second year of implementation, in part because principals articulated expectations more clearly and monitored projects more closely.

"The way that the participants are evaluated, there's no way to make it fair . . . " (comment on pilot-year payout). During the course of our study, teachers expressed reservations about the fairness of both observation-based evaluations and student test scores as measures of their performance. Although most teachers in our study sites reported that their administrators rendered accurate judgments of their performance in the classroom and tended to characterize the payment associated with those evaluations as fair, a smaller subset questioned the extent to which *any* observation-based evaluation system can accurately capture teacher performance. These concerns may help explain why about a third

of the teachers reported that payouts for the evaluation component were unfair. Case study participants who viewed the evaluation-based payouts as unfair described the evaluation process as "subjective and unfair" in part because the observations were brief "snapshots" of performance.

Similarly, about a third of survey respondents indicated that the student performance measures were unfair. Our qualitative data suggest that teachers questioned the validity of the student performance models. Furthermore, many teachers did not perceive the student performance targets to be attainable given the characteristics of their student populations and the tendency for targets to become more demanding from one year to the next. Statements across implementation years illustrate these concerns:

> We really could not do a lot to control whether or not we made AYP, especially last year. We're not going to make any of our goals, so that money is down the toilet.

> But again, I'll keep on hitting on the AYP. It's harder; it is more difficult for us to hit AYP than it is [for] most other schools.

> Our school never gets AYP. That's the only problem we run into because we have such a high special ed population, you know, and it's hard for special ed students to take a test and get the scores that . . . a regular ed kid does. It sounds like an excuse but it's not really an excuse because some kids just aren't meant to take a test.

Payouts and Participation in FIRST

Because the FIRST program was voluntary for teachers and could only realize its intended goals if they chose to participate, we were interested in understanding how payouts may have affected the ongoing participation of FIRST teachers. Our data suggest that most teachers reported that their pilot-year payouts encouraged them to remain enrolled in FIRST. Not surprisingly, these responses were directly related to the size of the payouts. As shown in Figure 4, among those who received payouts greater than US$3,500, 75% reported that their payouts somewhat or strongly encouraged their continued participation; only 2% reported that their payouts discouraged participation. While more than 60% of teachers who received less than US$3,500 reported that their payouts somewhat or strongly encouraged ongoing participation, 23% reported that the payouts discouraged participation.

Furthermore, teachers who viewed the payouts as commensurate with effort, linked to performance, and fair were more likely to report that payouts encouraged their participation. School level also appeared to be related to the payouts' influence on participation. Elementary school teachers were *much* more likely than middle school and high school teachers to report that the payouts had a positive impact on their decision to remain in FIRST. Just more than 70% of elementary teachers, compared with about 40% of middle and high school teachers, reported that the payouts encouraged participation in FIRST. Teacher experience was not related to whether FIRST payouts influenced their decision to remain in the program.

Our field data confirm that teachers held a range of views regarding the impact of payouts on participation. For some teachers, the disappointment of the smaller-than-anticipated payouts combined with high tax rates discouraged participation. These teachers tended to say, "Okay, never mind." For others, the payouts were sufficient to sustain their interest in participating in FIRST. Some stated directly, "The reason I participate is so I can get paid." Over the course of our study, teachers expressed appreciation for the opportunity to earn any additional money particularly when they were facing furloughs and wage freezes. In their words,

> And especially with getting furloughed and all this other stuff, it's nice to have that extra little bit of money.

> Let's be honest, we're not getting any raises, this is our chance to really get extra [money].

> Whereas some people had been a little reluctant to join, they started 'cause there's so many cuts and we didn't get a cost of living raise and things like that so they're like at least this is something.

Most teachers reported that their participation decisions were driven by multiple considerations,

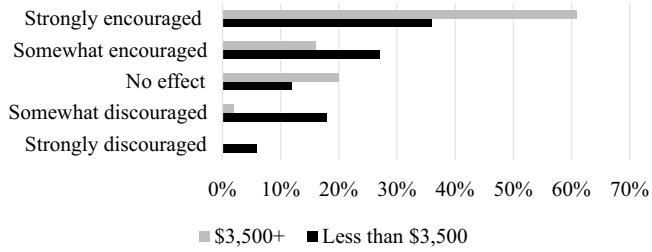

FIGURE 4. *Impact of pilot-year payouts on teachers' decisions to continue participating in FIRST.*
Note. Survey question: What influence, if any, did your FIRST payout have on your decision to participate or not participate in FIRST in subsequent years? FIRST = Financial Incentive Rewards for Supervisors and Teachers.

including but not limited to the payouts. Across the first 2 years of implementation, FIRST participants frequently reported that FIRST was "worth doing" not only for the financial return but also for the professional growth that comes with participation.

Discussion: Applying and Refining the Orienting Theory

This article addresses a major gap in the literature by providing empirical evidence on how payouts associated with an educator incentive program were distributed to teachers and how teachers reacted to their payouts. While our study is limited to a single voluntary program at a particular point in time, it generates important empirical and conceptual insights. Empirically, our study documents how teachers perceive financial incentives and how their payouts may affect their willingness to participate in these incentive programs. Conceptually, our study suggests that an expanded version of expectancy theory, augmented by concepts from goal setting theory, might strengthen our ability to analyze economic incentives in education contexts. In this section, we briefly summarize our key findings and discuss their implications for theory as well as for policy and future research.

Summary of Key Findings

While some incentive programs have allocated rewards uniformly across participants (Malen et al., 1987; J. Marsh, 2012; J. A. Marsh et al., 2011; Murnane & Cohen, 1986), other programs have sought variance. Like some other Cohorts 1 and 2 TIF programs (U.S. Department of Education, 2012), FIRST differentiated payouts. Generally speaking, the payout amounts were lower than expected; average awards tended to decline over time and no program participant earned the maximum award amount in either year. Across both years, teacher payouts were largely a function of eligibility for various components of the program (VAM, hard-to-staff) and whether schools met AYP and made growth targets; few did. As noted, pilot-year payouts often compensated for implementation errors as well as rewarded performance. Because the payouts were taxed at a 40% rate, the actual payout amounts were considerably smaller than teachers anticipated.

Teachers had a range of reactions to the payouts in terms of the valence of the rewards and the degree to which their payouts were commensurate with the effort invested in the program, the extent to which they were linked to performance, and the fundamental fairness of the awards. As one might predict, teacher reactions varied by the size of the payout; teachers receiving higher payouts tended to be more positive than those receiving smaller awards. Interestingly, our findings indicate that responses to payouts did not vary by teacher experience. They did, however, vary by school level; elementary school teachers were more positive about their payouts than middle and high school teachers. This finding warrants further study to determine the reasons for this variance.

Although teachers reported that a variety of factors influenced their ongoing participation in FIRST, most reported that payouts mattered. Both our case study and survey data suggest that the payouts' negative influence on participation may have been the result of receiving

smaller-than-anticipated awards. Payouts were more likely to encourage participation for teachers who received relatively large awards. To the extent that award size is an indication of performance, this finding suggests that incentives may attract higher performing teachers, though the voluntary nature of FIRST limits our ability to draw conclusions about the potential of incentives to alter the teacher workforce. In addition, those who viewed the payouts as commensurate with effort, linked to performance, and fair were more likely to report interest in ongoing participation in the program.

Applying an Expanded Version of Expectancy Theory

Our study confirms some of the central elements of expectancy theory, and provides evidence that helps to refine how this theory, augmented by goal setting theory, may be applied to education. In the sections below, we discuss the theoretical implications of our findings.

Valence. Expectancy theory tells us that the valence of the reward is important. Extant research has demonstrated that the salience and size of the reward may shape how teachers view the valence of awards, influence the level of support for the programs, and mediate the motivational impact of the payouts. Our study reinforces these findings. It suggests that financial rewards matter to educators, particularly in a context of tight budgets, salary freezes, and staff furloughs. The vast majority of our survey respondents indicated that they increased their effort, at least in the short term, as a result of the program—perhaps because the budget constraints of the district made the rewards particularly salient to educators as a rare opportunity to augment their salaries, and perhaps because the size of the maximum potential awards available through the FIRST program met Lawler's (1990) recommendation that the incentives should constitute at least 10% to 15% of one's salary. Our survey and case study data are consistent with the small but growing body of research that suggests that the size of the award matters; participants who received higher payouts were more likely to express an interest in ongoing participation in FIRST. Furthermore, our survey findings suggest that larger rewards were associated with more positive assessments of the rewards (in terms of their fairness and the degree to which they were commensurate with effort) and greater interest in remaining in the program. Thus, our study underscores that teacher reactions are, at least in part, a function of the size of the rewards.

No participant earned the appealing maximum payout during the first 2 years. In reality, the average award tended to be less than half of the maximum and declined over time. Many teachers were disappointed with their payouts; the unexpected and substantial tax deductions contributed to their disappointment. While it is difficult to predict the impact of the discrepancy between potential and actual rewards on the efficacy of an incentive pay program, our study revealed a range of reactions: Most teachers were grateful to be getting anything extra; many were initially frustrated, but became more resigned; and others were disenchanted with, even insulted by, the minimal payouts they received. Because our data are confined to the first two rounds of payouts, we do not know how the payouts and responses to them might change over time. While payouts might increase with improved teacher performance, they could also decrease in the face of budget constraints and competing priorities and programs. Our findings suggest that the long-term success of a program may depend on how the payout amounts and the responses to the payouts evolve over time.

Connections Between Effort, Performance, and Reward. Expectancy theory, along with the concept of attainability from goal setting theory, predicts that educators are likely to be more supportive of incentives that have dependable connections between effort, performance, and reward. Empirical evidence from our study supports this idea. Teachers in our study were most positive about the payouts they viewed as attainable and dependable; they were most negative about the payouts that they perceived to be beyond their control regardless of the amount of effort exerted. Teachers were most positive about the stipends related to the leadership and professional development components. While educators may have responded favorably to these components because they represent familiar practices in school settings, these payouts were

also attractive because FIRST teachers could count on these component payouts if they satisfied the relatively straightforward requirements. Conversely, teachers were most negative about the payouts received for FIRST components based on student achievement. Many viewed these awards as unattainable, not under the control of individual educators, and not linked with performance.

Fairness. While expectancy theory does not directly address the concept of fairness, goal setting theory suggests and research demonstrates that perceived fairness of payouts may affect educators' support for financial incentive programs (C. J. Kelley & Finnigan, 2003; MacLeod et al., 2009; J. Marsh, 2012; Springer et al., 2008; Yuan et al., 2013). Our study confirms these observations and suggests that expectancy theory should be expanded to incorporate notions of fairness identified in goal setting theory. In our study, FIRST teachers were concerned about the principle of equal pay for equal investment in and comparable quality of work and the perceived inequities associated with eligibility for the program and the various components of it. Program participants also expressed concerns about the measures used to assess their performance. The awards based on VAM were considered unfair by some on the grounds that it is impossible to isolate one teacher's "value added" and that some teachers who contribute to student learning were excluded from this part of the award because they weren't teaching a tested subject or grade level or eligible for FIRST at all. The awards based on the GOTM were suspect on the grounds that teachers' payouts were contingent on the performance of other teachers and the demographics of the students—both factors that are beyond the teachers' control. To the credit of the developers, FIRST used multiple measures (student test scores and standards-based teacher evaluations) to capture teacher performance at both the individual and school levels. However, each of those measures is subject to criticism; teachers may have been particularly critical of the student achievement measures because a third or more of the potential FIRST award was based on these measures.

Taken together, these findings suggest that perceived fairness may be more than simply a fourth factor of expectancy theory, as depicted in Figure 1. Our evidence suggests that fairness interacts with valence, instrumentality, and expectancy to shape how teachers view the payouts. For example, perceptions of fairness may affect the degree to which educators value the rewards. Concerns about fairness could intensify participants' concerns about the weak connections between effort, performance, and award, which may undercut both expectancy and instrumentality. For instance, concerns about the fairness of the measures used to determine the award could undermine expectancy and threaten instrumentality. These data suggest that fairness is a pervasive issue that must be considered in the design and implementation of educator incentive programs if they are to work as intended.

Implications for Policy and Future Research

While this formative evaluation of one TIF program was not intended to produce definitive and generalizable conclusions about the design, implementation, or impact of educator incentive programs, our findings suggest that policymakers implementing educator incentive pay programs give careful attention to the salience and size of the award; the alignment between payouts, effort, and performance; and issues of fair practice. These factors raise design considerations that may affect the potential of educator incentive programs to operate as intended. Our findings suggest that from the standpoint of teachers, the maximum award amounts must be attainable, various conceptions of fairness must be considered and addressed, performance measures must be accurate and dependable, eligibility restrictions that limit participation in the program or components of it must be justified on fairness grounds, and the rules regarding payouts (i.e., eligibility criteria, performance measures, tax rates) must be clear up front. Given the increasing policy interest in incentive pay for educators and the appealing theory of action behind this approach to reform, our study also reveals the need for more research that (a) incorporates more representative and longitudinal data to explain teachers' short-term and long-term responses to alternative combinations of financial incentives and the manner in which those responses vary by school level, teacher experience, size of payout and other factors; (b)

focuses on how financial incentives may alter the composition of the teacher workforce; and (c) continues to refine the prominent theories that guide the design of and reliance on these approaches to improving human capital in schools.

Acknowledgment

The authors also wish to acknowledge the members of the research team for their assistance with data collection for this study. These individuals include Lauren Bivona, Laurie Blaisdell, Amanda Bowsher, Karmin Cortes, Brad Coverdale, Allison De La Torre, Marisa Goldstein, Laura Hyde, Claire Jacobson, Apitchaya Pimpawathin, and Jessica Sutter.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported through a partnership between the Prince George's County Public Schools and the Department of Education Policy Studies at the University of Maryland, College Park (UMCP). The authors are grateful for support from both organizations.

Notes

1. The school-wide growth over time model was divided into making adequate yearly progress (AYP; 30%) and meeting growth targets (70%). Schools that made AYP received US$750 (30% of the US$2,500 maximum). The payout for meeting growth targets depended on the percent of targets met (0%–24% = US$0; 25%–49% = US$900; 50%–74% = US$1,300; 75%–100% = US$1,750). The district set targets for each school.

2. Teacher value added fell into four categories with payouts of US$0; US$1,200; US$1,800; and US$2,500.

3. Professional development payouts were paid on a per session basis.

4. The teacher evaluation awards were split between growth and excellence. Based on *Framework for Teaching* (FFT) ratings, teachers' performance was grouped into five categories that yielded payouts of US$0, US$100, US$250, US$500, and US$750 for growth and for excellence.

5. These analyses are based on self-reported payout amounts, which we dichotomized: US$3,500 and above was considered a "high" payout; less was considered a "low" payout.

References

Barnett, J. H., & Ritter, G. W. (2008). When merit pay is worth pursuing. *Educational Leadership*, *66*(2). Retrieved from http://www.ascd.org/publications/educational-leadership/oct08/vol66/num02/When-Merit-Pay-Is-Worth-Pursuing.aspx

Bryan, J. F., & Locke, E. A. (1967). Goal-setting as a means of increasing motivation. *Journal of Applied Psychology*, *51*, 274–277.

Clotfelter, C., Glennie, E., Ladd, H., & Vigdor, J. (2008). Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina. *Journal of Public Economics*, *92*, 1352–1370.

Clotfelter, C. T., & Ladd, H. F. (1996). Recognizing and rewarding success in public schools. In H. F. Ladd (Ed.), *Holding schools accountable: Performance-based reform in education* (pp. 23–64). Washington, DC: Brookings Institution.

Danielson, C. (2007). *Enhancing professional practice: A framework for teaching* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Goldhaber, D. D., & Brewer, D. J. (1997). Why don't schools and teachers seem to matter? Assessing the impact of unobservables on educational productivity. *The Journal of Human Resources*, *32*, 505–523.

Gritz, M., & Theobald, N. D. (1996). The effects of school district spending priorities on length of stay in teaching. *The Journal of Human Resources*, *31*, 477–512.

Guarino, C. M., Santibañez, L., & Daley, G. A. (2006). Teacher recruitment and retention: A review of the recent empirical literature. *Review of Educational Research*, *76*, 173–208.

Hanushek, E. A., & Lindseth, A. A. (2009). *Schoolhouses, courthouses, and statehouses: Solving the funding-achievement puzzle in America's public schools*. Princeton, NJ: Princeton University Press.

Hatry, H. P., Greiner, J. M., & Ashford, B. G. (1994). *Issues and case studies in teacher incentive plans* (2nd ed.). Washington, DC: Urban Institute Press.

Heneman, H. G., III. (1998). Assessment of the motivational reactions of teachers to a school-based performance award program. *Journal of Personnel Evaluation in Education*, *12*, 43–59.

Kelley, C., Heneman, H. G., III, & Milanowski, A. (2002). Teacher motivation and school-based performance rewards. *Educational Administration Quarterly*, *38*, 372–401.

Kelley, C., Odden, A., Milanowski, A. T., & Heneman, H. G. (2000). *The motivational effects of school-based performance awards* (RB-29). Philadelphia, PA: Consortium for Policy Research in Education. Retrieved from http://www.cpre.org/sites/default/files/policybrief/870_rb29.pdf

Kelley, C. J., & Finnigan, K. (2003). The effects of organizational context on teacher expectancy. *Educational Administration Quarterly*, *39*, 603–634.

Ladd, H. (2011). Teachers' perceptions of their working conditions: How predictive of planned and actual teacher movement? *Educational Evaluation and Policy Analysis*, *33*, 235–261. doi:10.3102/0162373711398128

Lawler, E. E. (1990). *Strategic pay: Aligning organizational strategies and pay systems*. San Francisco, CA: Jossey Bass.

Lewis, J. L., & Springer, M. G. (2008, February). *Performance incentives in Texas: Why schools chose not to participate* (Working Paper No. 2008-18). Nashville, TN: National Center on Performance Incentives. Retrieved from http://www.performanceincentives.org/news/detail.aspx?linkid=234&moduleid=38

Locke, E. A., & Latham, G. P. (2006). New directions in goal setting theory. *Current Directions in Psychological Science*, *15*, 265–268.

MacAllum, K., Wells, J., & Ristow, L. (2011). *The Ohio Teacher Incentive Fund external evaluation: Final year 5 report*. Rockville, MD: Westat. Retrieved from http://www.westat.com/Westat/pdf/evaluation/Westat_OTIF_Year_5_Report_06-29-11_Final.pdf

MacLeod, M., Sielke, C., & Valentine, T. (2009, April). *Teachers' perceptions of positive and negative outcomes of differentiated pay*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.

Malen, B., Murphy, M. J., & Hart, A. W. (1987). Restructuring teacher compensation systems: An analysis of three incentive strategies. In K. Alexander & D. Monk (Eds.), *Attracting and compensating America's teachers* (pp. 91–141). Cambridge, MA: Ballinger Publishing Company.

Malen, B., Rice, J. K., Baumann, P., Beaner, T., Chen, E., Dougherty, A., . . . Robinson, C. (2009). *Building the plane in flight: A formative evaluation of Financial Incentive Rewards for Supervisors and Teachers (FIRST)*. Upper Marlboro, MD: Prince George's County Public School System.

Malen, B., Rice, J. K., Jackson, C., Hoyer, K., Hyde, L., Bivona, L., . . . Sutter, J. (2011). *Implementation, payouts, and perceived effects: A formative analysis of Financial Incentive Rewards for Supervisors and Teachers (FIRST)*. Upper Marlboro, MD: Prince George's County Public School System.

Marsh, J. (2012). The micropolitics of implementing a school-based bonus policy: The case of New York City's compensation committees. *Educational Evaluation and Policy Analysis*, *34*, 164–184.

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*. Santa Monica, CA: RAND Corporation.

Murnane, R. J., & Cohen, D. K. (1986). Merit pay and the evaluation problem: Why most merit pay plans fail and a few survive. *Harvard Education Review*, *36*, 1–17.

Murnane, R. J., & Olsen, R. J. (1990). The effects of salaries and opportunity costs on length of stay in teaching: Evidence from North Carolina. *The Journal of Human Resources*, *25*, 106–124.

Murphy, J. T. (1980). *Getting the facts: A fieldwork guide for evaluators and policy analysts*. Santa Monica, CA: Goodyear Publishing.

Odden, A., Kelley, C., Heneman, H., & Milanowski, A. (2001). *Enhancing teacher quality through knowledge and skills-based pay*. Philadelphia, PA: Consortium for Policy Research in Education. Retrieved from http://cpre.wceruw.org/publications/rb34.pdf

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). Thousand Oaks, CA: SAGE.

Rice, J. K. (2003). *Teacher quality: Understanding the effectiveness of teacher attributes*. Washington, DC: Economic Policy Institute.

Snyder, T. D., & Dillow, S. A. (2011). *Digest of Education Statistics 2010* (NCES 2011-015). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Springer, M., Podgursky, M., Lewis, J., Ehlert, M., Ghosh-Dastidar, B., Gronberg, T., . . . Taylor, L. (2008). *Texas Educator Excellence Grant (TEEG) Program: Year two evaluation*. Austin: Texas Education Agency.

U.S. Department of Education. (2012). *Teacher incentive fund: First implementation report, 2006 and 2007 grantees*. Washington, DC: Evaluation and Policy Development, Policy and Program Studies Service, Office of Planning, U.S. Department of Education.

Vroom, V. H. (1964). *Work and motivation*. New York, NY: Wiley.

Yuan, K., Le, V., McCaffrey, D. F., Marsh, J. A., Hamilton, L. S., Stecher, B. M., & Springer, M. G.

(2013). Incentive pay programs do not affect teacher motivation or reported practices: Results from three randomized studies. *Educational Evaluation and Policy Analysis*, *35*, 3–22.

Authors

JENNIFER KING RICE is an associate dean and professor in the College of Education at the University of Maryland, College Park. Her research draws on the discipline of economics to explore education policy questions concerning the efficiency, equity, and adequacy of U.S. public education. Her current work focuses on teachers as a critical resource in the education process.

BETTY MALEN is a professor in the College of Education at the University of Maryland, College Park. Her research brings the discipline of political science to the examination of the political determinants and the substantive effects of prominent education reforms and the impact of various policy shifts on the distribution of power within and across federal, state, and local units of the education system.

CARA JACKSON is a Harvard Strategic Data Project fellow and assistant director of accountability for the Urban Teacher Center. Her research focuses on teacher quality, educator evaluation, and uses of data to improve student learning.

KATHLEEN MULVANEY HOYER is a doctoral student in the Department of Teaching and Learning, Policy and Leadership at the University of Maryland, College Park and a research analyst at Activate Research. Her research interests include federal education policy and its interaction with state and district policies and practices. She is particularly interested in school finance, teacher policy, and the pursuit of educational equity.

Manuscript received October 25, 2012
First revision received May 23, 2013
Second revision received January 9, 2014
Accepted January 24, 2014