

How effective are financial incentives for teachers?

Linking teacher pay to student performance has become popular, but evidence on its effectiveness is mixed

Keywords: teacher incentives, student achievement, pay-for-performance, merit pay

ELEVATOR PITCH

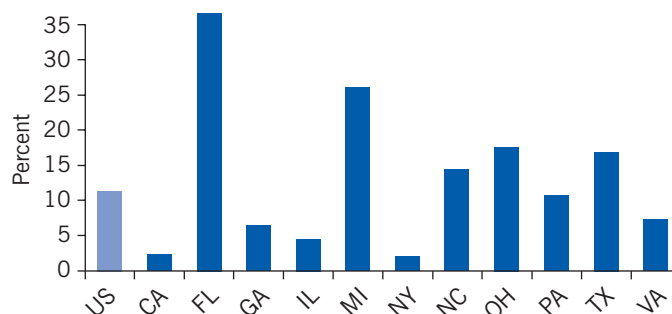
Concerns about poor student performance have led schools to diverge from traditional teacher compensation and base a portion of pay on student outcomes. In the US, the number of school districts adopting such performance-based financial incentives has increased by more than 40% since 2004. Evidence on individual incentives in developed countries is mixed, with some positive and some negligible impacts. There is less evidence for developing countries, but several studies indicate that incentives can be highly effective and far cheaper to implement. Innovative incentive mechanisms such as incentives based on relative student performance show promise.

KEY FINDINGS

Pros

- + Incentives can effectively improve student performance if they are designed well.
- + In developing countries, paying teachers for student performance has been shown to be highly effective at low cost.
- + Incentives based on the collective performance of small groups of teachers strike a balance between loss of effectiveness from free-riding teachers and gains in effectiveness from teachers cooperating with each other.
- + Innovative incentive mechanisms based on loss rather than gain or on relative student performance show promise for high effectiveness but are yet to be rigorously evaluated.

Across US states, the share of districts with performance incentives varies widely, 2011/2012 (largest states)



Note: New Jersey is missing due to insufficient data.

Source: Schools and Staffing Survey 2011–2012. Online at: http://nces.ed.gov/surveys/sass/tables/sass1112_2013311_d1s_005.asp

I Z A
World of Labor

Cons

- Overall, evidence on individual incentives in developed countries is mixed, with some positive and some negative impacts.
- In countries with high teacher salaries, incentives need to be large to elicit a response, which could make them too expensive for general use.
- Incentives based on the collective performance of large groups of teachers have been shown to have little impact on achievement and in some cases even generate negative impacts.
- There is no evidence that incentives tied to specific exams result in improvements in other measures of academic performance, suggesting a lack of general improvements in knowledge.

AUTHOR'S MAIN MESSAGE

Financial incentives for teachers can be effective if appropriately designed, but poorly designed incentives yield little benefit. Policymakers should avoid threshold-based incentives, such as meeting a target or doing better than other teachers, and instead favor systems based on incremental improvements in student performance. To avoid having teachers focus on any one specific measure at the expense of broad learning, incentives should be aligned with multiple outcomes that are both objective and subjective. If group incentives are used rather than individual incentives, the groups should be kept small: based on grade and subject, for example.

MOTIVATION

Traditionally, teachers in many parts of the world are compensated based on credentials (degrees and certifications) and experience. However, research has shown that the returns to experience are limited and that credentials have little impact on student performance. Nonetheless, teacher quality is very important. Because of this disconnect between teacher compensation and teacher performance, the idea of financial incentives for teachers (often called “performance,” “merit,” or “incentive” pay) aligned with measures of student performance has become increasingly popular.

According to the Schools and Staffing Survey of the US Department of Education, the share of US public and charter school districts with financial incentives for teaching excellence increased more than 40% from 2004 through 2012. There is also wide variation across states: some have no districts with incentive programs, and in others nearly half the districts offer incentives. Incentives have also been implemented in many other countries, including Denmark, India, Israel, Kenya, Hungary, and Norway.

In addition to financial incentives for direct improvements in student performance—the focus of this paper—other types of incentives include recruitment for hard-to-staff schools, incentives to acquire certain credentials, and incentives to recruit teachers in fields with shortages.

DISCUSSION OF PROS AND CONS

Key goals and variations in incentive pay

The motivating concept underlying teacher incentives is to pay teachers based on their productivity. The goal is to have two key impacts. First, to encourage teachers to exert more “effort,” broadly defined to include both quantity and quality. For example, to enhance quantity, teachers might spend more time on classroom instruction or after-school tutoring. They can enhance quality by adopting innovative teaching techniques, analyzing data to improve student performance, or experimenting with different teaching methods. Second, to recruit high-quality teachers. Some economists have theorized that incentives will attract people to the teaching profession who are better at improving student performance.

In practice, there are nuances to the implementation of incentive pay that can have large implications for effectiveness. Thus any policymaker must think carefully about the most appropriate types of incentives to provide.

Incentives based on individual or group performance

The first issue is whether to provide incentives based on the performance of teaching groups or individual teachers. Individual awards are provided to teachers based on how well they improve their own students’ performance. Group incentives provide rewards based on the average performance of a group of teachers. Most often, the group comprises all of the teachers in a school, school-grade, school-subject, or school-subject-grade. In some cases, teachers are grouped into smaller teams.

There are two key economic distinctions between these award types that drive their effectiveness. An important concern is that group incentives promote *free-riding* (a teacher free-rides if they reduce their effort toward achieving a common goal in response to an increase in the contribution of other group members): some teachers do not increase their effort as much as they would for individual incentives because they can take advantage of the improvements

in effort made by other group members. On the positive side, group incentives encourage cooperation among teachers, whereas individual performance promotes competition. Since teachers tend to benefit from the help of their colleagues, and a collegial school environment is better for productivity, there is worry that individual incentives could damage these relationships.

Metrics for measuring performance

The second key design characteristic is the metrics used to identify award winners. Typically, at least a portion of the incentives is aligned with a test score measure. Since basing incentives on unadjusted test scores tends to reward teachers for having high-ability students, rather than for improving student performance, districts have relied on test score growth to assess teacher performance. More complex statistical models called *teacher value-added* models are also widely used and purport to identify the direct contribution of teachers to student achievement growth. These models use statistical adjustments to student improvements in test scores to isolate the teacher's contribution to student achievement. Using test scores alone for teacher evaluation is problematic as it does not distinguish between the teacher's effectiveness and the existing ability of the students. Nonetheless, rewarding teachers based on student performance on a specific exam encourages teachers to target that exam alone, potentially with little impact on broader learning. Thus, in addition to test scores, it is common for districts to base pay on multiple outcomes, such as classroom observations and principal evaluations.

Structure of the incentive system

The final design characteristic is the structure of the incentive system. Incentives can be implemented through three methods: absolute targets, rank-order tournaments, and piece-rates. Absolute targets provide teachers with bonus pay if their students achieve certain outcomes regardless of how other teachers perform. For example, the Advanced Placement Incentive Program in Texas awarded teachers based on students passing Advanced Placement exams.

Rank-order tournaments award teachers for performing better than a certain percentage of other teachers on the metric. An example is Houston Independent School District's ASPIRE program, which pays teachers bonuses if they receive value-added scores above the 50th percentile; bonuses double for scores above the 75th percentile. Combined with absolute targets these constitute *threshold-based incentive systems*. In sum, threshold compensation implies paying teachers for reaching certain targets, such as doing measurably better than other teachers. An example would be providing additional compensation only to teachers who perform better than the average teacher.

Finally, *piece-rate compensation systems* pay teachers for each unit gain (incremental improvement) in student performance. For example, a piece-rate system might pay teachers \$100 multiplied by their value-added score. While economic theory suggests that piece-rate systems may be more effective and have fewer perverse incentives than threshold-based systems, districts tend to prefer the rank-order tournaments as they ensure budget security. Any relative system will place a cap on total payouts, while piece-rate systems, or systems with absolute targets, could generate far larger than expected payouts. An intriguing compromise between these two methods, called *pay for percentile*, has recently been proposed [1]. The idea is to pay teachers based on how students perform relative to a set of observably similar comparison students. While it has been demonstrated theoretically that such a system aligns incentives so that teachers provide optimal levels of effort, this has yet to be shown empirically.

Evidence on incentive pay

A fundamental difficulty with evaluating teacher incentives is that school systems that choose to have incentives may differ from those that do not in unobservable ways. For example, a typical concern is that districts may be more inclined to implement incentive pay programs if they are having problems recruiting effective teachers. Thus, any measured impacts of the plan would pick up the existing, low-performing state of the district. To get around this problem, much of the academic research on incentives has turned to using randomized controlled trials to evaluate teacher incentives. Randomization eliminates the unobserved differences between teachers affected by incentives and those not affected. However, randomized controlled trials are often limited in scope, and teachers may respond differently during these experiments than at other times, knowing that the experiments are temporary. Thus, in evaluating these programs one should also examine other evidence-based non-experimental studies that use methods that estimate causal impacts.

Figure 1 lists the studies that are considered here and provides some key takeaway messages, including *effect size* estimates (a common measure of performance in exams, measured in standard deviation units; typically a one standard deviation improvement is equivalent to improving by 25 to 30 percentile rankings) and whether they are statistically significant at the *95% confidence level* (i.e. there is less than 5% chance that the true impact is zero). Because most studies report multiple estimates—at different times relative to the start of the program and for both incentivized and non-incentivized exams—the table provides impact estimates for tests directly linked to the incentives using the average over all years of the study, if provided. Otherwise, the effect size estimate is for the last year of the study. Further, the estimate with the most extensive set of control variables is shown. In general, the studies tend to show positive results, though in many cases the estimates are close to zero and not statistically significant.

Evidence on US incentive pay programs

Most research on teacher incentive pay has been conducted in the US. In particular, there have been several evaluations of incentive pay schemes using various experimental designs. The incentive systems were implemented in many locations across the country, though all were in urban or suburban areas. Thus the programs tended to be in districts with large ethnic or racial minority and low-income populations. In all of the incentive systems studied and described here, except the Chicago Heights experiment, payments were based on teachers meeting thresholds rather than piece-rate incentives.

A widely publicized randomized controlled trial examined an incentive program in the city of Nashville, Tennessee, that provided teachers with large bonuses of up to \$15,000 for student improvements in mathematics performance [13]. This fixed-threshold system had relatively high thresholds: teachers needed to reach at least the 85th percentile of value-added scores to receive any award. The study found no statistically significant impact on mathematics scores from the awards. However, problems with some features of the incentive system might have reduced its effectiveness. First, the high thresholds may have discouraged many teachers from responding. Second, the focus on mathematics leaves open the question of impacts on other subjects. Third, the incentives were based entirely on test score performance. While this is an advantage in some sense as it allows the study to isolate this particularly focused incentive, it nonetheless limits what can be learned about incentives more generally.

The experiment in Chicago Heights, Illinois—a suburb of Chicago—found a similar lack of impact from individual incentives [5]. The results show no significant impact of both individual

Figure 1. Key research on teacher incentives

Study	Location (program name)	Incentive type	Maximum award	Mathematics effect size (s.d.)	Language/ reading effect size (s.d.)
Dee & Wyckoff (2013) [2]	Washington, DC, USA (IMPACT)	Individual awards based on observations, school involvement, and test score value-added	\$27,000 per year permanently	0.24*	
Fryer (2013) [3]; Goodman & Turner (2013) [4]	New York City, USA	Schools were permitted to decide on metric, but in practice almost all schools used schoolwide group awards based on test scores	\$3,000 onetime bonus that varied by position	-0.03*	-0.02
Fryer et al. (2012) [5] ^a	Chicago, Heights, IL, USA	Individual or small group award based on "pay for percentile" method Barlevy & Neal (2012) [1]	\$8,000 onetime bonus	0.09	—
Glazerman & Seifullah (2012) [6]	Chicago, IL, USA (Teacher Advancement Program)	Individual award based on observations, school involvement, and test score value-added	\$6,400 onetime bonus	-0.03	0.01
Glewwe et al. (2010) [7]	Busia and Teso, Kenya	Group awards at the school level based on test-score levels	Prizes worth up to \$51	0.077	
Imberman & Lovenheim (2015) [8] ^b	Houston, TX, USA	Group awards at the subject-grade level based on teacher value-added	\$7,700 onetime bonus	0.10*	0.03
Lavy (2002) [9] ^c	Israel	Group awards at the school level based on improvements in school-wide average scores	\$1,000 onetime bonus	0.011*	
Lavy (2009) [10]	Israel	Individual awards based on test scores	\$7,500 onetime bonus	d	
Muralidharan & Sundararaman (2011) [11]	Andhra Pradesh, India	Individual or group awards at the school level based on test score gains	\$11 per percentage point increase in average gain	0.28*	0.17*
Sojourner et al. (2014) [12]	Minnesota USA (Q-Comp)	Mix of individual and group awards based on multiple factors including observations and test scores	Varied by district; average maximum of \$2,200 onetime bonus	0.00	0.03*
Springer et al. (2012) [13]	Nashville, TN USA (POINT Experiment)	Individual awards based on teacher value-added	\$15,000 onetime bonus	0.05	—

Notes: If only one estimate is shown, it is an average over multiple measures. a. Study that had not completed peer review at the time of writing, and thus results are not final. b. Study that does not identify the impact directly, but estimates impacts indirectly through the effect of changing the teacher's impact on likelihood of award receipt. c. Estimates are student-weighted average of religious- and secular-school effects. d. Does not provide information to calculate effect size but finds positive impacts for both subjects. *Statistically significant result. s.d. = standard deviation. For full source details, go to <http://wol.iza.org/articles/how-effective-are-financial-incentives-for-teachers>

and small-team group incentives based on giving teachers up to \$8,000 for student test performance. Despite these non-significant results, two unique aspects of this study are worth mentioning. First, this is the only study that uses the pay for percentile incentive system. Second, despite finding no impact for incentives that provided teachers with monetary gains, the study also tested the impacts of monetary losses. There is substantial economic evidence that people care more about losing money than about gaining money, even if the amounts are the same—a concept known as *loss aversion*. Prior to the start of the school year, all teachers received a bonus and signed a contract requiring them to pay it back at the end of the school year if their students did not perform sufficiently well on an end-of-year exam. This simple change in the structure of the incentive program generated very large positive impacts on test scores. Nonetheless, while intriguing, such a method may be hard to implement in practice.

Another randomized experiment in individual incentives was Chicago's implementation of the Teacher Advancement Program, which paid teachers up to \$6,400 based on a mix of a student performance improvement measure (value-added scores), class observations, and teachers' involvement in the school [6]. Unlike the previous two experiments, entire schools were randomized into earlier or later adoption of the program. This method better reflects how adoption would occur in practice, as typically whole schools or districts would adopt an incentive system en masse. Even so, as with the other studies, this one finds no significant impacts of incentives after one year.

Two other US studies do find positive impacts of incentives on student performance. One study evaluates a unique characteristic of the IMPACT incentive program implemented in Washington, DC [2]. This program provided teachers with an opportunity to earn a one-time bonus plus a permanent salary increase of up to \$27,000 a year, which makes it considerably more expensive than typical programs that provide temporary bonuses but also provide an especially large incentive. As in the Chicago Teacher Advancement Program, teachers received the incentive award based on a metric that incorporated the teacher's value-added scores, classroom observations, and the teacher's involvement in the school. While the program was not implemented experimentally, the study took advantage of an aspect of the system's design that provides a natural experiment: To qualify for the permanent salary increases, teachers had to be rated "highly effective" two years in a row. This means that teachers who were just barely rated highly effective in the first year had a much stronger incentive to perform in the second year than those who just barely missed, even though the two groups are otherwise virtually identical. The study compared these two groups of teachers and found a significant positive impact on teacher performance.

A second study looks at Minnesota's Q-Comp program [12]. Unlike the other programs discussed so far, Q-Comp gave school districts substantial flexibility in designing the incentives, including selecting what metrics to use and whether to base awards on group or individual performance. Thus, this study does not differentiate between incentive types. Nonetheless, by comparing districts based on their timing of adoption and whether they adopt at all, the study finds small positive effects for reading but no statistically significant impacts for mathematics.

The research described above indicates that the impacts for individual incentive awards in the US are mixed and that, at best, awards need to be very large to be effective. Nonetheless, it remains possible that group-based awards can capitalize on encouraging cooperation among teachers despite the potential for free-riding. Once again, though, the evidence is mixed. The best evidence on the impacts of group awards in the US comes from two studies that looked at schools in New York City that were randomly assigned to an incentive program [3], [4].

While the program was designed to give schools flexibility in defining how incentives would be distributed (though they had to be awarded based on test scores), in practice nearly all schools adopted incentives of about \$3,000 per teacher based on average school-wide achievement. The studies find no significant positive impact on mathematics or reading scores and a small but significant negative impact in middle schools. However, one of the studies points out that smaller schools had better responses to incentives, suggesting that free-riding, which would be a bigger problem in large schools, plays an important role in group incentives.

Finally, a study that looks at Houston's ASPIRE program tested for this free-riding issue more directly [8]. The study focuses on an incentive system for high school teachers that provides awards at the subject-grade level (for example, science teachers in grade 9). Thus, group sizes differ considerably. The awards are based entirely on test score value-added and were as large as \$7,700 per teacher. The study finds substantial evidence of free-riding in large groups of teachers, indicating that the most effective group incentives are those that put teachers in teams of five or fewer. Though the study does not estimate direct impacts of the incentives, impacts are ascertained indirectly from the free-riding estimates and show large positive effects of incentives on targeted exams. However, there was no impact on exams in the same subjects that were not linked to the incentives. By highlighting the important problem that teachers may direct their efforts narrowly to the exams rather than to overall learning, this finding provides support for adopting a range of metrics for awards.

International evidence on teacher financial incentives

Outside the US, there has been much less research on teacher incentives. Even so, there are some key studies that show that the international evidence on incentives is much more positive than the US evidence. Two studies in Israel find substantial positive impacts of teacher incentives on student performance [9], [10]. Since Israel is a developed country with an education system similar to those in the US and European countries, this additional evidence can be combined with the US evidence about how incentive pay works in developed countries. The studies focus on two incentive systems that reward teachers for how well their students perform on the Bagrut, a combined high-school exit and college entrance exam. Though the studies are not randomized experiments, they nonetheless were conducted in ways that permit estimating causal effects. The first study estimates the impact of a relatively low-stakes school-wide (group) incentive program that gave teachers up to US \$1,000 each. The second looks at an individual teacher incentive program that had much higher payouts—as much as US \$7,500. In both cases, the incentives had positive and statistically significant impacts on student performance.

Knowledge of impacts in developing countries is far more limited. Nonetheless, two important experiments offer insights that suggest that incentives can be highly effective and far cheaper to implement in developing countries. A randomized controlled trial conducted in the state of Andhra Pradesh in India assigned teachers to one of three groups: no incentive; a school-wide incentive; or an individual incentive [11]. As a percentage of teacher salaries, the incentives were substantial, but in absolute terms they were inexpensive—typically less than US \$100 per teacher. The incentives were based on a piece-rate system rather than a threshold system. The study finds large and statistically significant impacts on mathematics and language achievement from individual incentives and a still sizable and significant, but 50% smaller, impact for group incentives. It should be noted that, unlike schools in the US, these schools are small, averaging only three teachers per school. The New York City incentive schools, for example, averaged 16

incentivized teachers. Thus, the finding of group incentive effects here is consistent with the Houston study that shows small groups respond more. The India study attributes much of this impact to the fact that incentivized teachers put more effort into student preparation for the exams, including through additional test preparation, more homework and class work, extra tutoring outside of school hours, and more attention to weaker students.

In the second developing country experiment, conducted in Kenya, schools were randomly assigned to an incentive program that offered prizes to teachers worth up to US \$51 based on average test score performance in the school [7]. The incentives were successful at improving student performance on the incentivized exams. However, as in the Houston study, there was little evidence of impacts on non-incentivized exams in the same subjects.

Figure 2 provides a broad overview of the research discussed in this paper.

Figure 2. Summary of impacts of teacher financial incentives

Incentive type	US and Israel		India and Kenya	
	Results	Locations	Results	Locations
Individual incentives	Mixed statistically insignificant and positive	6	Positive	1
Group incentives	Mixed negative and positive	5	Mixed statistically insignificant and positive	2
Threshold incentives	Mixed negative and positive	7	Positive	1
Piece-rate incentives	Mixed statistically insignificant and positive	2	Positive	1
Directly incentivized exams	Mixed negative and positive	9	Mixed statistically insignificant and positive	2
Non-incentivized exams	Statistically insignificant	1	Statistically insignificant	1
Multiple outcomes	Mixed statistically insignificant and positive	3		—
Test scores only	Mixed negative and positive	6	Positive	2
Total number of locations		9		2

Source: Author's own.

I Z A
World of Labor

LIMITATIONS AND GAPS

While there are a number of excellent studies evaluating teacher incentive pay, much remains unknown. First, the research is heavily concentrated in the US. Only a handful of rigorous

studies have been conducted in other countries, and none of them are in Europe, East Asia, or South and Central America. Second, most of the incentive schemes studied are based on thresholds requiring teachers to reach certain achievement levels to receive awards. This is due mainly to the attractive feature that such methods (especially those based on rankings of teachers) provide: certainty in budgeting. Theoretically, however, threshold incentives would be expected to be less effective than piece-rate methods, which pay for each unit of additional performance. Third, direct comparisons of group and individual incentives are rare. While both have been studied in different contexts, comparisons in the same location are limited. Fourth, there is a lack of empirical evidence on how basing incentives on multiple outputs compares to basing incentives on single outputs, like teacher value-added.

SUMMARY AND POLICY ADVICE

In general, the evidence on the impacts of financial incentives for teachers is mixed. While financial incentives appear to be quite successful in developing countries, the results are unclear in developed countries like the US and Israel, though those too tend to weigh more toward positive effects than negative. Even so, in cases where there are positive impacts, the effects appear to be concentrated on the directly incentivized exams, which indicates that financial incentives may not improve general learning if they are narrowly targeted. Studies that look at incentives based on multiple outcomes tend to show more positive effects.

Several recommendations can be derived from the studies reviewed here to guide policymakers considering implementing teacher incentive pay to improve student performance. First, the choice of metrics and the incentive structure of the system matters for its effectiveness, and poorly designed systems can even make outcomes worse. Second, incentives should be based on multiple outcomes, of which student performance improvement (teacher value-added) is just one of several metrics, at least one of which should be subjective (principal evaluation or classroom observations). Third, when possible, thresholds and rank-order tournaments should be avoided in favor of piece-rate systems. Pay for percentile is a promising method but remains to be empirically tested.

For group-based incentive systems, the size of groups should be kept low. Incentives at the school level are typically ineffective, but the evidence suggests that small groups can generate improvements in student performance.

Acknowledgments

The author thanks an anonymous referee and the IZA World of Labor editors for many helpful suggestions on earlier drafts.

Competing interests

The IZA World of Labor project is committed to the *IZA Guiding Principles of Research Integrity*. The author declares to have observed these principles.

© Scott A. Imberman

REFERENCES

Further reading

Neal, D. “The design of performance pay in education.” In: Hanushek, E. A., S. J. Machin, and L. Woessmann (eds). *Handbook of the Economics of Education: Vol. 4*. Amsterdam: Elsevier, 2011; pp. 495–550.

Podgursky, M. J., and M. G. Springer. “Teacher performance pay: A review.” *Journal of Policy Analysis and Management* 26:4 (2007): 909–950.

Key references

- [1] Barlevy, G., and D. Neal. “Pay for percentile.” *American Economic Review* 102:5 (2012): 1805–1831.
- [2] Dee, T., and J. Wyckoff. *Incentives, Selection, and Teacher Performance: Evidence from IMPACT*. NBER Working Paper No. 19529, 2013.
- [3] Fryer, R. G. “Teacher incentives and student achievement: Evidence from New York City public schools.” *Journal of Labor Economics* 31:2 (2013): 373–407.
- [4] Goodman, S. F., and L. J. Turner. “The design of teacher incentive pay and educational outcomes: Evidence from the New York City bonus program.” *Journal of Labor Economics* 31:2 (2013): 409–420.
- [5] Fryer, R. G., S. D. Levitt, J. A. List, and S. Sadoff. *Enhancing the Efficacy of Teacher Incentives through Loss Aversion: A Field Experiment*. NBER Working Paper No. 18237, 2012.
- [6] Glazerman, S., and A. Seifullah. *An Evaluation of the Chicago Teacher Advancement Program (Chicago TAP) after Four Years (Final Report)*. Washington, DC: Mathematica Policy Research, Inc., 2012.
- [7] Glewwe, P., N. Ilias, and M. Kremer. “Teacher incentives.” *American Economic Journal: Applied Economics* (2010): 205–227.
- [8] Imberman, S. A., and M. F. Lovenheim. “Incentive strength and teacher productivity: Evidence from a group-based teacher incentive pay system.” *Review of Economics and Statistics* 97:2 (2015): 364–386.
- [9] Lavy, V. “Evaluating the effect of teachers’ group performance incentives on pupil achievement.” *Journal of Political Economy* 110:6 (2002): 1286–1317.
- [10] Lavy, V. “Performance pay and teachers’ effort, productivity, and grading ethics.” *American Economic Review* (2009): 1979–2011.
- [11] Muralidharan, K., and V. Sundararaman. “Teacher performance pay: Experimental evidence from India.” *Journal of Political Economy* 119:1 (2011): 39–77.
- [12] Sojourner, A. J., E. Mykerezi, and K. L. West. “Teacher pay reform and productivity panel data evidence from adoptions of Q-Comp in Minnesota.” *Journal of Human Resources* 49:4 (2014): 945–981.
- [13] Springer, M. G., D. Ballou, L. S. Hamilton, V.-N. Le, J. R. Lockwood, D. F. McCaffrey, M. Pepper, and B. M. Stecher. *Final Report: Experimental Evidence from the Project on Incentives in Teaching (POINT)*. Nashville, TN: National Center on Performance Incentives, 2012.

The full reference list for this article is available from the IZA World of Labor website (<http://wol.iza.org/articles/how-effective-are-financial-incentives-for-teachers>).