Educational Evaluation and Policy Analysis
June 2016, Vol. 38, No. 2, pp. 199–221
DOI: 10.3102/0162373715609687

Effective Teacher Retention Bonuses: Evidence From Tennessee

Matthew G. Springer
Walker A. Swain
Luis A. Rodriguez
Vanderbilt University's Peabody College
Tennessee Consortium on Research, Evaluation, and Development

We report findings from a quasi-experimental evaluation of the recently implemented US$5,000 retention bonus program for effective teachers in Tennessee's Priority Schools. We estimate the impact of the program on teacher retention using a fuzzy regression discontinuity design by exploiting a discontinuity in the probability of treatment conditional on the composite teacher effectiveness rating that assigns bonus eligibility. Point estimates for the main effect of the bonus are not different from zero. However, for teachers of tested subjects and grades, the program has a consistently positive effect that is both statistically and substantively significant. We hypothesize that the null finding for the main effect is driven by teachers of untested subjects and grades given the amount of weight Tennessee's teacher evaluation system attributes to school-level performance. This creates a strong incentive to exit the Priority Schools that are by definition low performing. Implementation concerns, including the timing of application process and observed noncompliance in bonus distribution, present obstacles for both the program's effectiveness and its evaluation.

Keywords: *financial incentives, teacher retention, teacher mobility, educational equity*

Introduction

The inequitable distribution of highly effective teachers across schools is a major concern of policy leaders and practitioners interested in the condition of American public schooling. One of the most salient findings in recent education research is that differences in teacher quality result in substantially different outcomes for students in school and beyond (e.g., Aaronson, Barrow, & Sander, 2007; Chetty, Friedman, & Rockoff, 2014; Hanushek, Kain, O'Brien, & Rivkin, 2005; Kane, Rockoff, & Staiger, 2008; Rockoff, 2004; Sanders & Rivers, 1996). In economic terms, it has been estimated that the net present value of future earnings for a student having access to a teacher one standard deviation above average effectiveness approaches half a million dollars (Hanushek, 2011). However, it is equally well documented that minority and low-income students, and particularly those in schools with high concentrations of students in poverty or racial minorities, are considerably less likely to be taught by teachers with strong credentials or high estimated effectiveness (e.g., Clotfelter, Ladd, & Vigdor, 2011; Goldhaber, Lavery, & Theobald, 2015; Feng & Sass, 2012).

In addition to their initial maldistribution, highly qualified teachers leave disadvantaged schools at higher rates than their less-qualified counterparts. It is, therefore, not surprising that policymakers, researchers, and practitioners interested in promoting learning, economic growth, and the closing of achievement and opportunity gaps would explore policies designed to promote the retention of effective teachers in historically disadvantaged schools. Indeed, since 2010, the number of states adopting retention incentive policies has more than doubled, with the largest growth among states participating in the federally funded Race to the Top (RTTT) initiative.

The inequitable distribution of highly effective teachers across schools within districts is arguably a consequence of uniform teacher salary schedules in conjunction with differences in nonpecuniary characteristics of schools (e.g., condition of school building, principal leadership, safety, distance from home, and perhaps most importantly the makeup of the student body).[1] To equalize teacher quality, federal and state policymakers have demonstrated interest in designing practical incentives to help offset differences in nonwage job characteristics (Kirshstein, Berger, Benatar, & Rhodes, 2004; Prince, 2002, 2003), including signing bonuses, certification stipends, tuition reimbursement, loan forgiveness, tax credits, and housing subsidies.[2]

The theory behind retention incentives assumes that offering financial incentives will help retain more teachers (ideally in the upper tail of the ability distribution) in hard-to-staff or chronically low-performing schools. Proponents of financial incentives commonly argue that schools need to be able to respond to labor market conditions by offering additional incentives when trying to retain teachers at less competitive campuses, which are typically located in rural or densely urban areas. Unfortunately, there is relatively little research on the amount of additional compensation needed to offset differences in nonpecuniary workplace characteristics. Retention bonuses in practice have ranged from US$250 to more than US$20,000, whereas a simulation study by Feng (2009) estimated that an additional US$10,000 a year in pay for teachers in hard-to-staff schools would improve retention rates to that of teachers in average schools.[3]

This study adds to the current knowledge base in a number of important ways. First, this is one of the first studies to evaluate a retention bonus program that targets highly effective teachers using a rigorous causal research design. Second, prior studies have often examined bonuses introduced in the context of a broad set of reforms and cannot isolate the impact of the retention bonus component. We assess the impact of a large pilot bonus program, which the state implemented apart from any other major compensation or school policy changes. Third, the current study advances our understanding of teacher retention bonuses by implementing a cost-effectiveness analysis.

We estimate the impact of the program on teacher retention using a fuzzy regression discontinuity (RD) design. We exploit a discontinuity in the probability of treatment conditional on the continuous composite teacher effectiveness rating that assigns bonus eligibility. Our primary research question is as follows:

> **Research Question 1:** To what extent does a US$5,000 retention bonus affect retention of high-performing teachers in Priority Schools that elected to participate in the program?

We also investigate if the retention bonus program produces long-run net benefits for students at participating schools, as well as potential benefits for state budgets due to improved earnings potential of students and avoided costs of teacher turnover.

Point estimates for the main effect of the bonuses are not different from zero. However, for teachers of tested subjects and grades, the program has a consistently positive effect that is both statistically and substantively significant. We argue that the null finding for the main effect is driven by teachers of untested subjects and grades given the amount of weight Tennessee's teacher evaluation system attributes to school-level performance for untested-subject teachers. We also note that the stark contrasts in estimated effectiveness between bonus recipients and their likely replacements make student benefits likely across a range of plausible program impacts. Implementation concerns, including the timing of application process and observed noncompliance in bonus distribution, present obstacles for both the program's effectiveness and its evaluation.

Tennessee's Retention Bonus Program

The distribution of highly effective teachers in the Tennessee public school system, as defined by a value-added measure of teacher effectiveness, is working to the detriment of students in schools with large concentrations of economically disadvantaged and non-White students.

During the 2011–2012 school year, approximately 17% of teachers left their school and the attrition rate for the state's most-effective teachers is around 7%. The attrition rate for highly effective educators increases to 10% when focused on urban districts in the state and 23% when focused on the bottom 5% of schools in the state (more than 3 times greater than the statewide attrition rate of highly effective teachers).

In spring 2013, in an effort to combat these high rates of teacher turnover among highly effective teachers in chronically low-performing schools, the Tennessee Department of Education (TDOE) and the Tennessee Governor's Office announced a 1-year teacher retention bonus program for Priority Schools. Under the program, all Priority Schools were eligible to participate by applying to offer US$5,000 retention bonuses to any Level 5 teacher who was teaching in a Priority School.[4] For many of the teachers in Tennessee Priority Schools, a US$5,000 bonus constitutes approximately a 10% salary increase or the equivalent of a teacher with a master's degree moving from 10 to 15 years of experience on a district salary schedule.

Level 5 teachers at Priority Schools who accepted retention bonuses were required to complete the 2013–2014 school year at a Priority School to keep the bonus. For the purposes of this program, a teacher is defined as a classroom teacher with assigned students and associated evaluation scores. It excludes principals, school counselors, and school services personnel. Itinerant teachers can receive a prorated amount of the retention bonus based on the number of days per week that he or she is actually working in a Priority School.

Priority School Status

In 2012, the TDOE secured waivers from certain portions of the federal No Child Left Behind (NCLB) law. The waiver allowed Tennessee to replace NCLB's Adequate Yearly Progress proficiency targets with a system that focuses on "ensuring growth for all students every year and closing achievement gaps by ensuring faster growth for those students who are furthest behind" (TDOE, 2012, p. 52). In addition, the state identifies individual schools based on these relative performance measures, ranging from high-performing "reward" schools to low-performing "priority" schools.

Tennessee identified 83 Priority Schools based on a composite proficiency rate (success rate) for all students in a school. The bottom 5% of schools in the state receive priority status. The composite proficiency rate that determines a school's eligibility for priority status is based on the following formula, where math, reading/language arts, and science are for Grades 3 through 8 and Algebra I, English I and II, and Biology I, and graduation rate are for high schools:

$$\frac{\begin{array}{c}\#\text{proficient or advanced students in math}+\\ \text{reading language arts}+\text{science}+\text{algebra I}+\\ \text{English I}+\text{English II}+\text{Biology}+\#\text{ HS Graduates}\end{array}}{\begin{array}{c}\#\text{ tested students in math}+\text{reading language arts}+\text{science}+\\ \text{algebra I}+\text{English I}+\text{English II}+\text{Biology}+\\ \#\text{ students in HS graduation cohort}\end{array}}$$

The success rates used for determining Priority Schools include up to 3 years of data. Success rates are calculated for schools with at least 2011–2012 school year data. Only schools that are active in the 2011–2012 school year, with at least 2010–2012 and 2011–2012 data, receive Priority School status.

Teacher Evaluations

In January 2010, the Tennessee General Assembly passed Senate Bill 5, also known as the First to the Top Act, thereby reforming dozens of areas of state education policy. The ambitious reforms helped Tennessee win a US$501 million award granted under the RTTT competition to implement and institutionalize innovative policy changes statewide. One of the most contentious provisions of the new law required that all school personnel be evaluated annually and personnel decisions be based, in part, on those evaluations.[5]

As of July 2011, the Tennessee State Board of Education approved four teacher evaluation models—the Tennessee Educator Acceleration Model (TEAM), Project COACH, Teacher Effectiveness Measure (TEM), and Teacher Instructional Growth for Effectiveness and Results (TIGER). The evaluation models all follow the requirements set forth by Tennessee's Teacher Effectiveness Advisory Committee and adopted by the State Board of Education and have the same goals—to monitor teacher performance and encourage teacher development, though implementation from one model to the next is quite different. As of the 2012–2013 school year, more than 80% of teachers across Tennessee use TEAM as their evaluation model, whereas TEM is the second most frequently used (11%), followed by Project COACH (5%) and TIGER (2%; Ehlert, Pepper, Parsons, Burns, & Springer, 2013). In the current study, 80% of schools use the TEAM model, 11% use TEM, 6% use Project COACH, and 3% use TIGER.

Annual evaluations differentiate teacher performance based on an overall level of effectiveness score (what we refer to as overall teacher evaluation rating or teacher rating for short). The overall teacher evaluation rating is calculated using individual and school-level student growth scores and achievement data as well as teacher observations for teachers in tested and untested subjects and grades. For teachers of tested subjects and grades, state law specifies that 50% of the evaluation score be based on student achievement data. Of this 50%, 35% is comprised of value-added student achievement data as calculated using the Tennessee Value-Added Assessment System (TVAAS), and 15% is based on alternative measures of student achievement approved by the State Board of Education and selected through joint agreement by the educator and evaluator. The remaining 50% of an evaluation must be based on qualitative measures, including teacher observations, student perception surveys, personal conferences, and review of prior evaluations and work. For teachers of untested subjects and grades, 40% of the evaluation is comprised of student achievement data—25% based on school- or district-wide student growth as measured by TVAAS and 15% based on additional approved measures of student achievement. The remaining 60% of the overall evaluation scores are determined through qualitative measures similar to those used for teachers of tested subjects and grades.

The classroom observation process noticeably differs across the TEAM and the three approved alternative evaluation models. Both the TEAM and TIGER models use announced and unannounced classroom observations whereas the COACH model only relies on unannounced observations. The number and frequency of observations also varies across models—TEAM requires that apprentice teachers be observed for a minimum of 90 minutes across six observations whereas tenure teachers must be observed for a minimum of 60 minutes across four observations; TIGER requires teachers to be observed for a minimum of 60 minutes across six observations; and COACH requires that all teachers receive 10 observations for a minimum of 5 minutes per observation.[6] Finally, all four models have distinctive processes of providing feedback, though they all include written feedback provided by the observer and a post-observation meeting.

The effectiveness scores range from 0 to 500 in all four evaluation models and ultimately define the discrete performance categories to which teachers are assigned. Denoting X as the teacher score, for all models, teachers with $X < 200$ are categorized as "Significantly Below Expectation" (Level 1), teachers with $200 \leq X < 275$ as "Below Expectation" (Level 2), teachers with $275 \leq X < 350$ as "At Expectation" (Level 3), teachers with $350 \leq X < 425$ as "Above Expectation" (Level 4), and teachers with $X \geq 425$ as "Significantly Above Expectation" (Level 5). Ratings reports provided to teachers include the discrete rating but not the underlying score on the 0 to 500 scale. This is useful for interpreting our findings because it means that teachers with very similar underlying scores but different discrete ratings did not receive the necessary information to determine their closeness to the threshold.

Brief Review of Teacher Retention Bonus Literature

A number of studies have directly examined the influence of cash bonuses on retention and attrition rates of teachers at high-need schools with mixed results.[7] The nature, size, and context of the evaluated bonuses vary considerably, as do the methods used to assess their impacts. One of the difficulties retention bonus studies have faced is the fact that policymakers often introduce retention bonuses in the context of a broader set of reforms (i.e., Balch & Springer, 2014; Dee & Wyckoff, 2013; Hough, 2012). For example, in their evaluation of a pilot supplemental funding program to a group of educationally disadvantaged schools in North Carolina

(NC), Henry, Fortner, and Thompson (2010) found that approximately half of the money went toward salary bonuses that gave the schools a comparative advantage in hiring and retaining teachers. The authors note that in years of the pilot funding, teacher turnover decreased significantly at the schools with the supplement, in spite of having the most disadvantaged students in the state. Meanwhile, turnover rates increased at nonsupplement schools. However, although the RD design of the study allowed the authors to attribute the increased retention to the supplemental funding, they were unable to distinguish the effects of salary bonuses from other expenditures that might have made teachers more likely to stay.

Conversely, Clotfelter, Glennie, Ladd, and Vigdor (2008) were able to directly examine an US$1,800 annual teacher retention bonus offered in NC between 2001 and 2004 to certified math, science, and special education teachers in a set of low-performing and/or high-poverty secondary schools. The authors found modest, but significant effects on teacher turnover. The difference-in-difference-in-difference analytical strategy indicated that the bonuses reduced turnover rates of eligible teachers in eligible schools by 17%, or 5 percentage points. Survey results also indicated widespread misunderstandings about the nature of the retention incentive offered and skepticism among teachers and administrators that the size of the bonus would be sufficient. The NC bonus program differed from the Tennessee retention bonuses both in its smaller magnitude (US$1,800 vs. US$5,000) and the fact that it was not tied to any measure of teacher quality, but rather specified credentials (math, science, and special education teachers).

Relatedly, several recent studies have applied rigorous methods to estimate effects of recruitment incentives, in the form of large conditional scholarships and cash bonuses that also required teachers to remain in a high-need school for a specified time. Glazerman, Protik, Teh, Bruch, and Max (2013) experimentally evaluated a Talent Transfer Initiative (TTI) that offered high value-added teachers a US$20,000 bonus—paid in installments over a 2-year period—if they transferred into and remained in schools that had low-average test scores. The transfer incentive increased both transfer and retention of targeted teachers during the payout period across 10 school districts in seven states. However, not surprisingly, the difference was no longer statistically significant after payments stopped. Similarly, Steele, Murnane, and Willett's (2009) evaluation of California's Governor's Teaching Fellowship (GTF) program, which offered US$20,000 conditional scholarships (US$5,000 per year over 4 years) to attract and retain academically talented, newly licensed teachers to low-performing schools, found the program had significant effect on teacher recruitment but did not differentially affect retention among GTF recipients.[8]

Finally, Dee and Wyckoff's (2013) analysis of salary bonuses in IMPACT, a high-stakes teacher evaluation system implemented in District of Columbia (DC) that was designed to improve teacher quality and student achievement, faces similar challenges to other studies of bonuses administered as part of broader reforms. Their study implements a rigorous set of analyses similar to those described in this article to offer important evidence with respect to the impact of bonuses. Using an RD design, the authors compared teachers near the IMPACT score threshold that separated "Effective" from "Highly Effective" teachers. Similar to the evaluation program in Tennessee, the DC system utilized a mix of observation and value-added metrics to generate a continuous composite score with sharp cut points to group teachers into consequential categories of effectiveness. Teachers were qualified for a large one-time bonus (up to US$25,000) after being rated "Highly Effective" for 1 year and a sizable and permanent base salary increase (as large as US$27,000 per year) on achieving "Highly Effective" status in a second consecutive year.

Although the IMPACT incentive had positive effects on teacher performance, impacts on retention of effective teachers were not statistically significant. At first glance, this null finding seems contradictory to the general trend toward larger effects for larger incentives, but there are important contextual issues that likely contributed to the inability to detect an effect. First, the comparison group (teachers who had barely missed the cutoff for "Highly Effective" designation in the prior year) was also subject to a substantial financial incentive, as they were eligible for the one-time bonus in the study year and the prospect of base pay increases in the coming years. Second, perhaps as a consequence of the new IMPACT policies, the retention rate for effective teachers above and around the cutoff was particularly high in the year studied (roughly 90%).

Data, Sample, and Analytic Strategy

The primary analytic strategy we use is an RD design exploiting the sharp cutoff in a teacher's overall evaluation rating that determines eligibility for the retention bonus in participating schools. Because administrators chose both whether to participate in the program and whether to offer bonuses to teachers, our primary findings can be understood as a treatment-on-the-treated (TOT) analysis in a fuzzy RD framework. We take advantage of rich administrative data supplemented by county-level economic data to add precision and efficiency to our estimates, and to examine potential differences between the participating sample and broader school population. We describe our data sources, analytic sample, and methodology in detail below.

Data Sources

This study utilizes administrative data obtained from the TDOE and maintained by the Tennessee Consortium on Research, Evaluation, and Development (the Consortium) at Peabody College's Vanderbilt University. We cleaned and merged relevant teacher and school information from multiple data sources to create a single data file for the 2011–2012 through 2013–2014 school years.

Our first data file captures demographic, job assignment, and salary information on all certified educators in Tennessee. This file is the combination of information from two different data systems containing certified staff information that serve different reporting and compliance functions in the state. We spent significant time capturing the most accurate information from both of these data sets while reconciling disagreements and longitudinal inconsistencies. These staffing files contain common demographic variables, such as years of teaching experience, highest educational level, race/ethnicity, and salary and job assignment information.

Our second data source is from the TVAAS and Tennessee's online teacher evaluation platform, CODE. The TVAAS data file, created by SAS Institute in Cary, NC, contains value-added estimates for teachers in Grades 4 through 8 in math, reading/language arts, science, and social studies and end of course reporting for high school educators in English I, II, III; Algebra I and II; Biology I; and U.S. history. Teacher-effect estimates are calculated for specific subject, grade, year pairings as well as for composites across subject, grades, and years. All scores are expressed in state normal curve equivalents, using the 2008–2009 school year as the reference year. Our analyses use data on teacher composite scores, which statistically combines subject, grade, and year TVAAS estimates.[9] Tennessee's online teacher evaluation data platform, CODE, houses teacher observation data from the TEAM rubric and other state-approved observation systems. The CODE platform also contains school growth ratings from TVAAS that serve as the third and final component of teachers' final evaluation rating.[10]

Our school-level information comes from multiple sources, including state school accountability reports, National Center for Education Statistics' Common Core of Data, and aggregated individual student- and teacher-level information at the school level. These school files typically used variables such as school levels, student enrollment, and proficiency rates as well as select student and teacher demographic information.

TDOE also provided our research team at the Tennessee Consortium with details on the design and implementation of the teacher retention bonus program. The teacher retention bonus program file contains teacher name, school name, and local education agency for all teachers that received a retention bonus. The file also contains a list of all Priority Schools with an indicator for whether they opted to participate in the program.

Sample

Our sample includes all teachers working in Priority Schools in Tennessee during the 2012–2013 school year. We are most interested in the schools that elected to participate in the retention bonus program and the teachers that worked in those schools. Bonus program participation

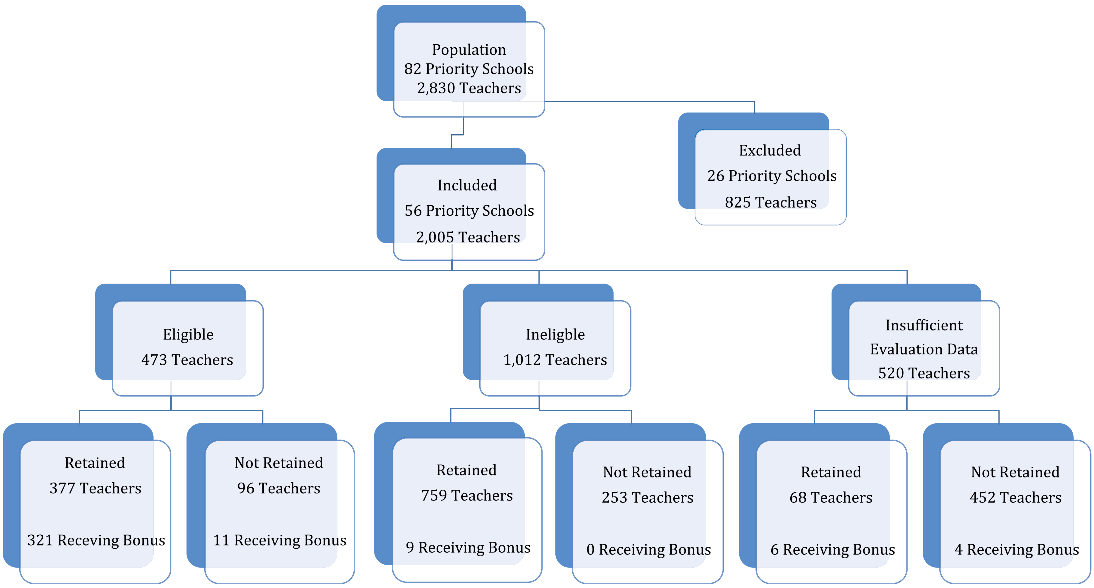


FIGURE 1. *Consort diagram.*

required that the school principal and district superintendent sign and submit a letter of commitment to the state that affirmed their agreement to all of the terms and conditions of the retention bonus program.[11] Similarly, the program was structured so that the burden of proof for bonus eligibility resided at the district level. That is, if the district determined a teacher was/was not eligible for a bonus, the person received/did not receive a bonus. The state did not conduct an independent audit of how funds were distributed. However, the state did play a significant role recruiting schools to participate in the program.

As displayed in Figure 1, there were 82 Priority Schools during the 2012–2013 school year that qualified to participate in the program. Of those 82 schools, 56 of them, employing 2,005 teachers, elected to participate. Although take-up may appear low given that participating schools had nothing to lose, the education sector has been notoriously skeptical of incentive pay programs. Low take-up rates have been documented in several Texas incentive pay initiatives (Springer et al., 2010; Springer et al., 2008; Springer et al., 2009) as well as New York City's School-Wide Performance Bonus Program (Marsh et al., 2011). In addition, when we ran a basic probit model to explore the relationship between observable school-level characteristics and school participation in the retention bonus program, we found that the percentage of Level 5 teachers in a school was the strongest predictor of participation. This is logical given that Level 5 teachers are most likely to benefit from the program.[12]

Figure 1 further delineates teachers in the 56 schools that volunteered to participate by a teacher's eligibility status for a US$5,000 retention bonus. Approximately 26% of teachers, or 520 teachers, did not have sufficient classroom observation data for program participation due to being flagged as partial-year exemption teachers, meaning (a) they did not teach an adequate number of days in their current school to receive an observation or (b) they changed schools during the academic year.[13] A total of 1,012 teachers, or about one half of the teacher sample, were not eligible for the bonus because they received a Level 4 or lower overall performance evaluation rating, though 9 of the 759 teachers receiving below a Level 5 rating who returned to a Priority School the following year still received a US$5,000 retention bonus. Of the 473 Priority School teachers that earned a Level 5 rating for the 2012–2013 school year, 80% (377 teachers) remained within a Priority School, of which 321 (or 85%) received a US$5,000 bonus.[14]

TABLE 1
Summary Statistics on Schools by Participation Status of School

	(1)	(2)	(3)	(4)
	Full sample	Participated	Did not participate	Difference (2) − (3)
School level				
Elementary	55.56%	54.00%	59.09%	−5.09%
Middle	65.28%	66.00%	63.64%	2.36%
High	12.50%	10.00%	18.18%	−8.18%
Urbanicity				
City	90.24%	92.86%	84.62%	8.24%
Suburb	8.54%	5.36%	15.38%	−10.03%
Town	1.22%	1.79%	0.00%	1.79%
Rural	0.00%	0.00%	0.00%	—
School size	528.744	529.911	525.773	−4.138
n	82	56	26	

*Significant at the 10% level. **5% level. ***1% level.

Table 1 displays summary statistics on the characteristics of schools that participated and did not participate in the retention bonus program. Participant and nonparticipant campuses are relatively similar across school level (elementary, middle, and high school), urbanicity (city, suburb, town, and rural), and school size. More than 90% of participating campuses come from urban setting and are categorized as elementary or middle schools, whereas the average size of enrollment is about 530 students.

Table 2 displays descriptive information on students and teachers for the Priority Schools that participated and did not participate in the retention bonus program. Across all observable characteristics, the samples are rather similar, although participating campuses have slightly fewer White students (1.98% vs. 4.16%). Participating campuses also have modestly greater percentage of students qualifying for free and reduced-price lunch programs (90.46% vs. 86.42%) and female teachers (79.41% vs. 72.28%).

Analytic Strategy

The theory behind retention incentives implies that the opportunity to earn an additional US$5,000 in income for working in a Priority School for an additional year will induce bonus eligible teachers to remain in a Priority School the following year, as the economic benefit is greater given the standardized remuneration practices in the public education sector. We refer to the outcome of retention from the 2012–2013 to 2013–2014 school years as Y. Our treatment variable, whether a teacher received an overall teacher evaluation of 425 or greater, which designates them as a Level 5 teacher, will be denoted as T. We are interested in the impact of being eligible for a retention bonus, $X \geq 425$, on retention from the 2012–2013 to 2013–2014 school years.

In a sharp RD design framework, we would construct a comparison group for high-performing teachers (Level 5) in Priority Schools participating in the retention bonus program consisting of similar teachers that are not able to receive a bonus because they scored just below the Level 5 teacher performance threshold. Because a teacher's eligibility for the bonus program is determined by a score on a quantitative, continuous variable with a strict cutoff (a Level 5 teacher rating, which equates to a 425 or higher on the overall teacher evaluation rating variable), teachers slightly below the 425 cutoff that work in Priority Schools participating in the bonus program can serve as a control to estimate unbiased average treatment effects of the bonus program within specified bandwidths.[15] The number of points a teacher is above or below a Level 5 rating (highly effective), which equates to their overall performance score minus 425 points, becomes the running

TABLE 2
Summary Statistics on Students and Teachers by Participation Status of School

	(1)	(2)	(3)	(4)
	Full sample	Participated	Did not participate	Difference (2) − (3)
Student characteristics				
Female	48.22%	48.51%	47.49%	1.03%
White	2.60%	1.98%	4.16%	−2.18%*
Black	93.23%	93.52%	92.50%	1.02%
Asian	0.25%	0.23%	0.29%	−0.06%
Hispanic	3.76%	4.10%	2.90%	1.20%
Other	0.16%	0.16%	0.14%	0.02%
Free and reduced-price lunch	88.85%	90.04%	85.83%	4.21%*
Special education	16.96%	16.51%	18.09%	−1.57%
English-language learners	2.26%	2.40%	1.92%	0.48%
Teacher characteristics				
Female	77.21%	79.41%	72.28%	7.13%**
White	31.09%	30.14%	33.21%	−3.07%
Non-White	68.91%	69.86%	66.79%	3.07%
With bachelor's (BA)	33.67%	34.56%	31.66%	2.90%
With master's (MA)	36.46%	36.07%	37.32%	−1.25%
More than master's	29.87%	29.36%	31.03%	−1.67%
Teaching tested grade-subject	39.35%	39.03%	40.19%	1.15%
Average years of experience	11.70	11.73	11.66	0.07
Average salary	US$52,414.40	US$52,811.91	US$51,523.97	US$1,287.95
n	81	56	26	

*Significant at the 10% level. **5% level. ***1% level.

or forcing variable. When this number is equal to or greater than 0, the teacher is considered a highly effective instructor (a Level 5 teacher) and eligible for a US$5,000 retention bonus if he or she teaches in a Priority School the following year. When the value on the forcing variable is less than 0, a teacher's overall performance score is below 425 and he or she is not eligible for a retention bonus irrespective of their decision to work in a Priority School the following school year. Of course, this identification strategy may understate the true treatment effect if teachers that just missed being a Level 5 teacher return to a Priority School the following year in hopes that the program will be around for another year and that they earn a Level 5 teacher rating.

However, as displayed in Figures 1 and 2a to 2c, noncompliance is present, which makes a sharp RD identification strategy invalid. Fifty-six Level 5 teachers, or approximately 12% of our Level 5 teacher sample, did not receive a retention bonus even though they returned to a Priority School during the 2013–2014 school year (no-shows). Nine Level 4 teachers received a retention bonus even though they were not eligible under program guidelines (crossovers). To deal with endogeneity problems arising from partial compliance, we implement an instrumental variables estimation strategy using the exogenous assignment to the treatment (score on running variable) as an instrument for the effective participation in the retention bonus program. In this sense, treatment is no longer deterministically related to crossing a threshold but there is a jump in the probability of treatment at 425.

Thus, to estimate the impact of the bonus program on retention, we adopt a fuzzy RD design in which the treatment status is probabilistically determined as a discontinuous function of our running variable following procedures recommended in Lee and Lemieux (2009). The relationship between the probability of treatment and the performance score threshold can be written as

$$Pr(D=1 \mid X=x)=\gamma+\delta T+g(x-c),$$

where $T=1[X \geq c]$ indicates whether the assignment variable exceeds the eligibility threshold c and D indicates whether or not a teacher receives a retention bonus.[16] Because $D=\Pr(D=1|X=x)+v$, where v is an error term independent of X, the fuzzy RD design can be described as

$$Y=\alpha_l+\tau D+f(X-c)+e$$

$$D=\gamma+\delta T+g(X-c)+v.$$

We can substitute the treatment determining equation, which is estimated as a linear probability model, into the outcome equation to get the reduced form equation, which can be expressed as

$$Y=\alpha_r+\tau_r T+f_r(X-c)+e_r$$

where $\tau_r=\tau * \delta$. We address the bandwidth selection problem using Imbens and Kalyanaraman's (2012) optimal bandwidth calculation. We also estimate models using kernel-weighted local polynomial smoothing with triangular case weights. It is important to note that our estimation strategy identifies treatment-on-treated effect.

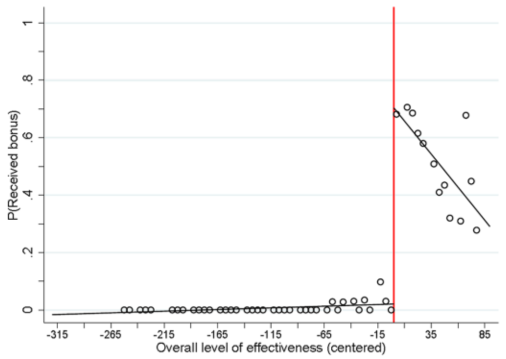
Assumptions of RD Identification Strategy. There are three fundamental assumptions critical to the RD identification strategy. First, unobserved characteristics vary continuously around the teacher effectiveness cutoff with observable characteristics used to determine bonus eligibility. We investigate whether there are baseline imbalances between treatment and control teachers using teacher covariates as the dependent variables. We are interested in whether there are jumps at the Level 4/5 performance threshold. We use Imbens and Kalyanaraman's (2012) optimal bandwidth calculation and perform these tests for the full sample of teachers as well as tested-subject teachers only and untested-subject teachers only.

As displayed in Table 3, we reject the hypothesis for all observables that the means of the treatment and the control condition teachers are statistically different. Observed baseline covariates are locally balanced on either side of the cutoff, which we use to infer that the treatment indicator is locally randomized. These estimates are not sensitive to the inclusion of 44 teachers that did not receive a bonus even though they achieved a Level 5 rating. We detect a significant difference in race/ethnicity (White vs. non-White) in the tested-subject teacher sample. However, this may be a result of multiple hypothesis testing, that is, one statistically significant comparison out of 22 inferences. Furthermore, the race indicator is not statistically significant when analyzing the retention bonus program impact which suggests that it is not a true threat to internal validity of study.

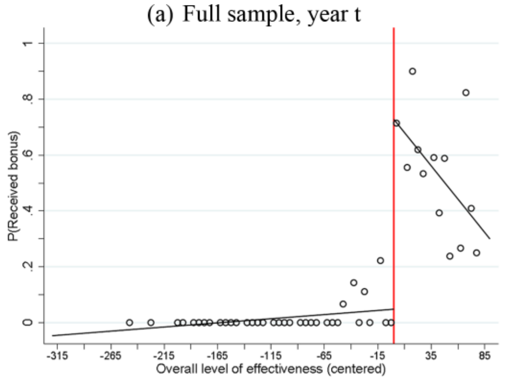
The second critical assumption is that our forcing variable, X, has not been manipulated to affect who receives treatment. This assumption is critical to the internal validity of our study. To examine if our forcing variable, X, has been manipulated, we are interested in whether the aggregate distribution of X is discontinuous. The concern is that self-interested teachers may try to influence their overall performance score rating to increase the likelihood of being eligible for a retention bonus. Although this is unlikely given that the program was implemented relatively late in the school year and multiple components that factor into a teacher's overall performance rating make it difficult for a teacher to artificially inflate performance scores, we implement the formal procedure developed by McCrary (2008).

Results from the McCrary sorting test, as displayed in Figures 3a to 3c, suggest a slight jump in X at c; however, the differences are not statistically significant at conventional levels for the full and reduced samples (i.e., tested-subject teachers only and untested-subject teachers only). Furthermore, formal hypothesis tests of the log difference in heights at the cut-point confirm that, in each case, we cannot reject the null hypothesis that the discontinuity at these thresholds is zero (full-sample estimate = 0.244, *SE* = 0.152; tested-subject teacher sample estimate = 0.135, *SE* = 0.282; untested-subject teacher sample estimate = 0.269, *SE* = 0.17). The discontinuity at the threshold is smallest for tested-subject teachers.

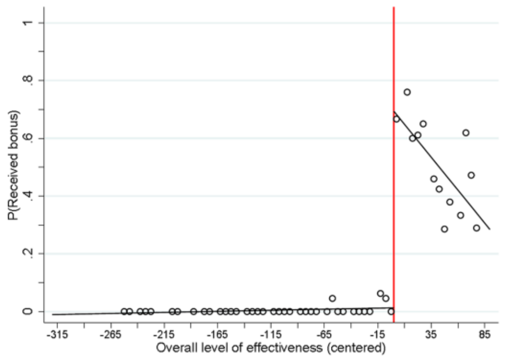
It has also been shown that the McCrary test has weak power to detect manipulation when manipulation heaps into particular values. One could argue this may be problematic because our running variable is not perfectly continuous



(a) Full sample, year t



(b) Tested-subject teachers, year t



(c) Untested-subject teachers, year t

FIGURE 2. *Compliance of bonus receipt.*

(see Figures 4a–4c). We investigate the implication empirically through a series of bounding exercises. We find that if all Level 5 tested-subject teacher scores were manipulated by 5 points, or alternatively if the bottom 15th percentile of Level 5 tested-subject teachers' scores were manipulated, the estimated effect among tested-subject teachers would no longer be statistically distinguishable from zero.[17]

Although manipulation is a real concern in any RD design study, we do not believe it is a serious internal validity concern in the current context. First, we believe our estimate is sensitive to a modest amount of fictitious manipulation because of the limited sample size. In fact, if the same bounding exercise applied to Level 4 teachers being erroneously categorized as such (i.e., they should have been Level 5 teachers), our impact estimates would have been nearly twice as large (est. = 0.40; p = .05). Second, as noted earlier, the program was announced after most formal evaluations had taken place. Third, a teacher's overall level of effectiveness is comprised of multiple measures making it less susceptible to systematic manipulation. Finally, we detect similar patterns in the running variable for the state as a whole and not just in schools participating in the retention bonus program.

A final assumption of the RD identification strategy is that there are no other programs or services with the same eligibility rules, which assures that the bonus program treatment is not confounded with some other treatment. In the current context, we are not aware of any other programs or services with the same eligibility rule that could confound treatment. However, we do want to make brief mention of the generalizability of this kind of study. It is plausible that teachers who are doing extremely well, for example, could respond differently to the bonus than those who are on the low end of the "significantly exceeds expectation" category. Our model predicting the likelihood of retention for all teachers based on their overall effectiveness rating indicates that the higher the rating teachers get, the more likely they are to stay.

Results

Although we find no significant overall effect for the offer of retention bonuses on Level 5 teacher retention, increases in the retention of teachers of tested subjects and grades are both statistically and substantively significant. Below we present primary-effect estimates, a series of robustness checks, and a cost-effectiveness discussion.

Graphical Evidence

It is common in RD studies to present findings in an unrestrictive, visual manner that closely

TABLE 3
Continuity of Teacher Observables by Sample (Imbens & Kalyanaraman's, 2012, Optimal Bandwidth Sample)

	Full sample		Tested-subject teachers		Untested-subject teachers	
	n	Estimate/*SE*	*n*	Estimate/*SE*	*n*	Estimate/*SE*
Female	603	−0.051 (0.091)	267	0.091 (0.151)	437	0.098 (0.316)
White	479	0.051 (0.121)	217	−0.398** (0.164)	356	0.083 (0.145)
Bachelor's	869	0.009 (0.091)	242	0.131 (0.183)	438	−0.036 (0.134)
Master's	718	−0.024 (0.113)	170	−0.269 (0.241)	401	0.128 (0.146)
More than master's	598	−0.026 (0.116)	177	−0.006 (0.226)	325	−0.063 (0.148)
Teaching tested grade-subject	519	−0.069 (0.123)	—	—	—	—
Years of experience	565	3.273 (2.435)	171	6.936 (4.117)	491	1.764 (2.663)
Salary	639	2,159.531 (2,598.395)	157	6,905.105 (4,496.034)	394	53.732 (3,299.798)

*Significant at the 10% level. **5% level. ***1% level.

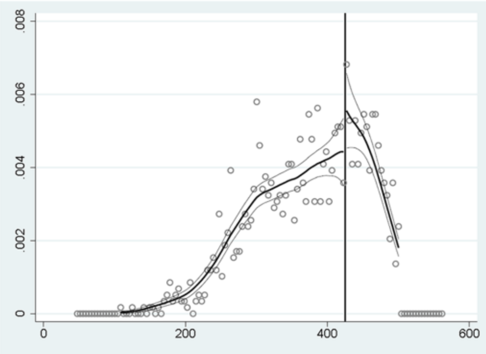
models parametric results from the main regression analysis. Figures 5a to 5c presents the conditional means for the probability of remaining in a high Priority School as a function of a teacher's underlying overall level of effectiveness score for each of the three samples of interest: all teachers (full sample), tested-subject teachers only, and untested-subject teachers only.

For the full sample of teachers, as displayed in Figure 5a, we find a slight decrease in the probability of Level 5 teachers remaining in a high Priority School. The size of the difference at the discontinuity is −0.04. For the tested-subject only sample of teachers, as displayed in Figure 5b, we find a noticeably different pattern in the data. Level 5 teachers are more likely to remain teaching in a high Priority School. The size of the difference at the discontinuity is 0.11. Finally, for the untested-subject only sample of teachers, as displayed in Figure 5c, we find an exaggerated version of the pattern for the complete sample of teachers. The size of the difference at the discontinuity is −0.12.

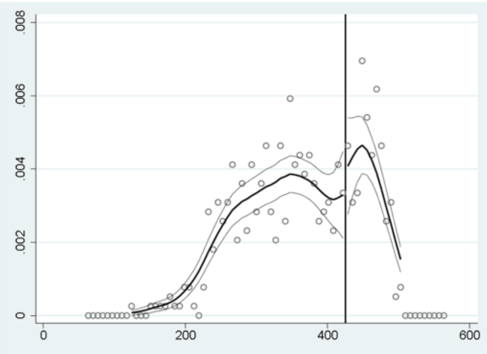
However, note that the gaps for the full sample and the untested-subject teachers are not statistically different from zero when controls are included in the models. The gaps for tested-subject teachers grow and are both statistically and practically significant. These patterns are consistent with the hypothesis that tested-subject teachers are responding to the retention bonus, whereas the opposite is true for untested-subject teachers. That is, given the amount of weight Tennessee's teacher evaluation system attributes to school-level performance for untested-subject teachers, the evaluation system creates a strong incentive for untested-subject teachers to exit the Priority Schools, which are by definition low performing.

Impact of the Retention Bonus Program

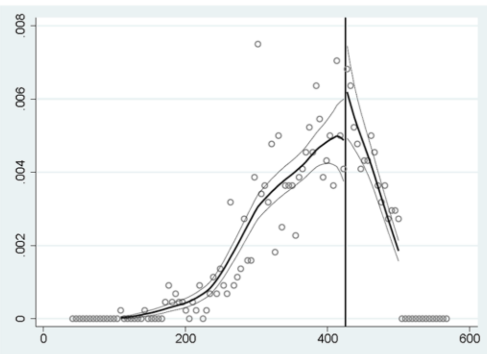
Table 4 reports estimates of the impact of the bonus program on teacher retention. Although we report estimates using several different bandwidth selections, and our results are robust to varying bandwidths, our discussion will focus on estimates from models where the sample is defined by the optimal Imbens and Kalyanaraman's (2012) bandwidth specification. We do not find a statistically significant program effect on teacher retention, though the point estimates are in the hypothesized direction. However, similar to the graphical evidence, when we produce estimates for tested-subject teachers and untested-subject teachers separately, we find that Level 5 tested-subject teachers who receive a retention bonus are approximately 20% more likely to remain teaching in a Priority School when compared with tested-subject teachers just below the Level 5 cutoff. The point estimates for untested-subject teachers are quite small and not different from zero at conventional levels, which suggests that teachers of tested subjects are driving the effect.



(a) Full sample, year t

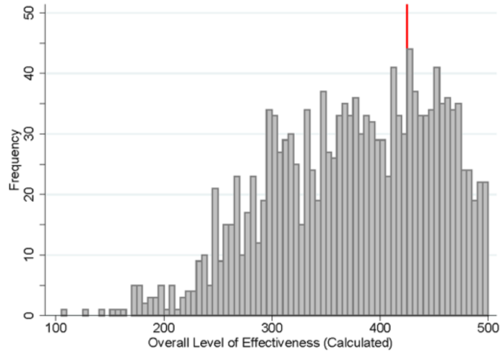


(b) Tested-subject teachers, year t

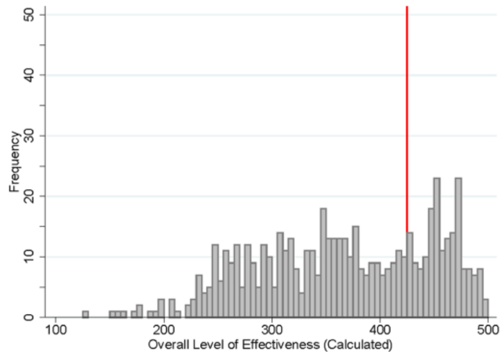


(c) Untested-subject teachers, year t

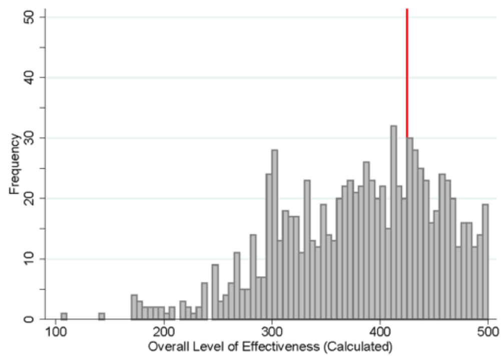
FIGURE 3. *Densities of the retention bonus program assignment variable.*



(a) Full sample, year t



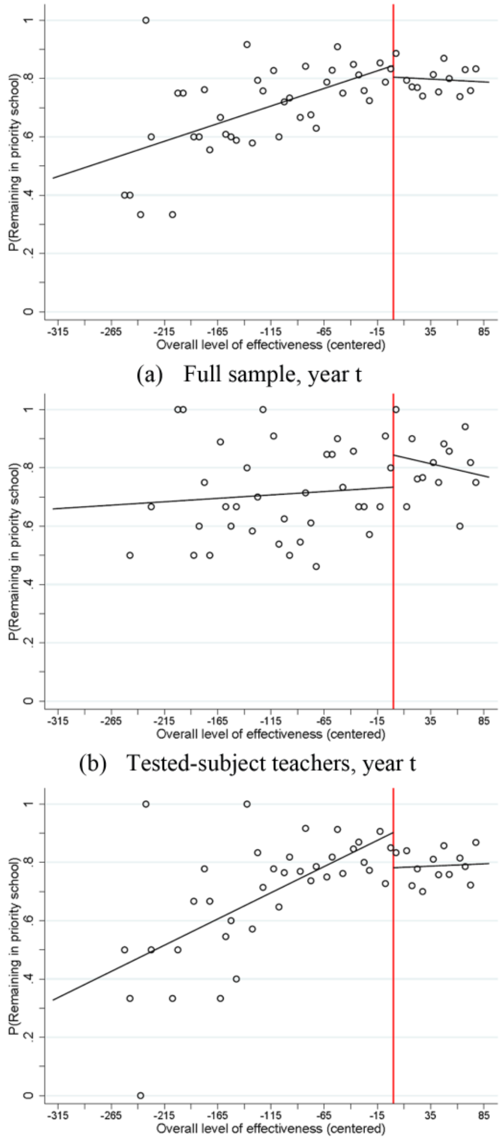
(b) Tested-subject teachers, year t



(c) Untested-subject teachers, year t

FIGURE 4. *Histograms of the retention bonus program assignment variable.*

We find qualitatively similar results when we estimate the same series of models using kernel weights (see Table 5). The same holds true for estimates from bivariate probit-modeling approach (see Table 5). It is important to note that the estimated coefficients from probit models do not quantify the influence of the right hand–side variables on the probability of success (i.e., remaining in a Priority School). Although the sign and level of significance of the estimates are informative, they are generally uninterpretable in isolation as they are parameters of the underlying latent model. As such, in the line below the biprobit estimates, we also report the marginal effects of receiving a bonus on the probability of remaining in a Priority School holding the regressors at their mean value.



(a) Full sample, year t

(b) Tested-subject teachers, year t

(c) Untested-subject teachers, year t

FIGURE 5. *Full, tested-subject, and untested-subject retention effects.*

It is plausible that rather than increasing the retention of Level 5 tested-subject teachers, the bonus policy only resulted in relatively higher rates of retention of Level 5 teachers by depressing retention of Level 4 teachers. One might imagine that the bonus program could have decreased the retention of Level 4 teachers who felt discouraged by the lack of recognition they received relative to other teachers who were just marginally higher rated. We find that this is not the case. The rate of departure for all teachers increased from the year prior to implementation (0.109 vs. 0.134), but this change was noticeably smaller for Level 4 teachers. As such, we argue that the detected effect is a result of increased retention of Level 5 teachers. On the contrary, more Level 4 teachers may choose to remain in high Priority Schools in hopes of achieving Level 5 status in the following year, thereby muting the magnitude of the reported effect.

To supplement our primary modeling approach, we also adopted a difference-in-difference strategy that focuses on the effects of the bonus on the composition of teacher effectiveness in the school as measured by the proportion of teachers classified as high scoring on TVAAS. We estimated models using the proportion of both Levels 4 and 5 TVAAS teachers in a school or the proportion of Level 5 TVAAS teachers in a school as the dependent variable. We exploit 4 years of data on teacher effectiveness (3 before and 1 after the introduction of the bonus) to estimate compositional effects, where we compare the change in average teacher value-added scores among participating eligible schools with those of similar nearly eligible schools that did not offer retention bonuses. Similar nearly eligible schools are defined as the bottom 10% of schools in the state according to the state's composite proficiency rate calculation. We rely on TVAAS scores as a measure of teacher effectiveness in this instance, because the more holistic teacher evaluation ratings were not implemented until the year before the bonus program was implemented.

For this supplementary analysis, we first compare eligible and ineligible schools, which represent intent-to-treat (ITT) estimates. That is, we do not account for whether or not Levels 4 and 5 teachers in the school receive a bonus. We find consistently positive, modest-sized effects of the program. There is an approximate 16 percentage point increase in the proportion of teachers rated either Level 4 or Level 5 on TVAAS in a bonus eligible school post-implementation. Similarly, there is an approximate 13 percentage point increase in the proportion of teachers rated Level 5 on TVAAS, showing a consistent pattern of increased retention for both of our value-added proxies for bonus eligibility.

We next produce treatment-on-treatment (TOT) estimates where the indicator for bonus eligibility is

TABLE 4
Local Linear Regressions of the Impact of Incentive on Retention

	Full sample		Tested-subject teachers		Untested-subject teachers	
	n	Estimate/*SE*	*n*	Estimate/*SE*	*n*	Estimate/*SE*
Optimal bandwidth	587	0.138 (0.088)	369	0.198* (0.110)	439	0.000 (0.108)
75% of optimal bandwidth	437	0.066 (0.104)	310	0.317** (0.124)	331	0.065 (0.117)
110% of optimal bandwidth	644	0.080 (0.088)	386	0.202* (0.107)	477	0.026 (0.100)
125% of optimal bandwidth	724	0.110 (0.080)	415	0.224** (0.100)	533	−0.004 (0.095)
Full sample	1,485	0.045 (0.053)	574	0.217*** (0.082)	911	−0.046 (0.071)

Note. Robust standard errors in parentheses. All models include a full set of teacher and school controls.
*Significant at the 10% level. **5% level. ***1% level.

TABLE 5
Sensitivity Analysis of Impact of Retention Bonus Estimates (Imbens & Kalyanaraman's, 2012, Optimal Bandwidth Sample)

	Full sample		Tested-subject teachers		Untested-subject teachers	
	n	Estimate/*SE*	*N*	Estimate/*SE*	*n*	Estimate/*SE*
Primary model	587	0.138 (0.088)	369	0.198* (0.110)	439	0.000 (0.108)
Primary model with triangular weights	734	0.105 (0.086)	416	0.243** (0.106)	536	0.025 (0.102)
Biprobit	587	1.192** (0.605)	369	1.224** (0.597)	439	0.052 (0.977)
Marginal effects from biprobit model above		0.206 (0.080)		0.210 (0.087)		0.013 (0.230)

Note. Robust standard errors in parentheses.
*Significant at the 10% level. **5% level. ***1% level.

replaced by an indicator for participation in the program, predicted by eligibility in a two-stage least squares instrumental-variable framework. As expected, the TOT estimates are larger than the ITT estimates. There is an approximate 34 percentage increase in the proportion of teachers rated either Level 4 or Level 5 according to TVAAS in a bonus eligible school post-implementation. Similarly, there is an approximate 27 percentage point increase in the proportion of Level 5 teachers.

Robustness Checks

We test the robustness of our primary findings through a series of placebo tests. We first estimate our primary analytic strategy using a series of placebo or false thresholds. The conventional method for conducting this test is inserting dummy variables for placebo thresholds into a single RD model utilizing the complete sample of teachers. We place placebo thresholds at fixed quartiles rather than fixed distances from the cutoff to ensure enough variability around the chosen threshold. We then test the joint significance using an F test. As displayed in Panel A of Table 6, we find that the estimated retention effects are only present at the Level 4/5 threshold; that is, we only detect an effect at the threshold with real world meaning to the teacher labor force in these schools. Furthermore, the estimate for the impact of the bonus program on retention is virtually identical to those reported in the primary model (i.e., 0.198 in primary model and 0.209 in model with placebo thresholds).

We next explore the robustness of our findings to various samples of teachers and schools where

TABLE 6

Placebo Tests of Impact of Retention Bonus Estimates (Imbens & Kalyanaraman's, 2012, Optimal Bandwidth Sample)

	Panel A: Placebo thresholds		
	Full sample	Tested-subject teachers	Untested-subject teachers
	Estimate/*SE*	Estimate/*SE*	Estimate/*SE*
Placebo threshold at 25th quartile (OLE = −108.5)	−0.099 (0.053)	−0.111 (0.090)	−0.048 (0.065)
Placebo threshold at 50th quartile (OLE = −42.6)	−0.064 (0.057)	−0.008 (0.076)	−0.070 (0.049)
Actual threshold (OLE = 0)	**0.071 (0.058)**	**0.209** (0.092)**	**0.002 (0.076)**
Placebo threshold at 75th quartile (OLE = 13)	−0.068 (0.057)	−0.081 (0.086)	−0.060 (0.076)
F test testing joint significance of placebo thresholds	6.08	2.61	2.78

	Panel B: Alternate samples					
	Full sample		Tested-subject teachers		Untested-subject teachers	
	n	Estimate/*SE*	*n*	Estimate/*SE*	*n*	Estimate/*SE*
Primary model	587	0.138 (0.088)	369	0.198* (0.110)	439	0.000 (0.108)
Same schools, year $t-1$ (year prior to implementation)	586	0.065 (0.055)	167	−0.035 (0.134)	484	0.035 (0.062)
Same schools, year $t-1$ (year prior to implementation) with year t bandwidth	432	0.024 (0.069)	254	0.066 (0.087)	369	0.094 (0.068)
High-poverty schools not eligible to participate in program, year t	3,247	0.006 (0.009)	1,591	−0.020 (0.015)	1,875	0.019 (0.011)
High-poverty schools not eligible to participate in program, year $t-1$	2,259	0.026** (0.013)	1,263	−0.012 (0.020)	1,236	0.029 (0.016)

Note. Robust standard errors in parentheses. OLE = overall level of effectiveness.
*Significant at the 10% level. **5% level. ***1% level. All models in Panel A condition on a linear spline of the assignment variable.

we would not expect to find an effect at the Level 4/5 threshold because the retention bonus program did not exist. As displayed in Panel B of Table 6, we first restrict our sample to schools that participated in the retention bonus program, but conducted our analyses in the school year prior to the implementation of the bonus program. We do not find a significant difference in retention rates of Level 5 teachers just above the cutoff compared with the retention rates of teachers just below the Level 5 cutoff. Given the narrow bandwidth selected for the tested-subject teacher sample (n = 167), we re-estimate the model using the year t bandwidth that includes 254 teachers. Once again, we do not detect an effect. The same holds true when we restrict the sample to all high-poverty schools in Tennessee that are not eligible to participate in the retention bonus program during the year in which the retention bonus program was implemented. We find similar results when we use the same sample but in the year prior to implementation.

Cost-Effectiveness Discussion

To allow for better comparison with alternative interventions that seek to elevate student achievement in struggling schools, we estimate the cost-effectiveness of the bonus program across a range of plausible impacts on teachers' retention decisions and the projected benefits their retention ultimately confers on the students they teach on remaining in a Priority School. The cost-effectiveness of the policy is primarily a function of the strength of the intervention in improving the retention of a Level 5 teacher in a Priority School. Teachers who accepted bonuses had overall teacher effectiveness ratings more than a full standard deviation above the state average, and the average teacher hired by Priority Schools was rated roughly two thirds of a standard deviation below the state average. Thus, for every teacher that is retained as a result of the bonus, students taught by that teacher rather than the likely replacement experience an increase in estimated teacher effectiveness of 1.64 standard deviations.

Although recent estimates of variance in teacher effectiveness differ (Hanushek & Rivkin, 2010), for simplicity, we follow Hanushek's (2011) arguably conservative effect size of 0.20 standard deviations in improved test performance for students assigned a teacher one standard deviation more effective (roughly the equivalent to Krueger's, 1999 estimated effect of reducing class size by one third). Thus, the projected effect size on student achievement for those taught by retained teachers the following year would range from roughly 0.20 to 0.33 standard deviations depending on whether the teacher would have been replaced by the average Tennessee teacher (1.03 *SD* less effective than the average bonus recipient) or the average teacher hired by a Priority School in previous years (1.64 *SD* less effective). To put this in perspective, Hanushek (2011) estimates that a single year of exposure to a teacher one standard deviation more effective can increase the total lifetime earnings of a 25 student class by greater than US$500,000, net present value.

In total, the state distributed slightly more than US$2.1 million in US$5,000 retention bonuses to 361 highly effective teachers, who agreed to stay at a Priority School during the 2013–2014 school year. After employer shares of taxes and administrative costs, the total cost to the state associated with each bonus paid was roughly US$6,000. However, if the goal of the investment was to retain highly effective teachers, it is informative to calculate the costs per teacher who would have otherwise left the Priority School. Although our fuzzy RD design estimated treatment effects are local to the cutoff score and not generalizable to Level 5 teachers who had particularly high scores, one can easily account for uncertainty in the effect estimates by estimating costs for a range of plausible program impacts. For example, if 20% of teachers who received bonuses stayed at a Priority School as a result of the program, the cost per teacher retained would be roughly US$30,000. If the teachers who were retained as a result of the bonus taught an average of 25 students, the cost per affected student would be roughly US$1,200. If 90% of the teachers who received bonuses would have stayed in the absence of the bonus, the cost per affected pupil would be US$2,400.

Compared with other interventions such as NCLB's supplemental education services (Heinrich et al., 2013; Springer et al., 2014), summer school (Jacob & Lefgren, 2004), or reduced class size (Krueger, 1999), the costs per affected pupil associated with the bonuses are modest, particularly given the relative effects on student achievement predicted for radical shifts in teacher effectiveness associated with the retention of highly effective teachers (i.e., rated Level 5). Notably, none of these alternative interventions costs are offset by reductions to the significant administrative financial burden associated with teacher turnover, including separation, hiring, and training costs (Synar & Maiden, 2012).

Although our projected effects on students may seem optimistic, there are several reasons to consider them plausible lower bounds. Recent studies have found significant effects on student achievement through peer effects on colleagues who work with highly effective teachers (Jackson & Bruegmann, 2009) and negative effects on students of teacher churn within and across schools (Ronfeldt, Loeb, & Wyckoff, 2012). None of these benefits are quantified in the estimates presented here. Also, all the above estimates presume the effects on retention only last one year. A more accurate portrayal of the long-run benefits of teacher retention on student achievement would also account for effects in subsequent years for the portion that continue to work in the Priority School each successive year.

Conclusion

Retention bonuses tied to estimates of teacher effectiveness could serve as a tool for policymakers to improve the quality of the teachers instructing disadvantaged students without implementing layoffs or other punitive measures. Because teachers across the effectiveness spectrum leave high-poverty, high-minority schools regularly on their own volition and are generally replaced by less experienced and effective teachers, bonuses that retain the teachers at the higher end of the effectiveness distribution can have substantial impacts on the quality of a school's faculty. In contrast to policies that would target teachers with poor evaluations or low value-added estimates for dismissal, introducing churn and instability, the retention bonuses mitigate unwanted turnover and have the potential to strengthen leadership and institutional knowledge among the schools' faculty while avoiding financial burdens associated with turnover.

As is true for any policy that relies on observations and test-score-based value-added estimates to differentiate teachers, the benefits of retention bonuses are only as strong as the measures of effectiveness are accurate. If, for example, the designation of "highly effective," based on the composite evaluation, is functionally random or even falls more frequently on less desirable teachers, then the policy would not have the desired effects on the teaching pool and could have discouraging effects on effective teachers who failed to receive the designation and monetary reward. However, the negative consequences of such miscategorizations in the context of retention bonuses are seemingly less severe than those for teacher quality policies that rely on terminations.

Although this study offers an important contribution to a relatively slender body of research using rigorous research designs to estimate the impact of teacher retention bonus programs, it is important to acknowledge several limitations. First, the late timing of implementation limited the opportunity for principals to take advantage of the program as a retention incentive. The program was not formally announced until March 2013, and many districts open the window for transfers as early as the first week of March. By the time eligible principals had applied to and been confirmed to offer bonuses, many teachers may have already made their decisions about whether or not to exit. Late timing may have decreased the level of awareness among eligible teachers; thus, the estimates presented in this study represent a conservative lower bound for what we might expect in the future. At the same time, it is worth noting that the late implementation limits the potential for principals and teachers to game the system by inflating observation scores to increase certain teachers' odds of receiving the bonus. Thus, although the timing of implementation limits both the sample and predicted effect size, it potentially strengthens the internal validity of our RD identification strategy.

Second, considerable noncompliance with the rules established for the distribution of the bonuses makes it difficult to rule out the potential that principals were offering bonuses selectively based on some alternative criterion that happened to align relatively well with the cutoff that provides the basis for the discontinuity. The use of the Level 5 assignment rule as an instrument allows us to isolate the variation in bonus recipients that was attributable to the functionally random distinction between teacher scores slightly above and below the cutoff for teachers to become a Level 5. However, if principals were aware of the cutoff and inflated teachers' observation scores who they wanted to stay or thought were more likely to stay, then the estimated effect of the bonus program could in part represent some unobserved difference between the recipient teachers and nonrecipients that was not captured by the full set of controls. As we stated above, though, this type of gaming bias is made less likely due to the small window of time around the implementation, and analyses of evaluation scores separate from value-added measures uncovered no consistent irregularities.

Despite these substantial implementation difficulties, and a relatively small sample of participating schools, we find some encouraging evidence of a causal link between the bonus offer and retention of high-quality teachers. Estimates are particularly positive among teachers of tested subjects and grades, for whom we have the most credible estimates of classroom effectiveness and are perhaps the most difficult to retain in schools facing strict oversight and accountability. On several measures, schools that participated in the bonus program appear to be slightly more disadvantaged than even the other eligible Priority Schools, ensuring that the bonuses provided additional compensation to effective teachers working in some of the most challenging settings in the state.

Moving forward, policymakers implementing similar programs could benefit from additional steps to ensure that principals and teachers in eligible schools are aware of the bonuses and are supported throughout the implementation process to ensure compliance with program guidelines. Earlier implementation and efforts to improve awareness would increase the likelihood of this type of performance-based retention bonus serving as both an incentive to stay at a hard-to-staff school and a reward for laudable work in a vital setting. Future research should seek to illuminate teacher and principal perceptions of the bonuses, barriers to them reaching the teachers they are designed to target, and the mechanisms by which they influence teachers to stay.

In the context of prior literature on the most comparable monetary retention incentives, the roughly 20% increase in rate of retention for the US$5,000 bonus could be viewed on its face as evidence for diminishing returns to increases in the size of the reward, as Clotfelter et al. (2008) found similar effects for a bonus roughly half as large. However, it is important here to note that in contrast to the NC bonuses, which had no requirements around measured effectiveness, the Tennessee bonus program targeted only teachers at the top end of the effectiveness spectrum. In their 2011 examination of the relationship between teachers' preservice credentials (an admittedly crude proxy for effectiveness), salary, and school characteristics, Clotfelter, Ladd, and Vigdor concluded that teachers with the strongest credentials require substantially larger salary differentials to remain in segregated high-poverty, high-minority schools than their less-qualified peers. Thus, the relatively larger bonuses in Tennessee could be understood as roughly equivalent to the NC bonuses, for the conceivably more marketable population of teachers classified as "significantly exceeding expectation."

Relatedly, it is worth reiterating the limitations of generalizability of the main findings presented in this article. First, as noted above, the Tennessee retention bonuses were offered only to teachers with the highest effectiveness rating (Level 5 on 1–5 scale) in the lowest performing schools (bottom 5% in the state on composite of test scores and graduation rates). It is plausible that smaller bonuses could yield similar results in even moderately less disadvantaged schools with smaller concentrations of high-need students. Similarly, if the aim was simply to reduce turnover among teachers who meet expectation (as opposed to significantly exceeding them), a more inclusive eligibility criterion could potentially yield larger effects. Furthermore, as is always the case for RD design, our results are local to teachers who were close to the eligibility cutoff, making it impossible to extrapolate findings to teachers who had the highest ratings in the highest rating category. Finally, we find no impacts for teachers of untested subject and grades, a group that represents the majority of the teacher labor force. Notably, these teachers face competing policy incentives due to the inclusion of school-level growth scores in their overall evaluation rating, which is by construction low in Priority Schools.

In sum, although our findings from Tennessee are consistent with a theory that even relatively small economic rewards can help mitigate the problematic exit patterns of teachers in high-need schools, they also highlight the need for further examination of the interaction of working conditions, nonmonetary policy incentives, and compensation to facilitate a more equitable distribution of the vital resource that is high-quality teachers. The small but substantively significant effects for highly effective teachers of often-prioritized tested-subject areas inspire optimism about the potential for salary supplements as a tool in combating inequality. Our relatively crude cost-effectiveness analyses also indicate that such expenditures could yield significant benefits to students under a range of plausible effects on both retention and achievement. However, the lack of impacts on teachers of untested subjects and grades paired with the relatively large percentage of teachers of tested subjects and grades who chose to leave and forgo the substantial payment underscore the complexity of keeping effective teachers in challenging environments with high concentrations of poor and minority students.

Authors' Note

Any errors remain the sole responsibility of the authors. The views expressed in this article do not necessarily reflect those of sponsoring agencies or individuals acknowledged.

Acknowledgments

The authors appreciate helpful comments and suggestions from Dale Ballou, Richard Blissett, Matthew Kraft, Eric Taylor, and seminar participants at Vanderbilt University and at the annual meetings of the Association for Education Finance and Policy and Association for Public Policy Analysis and Management. The authors acknowledge the many individuals at the Tennessee Consortium and the Tennessee Department of Education for providing data and expert insight to conduct their analyses, in particular, Susan Burns, Sara Heyburn, Bing Howell, Trish Kelly, Erin O'Hara, Art Peng, Matthew Pepper, and Nate Schwartz.

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 study was supported by the Tennessee Consortium on Research, Evaluation, and Development (the Consortium) at Vanderbilt University's Peabody College, which is funded by the State of Tennessee's Race to the Top grant from the United States Department of Education (Grant #S395A100032).

Notes

1. The sorting of teachers across schools exacerbates racial and poverty-related achievement gaps. Schools enrolling children from the most disadvantaged backgrounds are more likely to be staffed by teachers graduating from less competitive colleges, teachers instructing out-of-field, and novice teachers (Clotfelter, Ladd, & Vigdor 2011; Iatarola & Stiefel, 2003; Lankford, Loeb, & Wyckoff, 2002; Peske & Haycock, 2006). Teacher effect research consistently finds that novice teachers (e.g., first or second year teachers) produce smaller achievement gains for their students than more experienced teachers (Aaronson, Barrow, & Sander, 2007; Henry, Bastian, & Fortner, 2011; Kraft & Papay, 2014; Rivkin, Hanushek, & Kain, 2005). The net result is that children enrolled in schools with high concentrations of disadvantaged students have greater exposure to less-qualified instructors.

2. See Kolbe and Strunk (2012) for a typology of policies and practices regarding economic incentives.

3. Similarly, a simulation study by Clotfelter et al. (2011) suggests that teachers with better credentials demand a significantly larger pay differential to stay in disadvantaged schools than their peers with average credentials.

4. Tennessee Department of Education (TDOE) and the Governor's office also implemented a teacher signing bonus program. To help attract the most effective teachers to Priority Schools, a signing bonus of US$7,000 was offered to every new Level 5 teacher who transferred from a non-Priority School into a Priority School during the 2013–2014 school year. Only 59 teachers received the signing bonus. We do not evaluate this aspect of the program due to small sample size.

5. Teachers continue to push back on Tennessee's teacher evaluation process (Johnson, 2014).

6. Teachers with the top rating are only required to participate in one formal observation in the following year, with two walk-through observations.

7. A number of studies have investigated the effect of signing bonuses on teacher recruitment (Fowler, 2003; Liu, Johnson, & Peske, 2004) as well as the effect of a salary increase on teacher retention (Hough, 2012).

8. Two recent working papers add insights about the effectiveness of recruitment incentives for specific credentials. Cowan and Goldhaber (2015) find that a targeted teacher incentive policy in Washington State increased the proportion of National Board Certified Teachers in high-poverty schools, though they did not find evidence that this compositional change affected student achievement. Feng and Sass (2015) investigate the effect of Florida's critical teacher shortage loan forgiveness program, finding that the loans forgiveness decreased attrition of teachers in shortage areas and that teachers were more responsive to larger payments.

9. For more information on Tennessee Value-Added Assessment System (TVAAS), see http://www.tn.gov/education/topic/tvaas.

10. For more information on CODE, see http://team-tn.org/evaluation/data-system/

11. Program guidelines and participation sign-up procedures can be found here, https://news.tn.gov/sites/default/files/Bonus%20and%20retention%20application.pdf

12. The presence of nonparticipants is an external validity concern in evaluations of social programs. However, nonparticipants do not seem egregious in our application given that fewer than 10 schools (less than 12% of all eligible campuses) with Level 5 teachers

opted not to participate. Nonetheless, it is important to note that some schools with eligible teachers opted not to participate.

13. More information on partial-year exemption status can be found at http://team-tn.org/tag/pye/.

14. Data on the teacher retention bonus program contained one teacher record associated with two Priority Schools and five cases that did not merge on to data maintained by the Consortium. The duplicate teacher record was assigned to a single school where the assignment was based on where she appeared most frequently across various management information systems. After exhaustive attempts to reconcile the five anomalous cases, we decided to drop these cases from the analysis file as we could not locate pertinent information on these cases. We were also unable to obtain detailed information on why 56 Level 5 teachers did not receive a retention bonus.

15. This type of regression discontinuity (RD) design has been shown to produce unbiased, valid estimates of program effects approximating a randomized experiment (see, for example, Angrist & Lavy, 1999; Black, 1999; Cook, 2008; Hahn, Todd, & Van der Klaauw, 2001; Imbens & Lemieux, 2008).

16. As noted in Lee and Lemieux (2010), although the probability of treatment is modeled as a linear probability model, it does not impose any restrictions on the probability because $g(x - c)$ is unrestricted on both sides of the cutoff c, while T is a binary indicator. So there is no need to express model using a probit or logit model.

17. The point estimate is 0.099 and the standard error is 0.1744.

References

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in Chicago high schools. *Journal of Labor Economics*, *25*, 95–136.

Angrist, J. D., & Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class sized on scholastic achievement. *Quarterly Journal of Economics*, *114*, 533–575.

Balch, R., & Springer, M. G. (2014). *An evaluation of a pay for performance program in Texas* (Working paper). Nashville, TN: Vanderbilt University.

Black, S. (1999). Do better schools matter? Parental valuation of elementary education. *Quarterly Journal of Economics*, *33*, 62–87.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014). Measuring the impact of teachers I: Teacher value-added and student outcomes in adulthood. *American Economic Review*, *104*, 2633–2679.

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., & Vigdor, J. L. (2011). Teacher mobility, school segregation, and pay-based policies to level the playing field. *Education Finance and Policy*, *6*, 399–438. Retrieved from http://www.mitpressjournals.org/efp

Cook, T. D. (2008). Waiting for life to arrive: A history of the regression-discontinuity design in psychology, statistics, and economics. *Journal of Econometrics*, *142*, 636–654.

Cowan, J., & Goldhaber, D. (2015). *Do bonuses affect teacher staffing and student achievement in high-poverty schools? Evidence from an incentive for National Board Certified Teachers in Washington* (CEDR Working paper). Seattle, WA: Center for Education Data & Research.

Dee, T., & Wyckoff, J. (2013, October). *Incentives, selection, and teacher performance: Evidence from IMPACT* (NBER Working Paper No. 19529). National Bureau for Economic Research: Cambridge, MA.

Ehlert, M. W., Pepper, M. J., Parsons, E. S., Burns, S. F., & Springer, M. G. (2013). *Educator evaluation in Tennessee: Initial findings from the 2013 First to the Top Survey*. Nashville: Tennessee Consortium on Research, Evaluation, and Development.

Feng, L. (2009). Opportunity wages, classroom characteristics, and teacher mobility. *Southern Economic Journal*, *75*, 1165–1190.

Feng, L., & Sass, T. R. (2012). *Teacher quality and teacher mobility* (Andrew Young School of Policy Studies Research Paper Series 12-08). Atlanta, GA: Georgia State University.

Feng, L., & Sass, T. (2015). *The impact of incentives to recruit and retain teachers in "hard-to-staff" subjects: An analysis of the Florida Critical Teacher Shortage Program* (Calder Center Working Paper No. 141). Washington, DC: American Institutes for Research.

Glazerman, S., Protik, A., Teh, B., Bruch, J., & Max, J. (2013). *Transfer incentives for high-performing teachers: Final results from a multisite randomized experiment*. Washington, DC: U.S. Department of Education's Institute of Education Sciences.

Goldhaber, D., Lavery, L., & Theobald, R. (2015). Uneven playing field? Assessing the teacher quality gap between advantaged and disadvantaged students. *Educational Researcher*, *44*, 293–307.

Hanushek, E. A. (2011). The economic value of higher teacher quality. *Economics of Education Review*, *14*, 101–117.

Hanushek, E. A., Kain, J., O'Brien, D., & Rivkin, S. (2005, February). *The market for teacher quality* (NBER Working Paper No. 11154). Cambridge, MA: National Bureau for Economic Research.

Hanushek, E. A., & Rivkin, S. G. (2010). Generalizations about using value-added measures of teacher quality. *The American Economic Review*, *100*, 267–271.

Heinrich, C. J. & Nisar, H. (2013). The efficacy of private sector providers in improving public educational outcomes. *American Educational Research Journal*, *50*(5), 856–894.

Henry, G. T., Bastian, K. C., & Fortner, C. K. (2011). Stayers and leavers: Early-career teacher effectiveness and attrition. *Educational Researcher*, *40*, 271–280. doi:10.3102/0013189X11419042

Henry, G. T., Fortner, C. K., & Thompson, C. L. (2010). Targeted Funding for Educationally Disadvantaged Students. *Educational Evaluation and Policy Analysis*, *32*, 183–204. doi:10.3102/0162373710370620

Hough, H. J. (2012). *Salary incentives and teacher quality: The effect of a district-level salary increase on teacher retention*. Palo Alto, CA: Center for Education Policy Analysis.

Iatarola, P., & Stiefel, L. (2003). Intradistrict equity of public education resources and performance. *Economics of Education Review*, *22*, 69–78.

Imbens, G., & Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies*, *79*, 933–959.

Imbens, G., & Lemieux, T. (2008). Regression discontinuity design: A guide to practice. *Journal of Econometrics*, *142*, 615–635.

Jackson, C. K., & Bruegmann, E. (2009). *Teaching students and teaching each other: The importance of peer learning for teachers* (NBER Working Paper No. 15202). Cambridge, MA: National Bureau of Economic Research.

Jacob, B. A., & Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *The Review of Economics and Statistics*, *86*, 226–244.

Johnson, L. L. (2014, February 9). Tennessee teachers push back on evaluation process. *The Washington Times*. Retrieved from http://www.washingtontimes.com/news/2014/feb/9/tennessee-teachers-push-back-on-evaluation-process/

Kane, T., Rockoff, J., & Staiger, D. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review*, *27*, 615–631.

Kirshstein, R. J., Berger, A. R., Benatar, E., & Rhodes, D. (2004). *Workforce contingent financial aid: How states link financial aid to employment*. Indianapolis, IN: Lumina Foundation.

Kolbe, T., & Strunk, K. O. (2012). Economic incentives as a strategy for responding to teacher staffing problems: A typology of policies and practices. *Educational Administration Quarterly*, *48*, 779–813.

Kraft, M. A., & Papay, J. P. (2014). Can professional environments in schools promote teacher development? Explaining heterogeneity in returns to teaching experience. *Educational Evaluation and Policy Analysis*, *36*, 476–500.

Krueger, A. B. (1999). Experimental estimates of education production functions. *Quarterly Journal of Economics*, *114*, 497–532.

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher sorting and the plight of urban schools: A descriptive analysis. *Educational Evaluation and Policy Analysis*, *24*, 37–62. doi:10.3102/01623737024001037

Lee, D. S., & Lemieux, T. (2009, February). *Regression discontinuity design in economics* (NBER Working Paper No. 14723). Cambridge, MA: National Bureau for Economic Research.

Lee, D. S., & Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, *48*, 281–355.

Liu, E., Johnson, S. M., & Peske, H. G. (2004). New teachers and the Massachusetts signing bonus: The limits of inducements. *Educational Evaluation and Policy Analysis*, *26*, 217–236.

Marsh, J. A., Springer, M. G., McCaffrey, D. F., Yuan, K., Epstein, S., Koppich, J. E., . . . Peng, A. (2011). *A big apple for educators: New York City's experiment with schoolwide performance bonuses*. Santa Monica, CA: RAND Corporation.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, *142*, 698–714.

Peske, H. G., & Haycock, K. (2006). *Teaching inequality: How poor and minority students are shortchanged on teacher quality*. Washington, DC: Education Trust.

Prince, C. (2002). *The challenge of attracting good teachers and principals to struggling schools*. Arlington, VA: American Association of School Administrators.

Prince, C. (2003). *Higher pay in hard-to-staff schools: The case for financial incentives*. Lanham, MD: Scarecrow Press.

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

Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *American Economic Review*, *94*, 247–252. doi:10.1257/0002828041302244

Ronfeld, M., Loeb, S., & Wyckoff, J. (2012). How teacher turnover harms student achievement. *American Educational Research Journal*, *50*, 4–36.

Sanders, W., & Rivers, J. (1996). *Cumulative and residual effects of teachers on future student academic achievement*. Knoxville: Value-Added Research and Assessment Center, University of Tennessee.

Springer, M. G., Lewis, J. L., Ehlert, M. W., Podgursky, M. J., Crader, G. D., Taylor, L. L., . . . Stuit, D. A. (2010). *District Awards for Teacher Excellence (D.A.T.E.) Program: Final Evaluation Report*. Nashville, TN: National Center on Performance Incentives.

Springer, M. G., Lewis, J. L., Podgursky, Ehlert, M. W., Taylor, L. L., Lopez, O. S., & Peng, X. (2008). *Governor's Educator Excellence Grant (GEEG) Program* (Final Evaluation Report). Nashville, TN: National Center on Performance Incentives.

Springer, M. G., Lewis, J. L., Podgursky, M., Ehlert, M. W., Gronberg, T. J., Hamilton, L., . . . Peng, X. (2009). *Texas Educator Excellence Grant (TEEG) Program: Final Evaluation Report*. Nashville, TN: National Center on Performance Incentives.

Springer, M. G., Pepper, M., and Ghosh-Dastidar, B. (2014). Supplemental educational services and student test scores: evidence from a large, urban school district. *Journal of Education Finance*, *39*(4), 370–403.

Steele, J. L., Murnane, R. J., & Willett, J. B. (2009). *Do financial incentives help low-performing schools attract and keep academically talented teachers? Evidence from California* (NBER Working Paper No. 14780). Cambridge, MA.

Synar, E., & Maiden, J. (2012). A comprehensive model for estimating the financial impact of teacher turnover. *Journal of Education Finance*, *38*, 130–144.

Tennessee Department of Education. (2012). *Tennessee ESEA flexibility*. Nashville, TN: Author.

Authors

MATTHEW G. SPRINGER, PhD, is an assistant professor of public policy and education at Peabody College of Vanderbilt University and the director of the National Center on Performance Incentives. His research focuses on incentives, accountability, and compensation.

WALKER A. SWAIN is a PhD student at Vanderbilt University's Peabody College of Education and Human Development. His research focuses on the impacts of education and social policy on traditionally disadvantaged populations.

LUIS A. RODRIGUEZ is a PhD student at Vanderbilt University's Peabody College of Education and Human Development. His research focuses on program evaluation, teacher labor markets, and policies intended to improve access to effective teachers.

Manuscript received June 2, 2014
First revision received April 8, 2015
Second revision received June 26, 2015
Third revision received September 9, 2015
Accepted September 10, 2015