TENNESSEE CONSORTIUM ON RESEARCH, EVALUATION & DEVELOPMENT

WORKING PAPER

Effective Teacher Retention Bonuses: Evidence from Tennessee

JUNE 12, 2014

MATTTHEW G. SPRINGER
LUIS A. RODRIGUEZ
WALKER A. SWAIN

VANDERBILT PEABODY COLLEGE

The Tennessee Consortium on Research, Evaluation, and Development is a research and policy center at Peabody College of Vanderbilt University, which is funded through the State of Tennessee's Race to the Top grant.

Please visit www.tnconsortium.org to learn more about our program of research and recent publications.

* 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). We appreciate helpful comments and suggestions from Dale Ballou, Matthew Craft, and Brent Evans and seminar participants at the annual meeting of the Association for Education Finance and Policy. We would also like to acknowledge the many individuals at the Tennessee Consortium and the Tennessee Department of Education for providing data and expert insight to conduct our analyses, in particular, Susan Burns, Sara Heyburn, Bing Howell, Trish Kelly, Erin O'Hara, Art Peng, Matthew Pepper, and Nate Schwartz. Any errors remain the sole responsibility of the authors. The views expressed in this paper do not necessarily reflect those of sponsoring agencies or individuals acknowledged.

Direct all correspondence to Matthew G. Springer, Vanderbilt University, Peabody Box #43, 230 Appleton Place, Nashville, TN 37203-5721 (matthew.g.springer@vanderbilt.edu).

Tennessee Consortium
on Research, Evaluation, and Development

Peabody #44 • 230 Appleton Place • Nashville, Tennessee 37203

PHONE: 615-322-5538 FAX: 615-322-6018

www.tnconsortium.org

TENNESSEE CONSORTIUM ON RESEARCH, EVALUATION & DEVELOPMENT

VANDERBILT PEABODY COLLEGE

0. Abstract

We report findings from a quasi-experimental evaluation of the recently implemented $5000 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. We exploit 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 bonuses are consistently positive across all specifications, and for teachers of tested subjects the program appears to have an effect that is generally both statistically and substantively significant. 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.

1. 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.[1] One of the most consistent findings in recent education research is that differences in teacher quality result in substantially different outcomes for students in school and beyond. 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 poverty or racial minorities, are considerably less likely to be taught by teachers with strong credentials or high estimated effectiveness (e.g., Clotfelter, Ladd, and Vigdor 2011). Compounding the initial maldistribution, highly effective teachers leave disadvantaged students and schools at higher rates than their less effective counterparts (who also turnover rapidly). 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 traditionally disadvantaged schools.

As with any labor market, teachers make decisions about where to work or stop working based on some combination of salary differentials and non-pecuniary benefits (i.e. the satisfaction of working with a particular group of students and colleagues) or costs (i.e. the challenges of working in a high stakes, high needs, environment with limited parental support). However, teacher labor markets are distinct in at least two important ways that might fuel inequitable distributions of effective teachers. Salaries are increasingly low relative to what candidates might earn in other sectors of the economy (Hanushek and Rivkin 2006), and perhaps more importantly salaries, at least within districts, are flat (Podgursky and Springer 2011). The combination of these two factors magnifies the role of perceived working conditions as a driver of teacher mobility decisions, especially for effective teachers who may have more options available.

---

[1] Some teachers are consistently better at raising the achievement of their students than others. Value-added studies of teacher effectiveness consistently find large variation in teacher classroom performance (e.g., Aaronson, Barrow, and Sander, 2007; Hanushek et al., 2005; Kane, Rockoff, and Staiger, 2008; Sanders and Rivers, 1996; Rockoff, 2004). Top-performing teachers, as defined by those teachers at the 95th percentile, produce three times the achievement growth in students when compared to low-performing teachers (Hanushek, 2003). Hanushek and Rivkin (2004) reported that the achievement gap among high- and low-socioeconomic status students could be overcome if an economically disadvantaged student encountered an above average teacher for five consecutive years.

Public schools differ dramatically with respect to perceived working conditions, or attractiveness as places to teach, with schools that have higher concentrations of low-income, non-white, and low-performing students generally perceived as less desirable places to work (Murnane et al., 1991; Hanushek, Kain, and Rivkin, 1999; Loeb and Page, 2000; Boyd et al, 2005; Scafidi et al, 2007; Kukla-Acevedo, 2009; Loeb, Darling-Hammond, & Luczak, 2005). More experienced teachers can typically use seniority-based transfer provisions in collective bargaining agreements to choose where to teach, and they can be expected to use it to exit these less desirable placements.[2] Similarly teachers with stronger credentials are more likely to have quality employment opportunities outside of teaching, and teachers with track records of effectiveness are likely to attract job offers from schools in other districts.[3]

For the reasons outlined above, 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 (Lankford, Loeb, and Wyckoff, 2002; Iaterola and Steifel, 2003; Roza et al., 2007 Clotfelter, Ladd, and Vigdor 2011; Peske and 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, and Sander, 2007; Rivkin, Hanushek, and Kane, 2005; Henry et al. 2011). The net result is that children enrolled in schools with high concentrations of disadvantaged students have greater exposure to less qualified instructors.

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). When pay is equalized, but students and resources are segregated by ability, income, and race, teacher quality is disequalized across schools within a school district. In order to equalize teacher quality, federal and state policymakers have demonstrated a heightened interest in designing practical

[2] In a case study of five large, urban school systems, Levin, Mulhern, and Schunck (2005) reported that 40 percent of teaching vacancies were filled by incumbent teachers and that school administrators had very little or no voice in the hiring decision.

[3] A number of studies have found a negative relationship between academic ability and the likelihood of entering teaching (Manski 1985; Hanushek and Pace 1995; Vegas, Murnane, and Willett 2001; Podgursky, Monroe, and Watson, 2010). Murnane et al (1991) and Bacolod (2003) suggest a decline in the quality of teacher labor supply over time. Corcoran, Evans, and Schwab (2010) attribute the decline in teacher quality to labor market opportunities for women outside of the education sector, which Hoxby and Leigh (2003) attribute largely to compression of teaching wages.

incentives to help offset differences in non-wage job characteristics (Prince, 2002; Prince, 2003; Johnson, 2005; Kirshtein, Berger, Benatar, and Rhodes, 2004), including signing bonuses, certification stipends, tuition reimbursement, loan forgiveness, tax credits, and housing subsidies.[4]

The theory behind retention incentives assumes that offering financial incentives will help retain more teachers in the upper tail of the ability distribution in hard-to-staff or chronically low-performing schools. A common argument of proponents of financial incentives is 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, typically located in rural or densely urban areas. Unfortunately, there is little research on the compensating differential needed to offset differences in nonpecuniary workplace characteristics. As such, retention bonuses in practice have ranged from $250 to more than $20,000.[5]

In this paper, we report findings from a quasi-experimental evaluation of the recently implemented retention bonus program for teachers in Tennessee's priority schools. We estimate the impact of the program on teacher retention using a fuzzy regression discontinuity design. We exploit a discontinuity in the probability of treatment conditional on the composite teacher effectiveness rating that assigns bonus eligibility. Our primary research questions include:

- What types of priority schools participated and did not participate in Tennessee's teacher retention bonus program?
- To what extent does a $5,000 retention bonus impact retention of high-performing teachers in priority schools that elected to participate in the program?
- Does a $5,000 retention bonus impact subgroups of teachers differently?

Point estimates for the main effect of the bonuses are consistently positive across all specifications, and for teachers of tested subjects the program appears to have an effect that is generally both statistically and substantively significant. 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.

In the sections that follow, we provide a description of Tennessee's teacher retention bonus program and then offer a brief review of the empirical literature. We next describe the data and

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

[5] Some researchers suggest that bonuses should be in the range of $20,000, while a simulation study by Feng (2009) estimated that teachers in hard-to-staff schools would need to be paid an additional $10,000 a year to be retained at the same rate as those in average schools.

study sample. We also discuss our analytic methods. Finally we present the results and conclude with implications for policymakers, practitioners, and future research.

2. 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 (Tennessee Department of Education, 2007). During the 2011-12 school year, approximately 17 percent of teachers leave their schools every year in Tennessee and the attrition rate for the state's most-effective teachers is around 7 percent. The attrition rate for highly effective educators increases to 10 percent when focused on urban districts in the state and 23 percent when focused on the bottom 5 percent of schools in the state (more than 3 times greater than the statewide attrition rate of highly-effective teachers), who are the focus of this study.

In the spring of 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 teacher retention bonus program for priority schools. Under the program, all priority schools were eligible to participate by applying to offer $5,000 retention bonuses to any level 5 teacher who was teaching in a priority school.[6] For many of the teachers in TN priority schools, a $5000 bonus constitutes approximately a 10 percent salary increase, or the equivalent of teacher with a master 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 2013-14 school year at a priority school in order 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 pro-rated amount of the retention bonus based on the number of days per week that he or she is actually working in a priority schools.

Both the priority school designation and the systems of teacher evaluation are major components of the retention bonus program. Below we describe the priority school designation as defined by Tennessee's school accountability program. We then discuss the new teacher evaluation

---

[6] 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 $7,000 was offered to every new level 5 teacher that transferred from a non-priority school into a priority school during the 2013-14 school year. Only 59 teachers received the signing bonus, therefore, this aspect of the program is not considered in this evaluation due to small sample size.

systems that were implemented as part of Tennessee's federally-funded Race to the Top grant program.

2.1. *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). Additionally, 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 five (5) percent of schools in the state are assigned priority status. The composite proficiency rate used to determine a schools 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, Biology I, and graduation rate are for high schools:

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

The success rates used for determining priority schools include up to three years of data. Success rates are calculated for schools with at least 2011-12 school year data. Only schools that are active in the 2011-12 school year, with at least 2010-12 and 2011-12 data, were identified as priority schools.

2.2. *Teacher evaluation*

In January 2010, the Tennessee Generally Assembly passed Senate Bill 5, also known as the First to the Top Act, reforming dozens of areas of state education policy. As part of the federal Race to the Top competition, the ambitious reforms helped Tennessee win a $501 million award 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.[7]

As of July 2011, the Tennessee State Board of Education approved four teacher evaluation models – the Tennessee Educator Acceleration Model (TEAM); Project Coach; Teacher

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

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 (see Appendix A). As of the 2012-13 school year, more than 80 percent of teachers across Tennessee use TEAM as their evaluation model, while TEM is the second most frequently used (11 percent), followed by Project COACH (5 percent) and TIGER (2 percent) (Ehlert et al, 2013).

Tested subject area teachers are designated as a level 5 educator if their three year composite TVAAS score and their overall teacher observation rating place them in the highest performance category. Teachers of non-tested subject areas must be in the highest performance category on the overall teacher observation rating as well as the school-wide achievement-based performance category. In the 2012-13 school year, the percentage distribution of teachers by performance category was relatively similar among tested subject area and non-tested subject area teachers, though the concentration of lower-performing teachers is higher in priority school settings.

3. Brief Review of Relevant Literature

It is well documented that high quality teachers are one of if not the most important school-based component in the production of student achievement (D. J. Boyd, Grossman, Lankford, Loeb, & Wyckoff, 2009; Rockoff, 2004; etc.). For that reason, a strong body of research has sought to better understand what makes highly qualified or effective teachers decide to stay or leave a school, or exit the profession altogether (i.e. Clotfelter, Ladd, & Vigdor, 2011; Scafidi, Sjoquist, & Stinebrickner, 2007; Feng 2010; Boyd, Lankford, Loeb, & Wyckoff, 2008; Feng, Figlio, & Sass, 2010). We briefly discuss findings from recent teacher retention bonus research.

3.1. Teacher Retention Bonuses

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.[8] The nature,[9] size, and context of

[8] Our review of the literature is focused on studies that investigate the effect of incentives on teacher retention. A number of studies have investigated the effect of signing bonuses on teacher recruitment, including Glazerman et al (2013), Liu, Johnson, Peske (2004), and Fowler (2003).

[9] Another common form of recruitment or retention financial incentive is the use of scholarship programs for teachers that condition receipt of payment on teachers serving in disadvantaged schools for a specified period of time (Johnson 2005). Steele, Murnane, and Willett (2010) evaluation of California's Governor's Teaching Fellowship (GTF) program, which offered $20,000 conditional scholarship ($5,000 per year over 4 years) to attract and retain academically talented, newly licensed

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. Hough 2012; Dee & Wyckff 2013; Balch & Springer 2014). Hough (2012), for example, assesses the effect of a salary increase on teacher retention in the San Francisco Unified School District as part of the Quality Teacher and Education Act of 2008 (QTEA). QTEA introduced an overall salary increase ranging between $500 and $6,300 based on placement on the salary schedule, a $2,000 bonus for teaching in a hard-to-staff school, and a $2,500 retention bonus after the 4th year of the program and $3,000 after the 8th year. Hough finds that the QTEA salary increase did not affect retention of targeted teachers, though overall teacher retention rates increased following the implementation of the program, to which the author attributes the null finding.

Similarly, in their evaluation of a pilot supplemental funding program to a group of educationally disadvantaged schools in North Carolina, Henry and colleagues (2010) found that approximately half of the money went towards 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, while turnover rates increased at non-supplement schools. However, while 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 and colleagues' (2008) were able to directly examine a $1,800 annual teacher retention bonus offered in North Carolina 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 percent, 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 TN retention bonuses both in its smaller magnitude ($1,800 vs. $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).

---

teachers to low-performing schools. The program had significant effect on teacher recruitment but did not differentially affect teacher retention among GTF recipients and non-recipients.

Two recent studies have sought to estimate effects of retention bonuses offered based on measures of teacher effectiveness. First, Glazerman and colleagues (2013) evaluated a substantial monetary incentive offered through the Talent Transfer Initiative (TTI) across 10 school districts in seven states that was designed to recruit and retain high performing teachers in low performing schools. Using a random assignment scheme, teachers who demonstrated a sufficient level of value added effectiveness (roughly the top 20 percent for their subject and grade) were eligible for a $20,000 bonus – paid in installments over a two-year period – if they transferred into and remained in schools that had low average test scores. Results showed that the transfer incentive had substantial positive impact on teacher-retention rates during the payout period; retention rates were significantly higher for high performing teachers compared to their counterparts – 93 versus 70 percent. However, not surprisingly, the difference was no longer statistically significant after the payments stopped.

Finally, while their analysis of salary bonuses in IMPACT, a high-stakes teacher evaluation system implemented in 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, Dee and Wyckoff (2013) implement a rigorous set of analyses similar to those described in this paper to offer important evidence with respect to the impact of bonuses. Using a regression-discontinuity design, the authors compared teachers near the IMPACT score threshold that separated "Effective" from "Highly Effective" teachers. Like the evaluation program in TN, 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 qualified for a large one-time bonus (up to $25,000) after being rated "Highly Effective" for one year and a sizable and permanent base salary increase (as large as $27,000 per year) upon achieving "Highly Effective" status in a second consecutive year. The authors utilize the sharp discontinuity between teachers eligible for the base pay increase and those who are not to estimate the local average treatment effect of a significant long-term salary increase on performance and retention of high performing teachers. While the 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 percent).

In sum, there is substantial evidence that difficulties with retaining effective teachers at low-performing schools and schools with high concentrations of low-income and minority students contribute to an inequitable distribution of teacher quality. Descriptive analyses of reasons for teachers exit indicate that, while non-pecuniary factors play a large roll in effective teachers' decision to exit, salary is also a factor. However, though generally positive, evaluations of monetary interventions designed to increase retention have failed to reach consensus as to the appropriate size and scope of a bonus program to that will encourage effective teachers to stay.

4. Data and Sample

4.1. Data Sources

This study utilizes administrative data obtained from the Tennessee Department of Education (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 2011-12 through 2013-14 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 two stand-alone datasets 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 datasets 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 Tennessee Value-Added Assessment System (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 2008-09 school year as the reference year. Our analyses use data on teacher composite scores, which statistically combines subject, grade, and year TVAAS estimates.[10] 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 serves as the third and final component of teachers' final evaluation rating.[11]

Our school-level information comes from multiple sources, including state school accountability reports, National Center for Education Statistics' Common Core of Data, and aggregating individual student- and teacher-level information at the school level. These school files contain the typically-used information such as level of schooling, school size, 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.

4.2. Sample

The sample for this study includes all priority schools in the State of Tennessee and teachers that worked in these schools during the 2012-13 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 required that the school principal and district superintendent sign and submit a letter of commitment to the State of Tennessee that affirmed their agreement to all of the terms and conditions of the retention bonus program.[12] 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.

As displayed in Figure 1, there were 82 priority schools during the 2012-13 school year that qualified to participate in the program. Of those 82 schools, 56 of them, employing about 2,000 teachers, elected to participate. Figure 1 further delineates teachers in the 56 schools that volunteered to participate by a teacher's eligibility status for a $5,000 retention bonus. Approximately 28 percent of the sample, or 562 teachers, did not have sufficient classroom observation data for program participation. A total of 964 teachers, or nearly half of the teacher

---

[10] For more information on TVAAS, see http://www.tn.gov/education/TVAAS.shtml.

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

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

sample, were not eligible for the bonus because they received a level 4 or lower overall performance evaluation rating, though 9 of the 713 teachers receiving below a level 5 rating that returned to a priority school the following year were still given a $5,000 retention bonus. Of the 473 priority school teachers that earned a level 5 rating for the 2012-13 school year, 80 percent (377 teachers) were retained, of which 321 or 85 percent received a $5,000 bonus.[13]

[Insert Figure 1 Here]

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

[Insert Table 1 Here]

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, though participating campuses have slightly fewer white students (1.98 vs. 4.16 percent). Participating campuses also have modestly greater percentage of student qualifying for free and reduced price lunch programs (90.46 vs. 86.42 percent) and female teachers (79.41 vs. 72.28 percent).

[Insert Table 2 Here]

5. Analytic Strategy

Our primary research questions include:

- What types of priority schools participated and did not participate in Tennessee's teacher retention bonus program?
- Does a $5,000 retention bonus impact retention of high-performing teachers in priority schools that elected to participate in the program?
- Does a $5,000 retention bonus impact subgroups of teachers differently?

5.1. Relationship between school characteristics and program participation

[13] 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 information on these cases.

To explore the relationship between observable school level characteristics and school participation in the retention bonus program, we estimate a series of probit models. The probit analysis estimates the probability that a school participates in the retention bonus program. Specifically, these models take the following form:

$$P = \beta' \times X + e$$

where the dependent variable is P, a binary variable indicating whether a school elected to participate in the retention bonus program; X is a vector explanatory variables and e is the error term. The regression sample is made up of the 82 priority schools, 56 of which agreed to participate in the retention bonus program.

We incorporate several school and teacher characteristics into our analysis of the determinants of a school's decision to participate in the retention bonus program. The school determinants are the share of economically disadvantaged students, school type (elementary, middle, and secondary), and school size. The teacher determinants are the average years of teacher experience and the share of teachers who are male.

We control for the share of economically disadvantaged students because within-school variations in student characteristics can make it more difficult to measure the effectiveness of individual teachers, so schools in which the student body is more diverse may be less likely to participate in the program. Although all schools are priority schools, the share of economically disadvantaged students ranges from 74.8 to 100 percent, and it is important to note that lower percentage of economically disadvantaged students indicates greater diversity. We also control for school type given recent surveys suggest that elementary school teachers are less supportive of teacher incentive programs than are secondary school teachers (Jacob and Springer, 2008).

We control for school size because studies suggest that small groups are more likely than large groups to adopt egalitarian incentive structures (Encinsoa, Gaynor, and Rebitzer, 2007). If this is the case, the median teacher might not want their school to participate in the program if he or she had full information about the abilities of other teachers (as would be likely in a small school) and if there were significant variation in those abilities (Freeman and Gelber, 2010).

We include both the share of teachers who are male and the average years of teaching experience because the literature suggests that perspectives on incentive pay programs vary by gender and experience (Niederle and Vesterlund, 2007; Ballou and Podgursky, 1993; Goldhaber, DeArmond, and DeBurgomaster, 2010; Ekel and Grossman, 2002). Additionally, several studies on teacher attitudes toward incentive pay policies conclude that beginning teachers are more accepting of performance pay than are more experienced, veteran teachers (Ballou and Podgursky, 1993; Goldhaber, DeArmond, and DeBurgomaster, 2010; Jacob and Springer, 2007). Finally, we include the share of level 5 teachers in a school at time t-1. One might assume that a greater share of level 5 teachers would make a campus more likely to participate as high-performing teachers are most likely to benefit from the program.

5.2. Impact of the retention bonus program

The theory behind retention incentives implies that the opportunity to earn an additional $5,000 in income for working in a priority school for an additional year will cause retention bonus eligible teachers to be more likely to remain in a priority school the following year, as the economic benefit is greater given standardized remuneration practices in the public education sector. We refer to the outcome of retention from the 2012-13 to 2013-14 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 2012-13 to 2013-14 school years.

Following the Neyman-Rubin causal model, there are two possible outcomes for each individual teacher i at the conclusion of the school year. The first, denoted by $Y_i(1)$, is the potential outcome for teacher i when he or she is eligible for a retention bonus. The other, denoted by $Y_i(0)$, is the potential outcome when the teacher is not eligible for a retention bonus. Thus, the treatment effect for observation i is defined by:

$$\tau_i = Y_{i1} - Y_{i0}$$

However, a fundamental problem of causal inference is that we can never observe $Y_i(1)$ and $Y_i(0)$ at the same time for a single teacher. This remains true regardless of the methodology used to make an inference. Instead, we can only observe the outcome of interest after a teacher receives their overall performance score and they make a decision on whether to return to a priority school the following year. Thus, we are interested in estimating the average treatment effect of being eligible for a $5,000 bonus on retention, as defined by:

$$E[Y_i(1) - Y_i(0)]$$

In a sharp regression discontinuity 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.[14] 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 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 $5,000 retention bonus if they teach 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 they are 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 2, noncompliance is present, which makes a sharp regression discontinuity identification strategy invalid. Forty-four level 5 teachers, or approximately 11 percent of our level 5 teacher sample, did not receive a retention bonus even though they returned to a priority school during the 2013-14 school year (no-shows). Seven 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.

[Insert Figure 2 Here]

Thus, to estimate the impact of the bonus program on retention, we adopt a fuzzy regression discontinuity 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:

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

[14] This type of RD design has been shown to produce unbiased, valid estimates of program effects approximating a randomized experiment. See, for example, Cook, 2008; Imbens and Lemieux, 2008; Black, 1999; Angrist and Lavy, 1999; Hahn, Todd, and Van der Klaauw, 1999.

where $T = 1[X \geq c]$ indicates whether the assignment variable exceeds the eligibility threshold c and D is whether or not a teacher receives a retention bonus.[15] Since $D = \Pr(D = 1|X = x) + v$, 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 $\tau_r = \tau * \delta$. An obvious advantage of the reduced form approach is that it provides an estimate of the treatment effect and accompanying standard errors. As described in Imbens and Lemieux (2008), these estimates are identical to the ratio of the reduced form coefficients τ_r/δ, provided that the bandwidths used in the first and second stages are consistent (and that the same order polynomial is used for g (.) and f(.)).

If there is an underlying relationship between *X* and staying in a priority school, *Y*, this comparison will suffer from bias as long as our treatment group contains observations with *X* strictly above our cutoff value and the control group contains observations with *X* strictly below our cutoff value. By requiring that observations be closer to the cutoff value we can limit the amount of bias, but then the sample becomes smaller and we suffer a loss of precision, which is a palpable concern in the present context. Thus, we also estimate a local polynomial regression model where the functional form is cubic in the running variable.

We address the bandwidth selection problem using Imbens and Kalyanaraman's (2012) optimal bandwidth calculation; that is, how close should an observation be to *c* in order to be part of the analysis sample without sacrificing precision due to small sample size. We also estimate models using kernel-weighted local polynomial smoothing with triangular case weights. Triangular case weights are applied such that the impact on the estimate of cases farther from the cutoff declines at a linear rate.

We also estimate select specifications of our models with controls. It should not matter much whether controls are included, given our assumption that the relationship of Y to X is continuous and the supporting evidence that our treatment and control conditions are equivalent on

[15] As noted in Lee and Lemieux (2008), although the probability of treatment is modeled as a linear probability model, it does not impose any restrictions on the probability since 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.

observables. However, to the extent that the underlying relationship of Y to X is not linear, there may still be some bias remaining (our controls partly captures the effect of treatment and partly differences between the treatment and control groups that would have occurred anyway, for which we have imperfectly controlled). In this case, including covariates can reduce that bias and, in the usual way, their inclusion may improve precision.

5.3. Testing the validity of the RD identification strategy

There are three fundamental assumptions required by the RD identification strategy. First, unobserved characteristics vary continuously around the teacher effectiveness cutoff with observable characteristics used to determine bonus eligibility. Second, our forcing variable, X, has not been manipulated in order to affect who receives treatment. Third, there are no other programs or services with the same eligibility rule, which assures that the bonus program treatment is not confounded with some other treatment.

5.3.1. Equivalence between treatment and control conditions

We investigate whether there are baseline imbalances between treatment and control teachers by testing for differences on observable teacher characteristics using the full sample of teachers and the sample of teachers on either side of the cutoff. The optimal bandwidth calculation is defined by Imbens and Kalyanaraman (2012). As displayed in Table 3, we find a number of statistically significant differences when comparing the full sample of teachers (column 1) to teachers at and just to the right of c (column 4).[16] For example, a greater percentage of female and black teachers are rated level 5, while fewer white teachers and teachers with a bachelor's degree attain level 5 status. Imbalances are not unexpected given that our sample includes all priority school teachers.

[Insert Table 3 Here]

However, we are most interested in whether the observed baseline covariates are "locally" balanced on either side of the cutoff, which should be the case if the treatment indicator is locally randomized. As displayed in Table 3, across all comparisons and statistical tests, we reject the hypothesis that the means of the treatment and the control condition teachers are statistically different as indicated by the means reported in columns (3) and (4) and mean differences reported in column (7). These estimates are not sensitive to the inclusion of 44 teachers that did not receive a bonus even though they achieved level 5 rating.

5.3.2. Testing for manipulation

---

[16] In addition to simple mean comparisons using a Student's t-test, we used Hotelling's t-test, which is the analog to a t-test when multiple variables are considered simultaneously. We also ran a series of OLS and logit regressions with an indicator for treatment status.

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 individuals may try to influence their overall performance score rating in order to increase the likelihood of being eligible for a retention bonus. While this is unlikely given the timing of program implementation, we implement the formal procedure developed by McCrary (2007). Results from the McCrary sorting test, as displayed in Figures 3, suggest a slight jump in X at c; however the difference is not statistically significant at conventional levels. We found similar pattern for reduced samples (e.g., tested subject teachers only) as displayed in Figures 4 and 5 and, as a result, are reasonably confident that manipulation of the running variable did not occur.

[Insert Figures 3-5 Here]

As noted earlier the timing of program implementation further attests to this fact – the program was implemented relatively late in the school year and teachers did not necessarily have an opportunity to artificially inflate performance scores (particularly teacher observation ratings) in an effort to become eligible for a retention bonus.

5.3.3. Other programs or services

A final assumption of the RD identification strategy is that there are no other programs or services with same eligibility rules. This assures that T is not confounded with some other treatment. In this context, we are not aware of any other programs or services with the same eligibility rule that could confound treatment.

6. Results

6.1. Relationship between school characteristics and program participation

Table 4 displays estimates of the relationship between school characteristics and program participation. We find that the share of economically disadvantaged students in the school is a strong predictor of participation, though the estimate is not in the expected direction. Schools with a greater share of economically disadvantaged students, i.e., less within-school variation in student characteristics, were more likely to participate in the program. We also find evidence that middle schools were modestly less likely to participate in the retention bonus program when compared to high schools. Finally, it is clear that the share of level 5 teachers is a strong predictor of program participation. This is expected given that level 5 teachers are most likely to benefit from the program.

[Insert Table 4 Here]

6.2. Impact of the retention bonus program

Table 5 reports estimates of the impact of the bonus program on teacher retention. While we report estimates using a large number of different bandwidth selections, our discussion will focus on estimates from models where the sample is defined by the optimal bandwidth specification defined by Imbens and Kalyanaraman (2009). We do not find a statistically significant program effect on teacher retention when using a linear form, though the point estimates are in the hypothesized direction. When modeling a cubic form of the local polynomial regression, which is the empirically preferred form, the estimate is statistically significant and positive. Level 5 teachers that receive a retention bonus are 23 percent more likely to remain teaching in a priority school when compared to teachers just below the level 5 cutoff (see Figure 6).[17] This estimate is robust to a variety of bandwidth selections.

6.3. Heterogeneous treatment effects

We explore the impact of the bonus program on teacher retention separately for tested and non-tested subject teachers. A teacher is considered a tested subject teacher if they have valid data for the individual TVAAS component of the teacher evaluation system. A teacher is considered a non-tested subject teacher if a teacher is only rated on school value-added and classroom observations.

Table 6 displays estimates of the impact of the bonus program on tested subject teacher retention. Once again, our discussion focuses on the estimates when the analysis sample is defined by the optimal bandwidth specification. Estimates from the local linear regression are statistically significant, suggesting that a level 5 tested subject teacher is 24.3 percent more likely to remain teaching in a priority school when compared to tested subject teacher just below the level 5 cutoff. While the magnitude of the value on the treatment effect remains relatively similar when modeling a cubic form of the local polynomial regression, the lack of precision yields insignificant results. Additionally, when the full sample is included, the estimate is statistically significant and large, suggesting a very meaningful impact of the bonus on retention of highly-effective teachers.

Table 7 reports estimates of the impact of the bonus program on non-tested subject teachers. We find that the estimates are not statistically different from zero at conventional levels across all functional forms. Additionally, the magnitudes of the point estimates are quite small when compared to those reported for tested subject teachers or the sample as a whole. It appears tested subject teachers are driving the effect, which isn't unexpected given the amount of weight Tennessee's teacher evaluation system attributes to school level performance for untested subject teachers.

[17] Appendix A shows similar results when estimates are produced using kernel weights.

6.4. *Cost benefit analysis*

While the estimated effects of the retention bonus on teacher retention are modest in magnitude and sensitive to model specification, estimates of net program costs that account for benefits to students in the form of future earnings projections indicate the intervention would provide a net benefit even if its effects on retention were substantially overestimated. The robustness of the cost effectiveness of the policy is primarily a function of the strength of the intervention that retention of a Level-5 teacher in a Priority School represents. 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 teacher effectiveness of 1.7 standard deviations. To put this in perspective, Hanushek (2011) estimates the net present value of the increased earnings resulting from a teacher 1 standard deviation above the mean teaching 25 students for 1 year to be greater than $500,000. If we focus on marginal economic returns to students in discounted lifetime earnings (column 2 of Table 1), the program would only need to have influenced 5 Level-5 teachers to stay and teach another year to justify its total cost (roughly 2.1 million for 361 bonuses). Using Hanushek's most conservative estimate of the annual marginal earnings effects of a teacher quality increase (roughly $12,000 discounted lifetime earnings per student for a 1 standard deviation improvement), we find that the bonus program would pay for itself in long run tax revenue if only 10% of bonus recipients stayed because of the bonus and taught an average of 30 students for 1 year.

6.4.1. *Program Costs*

In total, the state distributed slightly over $2.1 million in $5,000 retention bonuses to 361 highly effective teachers, who agreed to stay at a priority school during the 2013-14 school year. After employer shares of taxes and administrative costs, the total cost to the state associated with each bonus paid was roughly $6000. 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.

While our fuzzy regression discontinuity design estimated treatment effects local to the cutoff score and not generalizable to Level-5 teachers who had particularly high scores, we attempt to account for uncertainty in the effect estimates by presenting estimates for a range of plausible program impacts. If 20 percent of teachers who received bonuses stayed at a Priority School as a result of the program, the cost per teacher retained would be roughly $30,000. If the teachers who were retained as a result of the bonus taught an average of 25 students the cost per effected student would be roughly $1,200. If 90 percent of the teachers who received bonuses would have stayed in the absence of the bonus, the cost per effected pupil would be $2,400 per effected student.

Compared with interventions, like NCLB's supplemental education services (Springer et al, 2014; Heinrich et al, 2013), summer school (Jacob and Lefgren 2004) or reduced class size (Krueger 1999), the costs per effected 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).

6.4.2. Program Benefits

The second broad component of the cost-benefit analysis accounts for potential benefits to students from Level-5 teacher retention. The effects of teacher retention on student achievement rely primarily on the value added of the retained Level-5 teachers relative to a range of potential replacement teachers (Table 8). These benefits are estimated in dollar terms using recent figures on the net present value of increased lifetime earnings associated with improved student achievement. Here we follow Hanushek's (2011) use of Chetty and colleagues' (2013)[18] estimates of long-term benefits of effective teachers, noting that our estimation of a teacher's relative effectiveness compared to the distribution of teachers in the state is based on a composite evaluation score rather than a single value-added measure.

The second column of Table 8 displays the present value of predicted increase in lifetime earnings for a class of 25 students taught by a retained teacher, relative to their predicted earnings if they were taught by likely replacements (average teacher in the state, average teacher in a priority school, or average new hire at a priority school). These student benefit estimates are derived from Hanushek's (2011) predicted annual marginal returns to a 1 standard deviation improvement in teacher effectiveness. The student income effects range from just roughly $550,000 to $900,000 per retained teacher, depending on the relative effectiveness of the teacher who would have replaced them had they left. We also estimate the potential for the long run tax revenue associated increased earnings to offset the government cost of implementing the program. The third column of Table 8 displays the present value of tax revenue generated by the income increase associated with each

[18] The initial version of Chetty and colleagues' 2013 study was subject to a considerable amount of critique, often focused on its discussion of the potential benefits to selective termination based on value-added scores (e.g. Ravitch 2012) and occasionally focused on concerns about the strength of the causal claims (Ballou 2012). Others highlighted the difficulty, which the authors note, of generalizing from data on teacher value-added measures that predated the introduction of high-stakes accountability systems (Winerup, 2012). However, to date, it remains the most methodologically rigorous attempt to quantify the monetary benefits of teacher quality.

teacher retained, using the current average sales tax rate in Tennessee of 9.45 percent. Estimated long-run revenue effects range from approximately $52,000 to $85,000 per teacher retained. The fifth column presents net benefit to the state budget per teacher retained less the net cost per teacher retained, which ranges from $26 to $60 if roughly 23 percent of bonus recipients would have otherwise left.

Table 9 displays the net cost to the government per $1 million spent on priority school retention bonuses for a range of potential impacts on retention and a range of predicted economic returns to achievement gains. Using Hanushek's preferred estimate of economic returns per standard deviation increase in teacher effectiveness, the bonus program provides a long-run net benefit to the state budget even if the program only results in 10 teachers staying per $1 million, if the retained teacher teach an average of 30 students for one additional year. With the lowest estimated economic returns to teacher effectiveness ($12,000 lifetime per teacher standard deviation), the smallest class size (20), and lowest effect level of impact on Level-5 teacher retention (10 teachers retained per $1 million spent), the state only loses $61,000 per $1 million spent. In other words, in the most conservative scenario, the state recuperates 94 percent of its expenses.

While these figures may seem optimistic, there are several reasons to consider these estimated benefits to be lower bounds. Recent studies have found significant effects on student achievement through peer effects on colleagues who work with highly effective teachers (Jackson and Bruegmann 2009), negative effects on students of teacher churn within and across schools (Ronfeldt, Loeb, and Wyckoff 2012), and a significant administrative financial burden associated with teacher turnover, including separation, hiring and training costs (Synar and Maiden 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. Of course, teachers who are retained in one year also have a higher likelihood of being retained in the following year than those who exit. 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. Thus, the estimates presented in Tables 8 and 9 represent a conservative lower bound for program net benefits.[19]

7. Conclusion

[19] While the calculated benefits rely heavily on long-term effects on student earnings, an alternative estimation strategy that disregards any student effects, focusing instead on reduced administrative costs associated with turnover also results in the bonuses representing a low cost intervention. Synar and Maiden (2012) estimate the costs associated with a teacher leaving (separation, hiring, and training) range from roughly $10-18 thousand dollars per teacher.

We report findings from a quasi-experimental evaluation of the recently implemented retention bonus program for teachers in Tennessee's priority schools. We estimate the impact of the program on teacher retention using a fuzzy regression discontinuity design. We exploit 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 bonuses are consistently positive across all specifications, and for teachers of tested subjects the program appears to have an effect that is generally both statistically and substantively significant.

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, less effective teachers, bonuses that retain the teachers at the higher end of the effectiveness distribution can have substantial impacts on the quality of a schools 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.

As is true for any policy that relies on observations and test-score-based value-added estimates to differentiate teachers, the benefits of retention bonus 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 in the case of teacher quality policies that rely on terminations.

While 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 April of 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. While the timing of implementation limits both the sample and predicted effect size, it potentially strengthens the internal validity of our regression discontinuity 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 non-recipients 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. Forthcoming interviews and surveys of administrators and teachers should help clarify whether the observed noncompliance was attributable to something other than rushed implementation.

In sum, despite these substantial implementation difficulties, and a relatively small sample of participating schools, we find some preliminary evidence of a causal link between the bonus offer and retention of high quality teachers. Estimates are particularly positive amongst teachers of tested subjects, 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 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.

0. References

Aaronson, D., Barrow, L., and Sanders, W. (2007). Teachers and student achievement in Chicago high schools. *Journal of Labor Economics*, 25(1), 95-136.

Angrist, J.D. and Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class sized on scholastic achievement. *Quarterly Journal of Economics*, 114(2), 533-575.

Ballou, D. (2012, February 16). Review of The Long-Term Impacts of Teachers: Teacher Value Added and Student Outcomes in Adulthood. National Education Policy Center. Retrieved from http://nepc.colorado.edu/thinktank/review-long-term-impacts

Balch, R. and Springer, M.G. (2014). An Evaluation of a Pay for Performance Program in Texas. Working Paper. Nashville, TN: Vanderbilt University.

Ballou, D. & Podgursky, M. (1993). Teachers' Attitudes Toward Merit Pay: Examining the Conventional Wisdom. *Industrial and Labor Relations Review*,*47*(1), 50-61.

Black, S. (1999). Do better schools matter? Parental valuation of elementary education. *Quarterly Journal of Economics*, 33(1), 62-87.

Bocolod, M.P. (2007). Do alternative opportunities matter? The role of female labor markets in the decline of teacher quality, 1960-2000. *The Review of Economics and Statistics*, 89(4), 737-751.

Boyd, D, Lankford, H., Loeb, S., & Wyckoff, J. (2008). The Impact of Assessment and Accountability on Teacher Recruitment and Retention: Are There Unintended Consequences? *Public Finance Review*, *36*(1), 88–111. doi:10.1177/1091142106293446

Boyd, D. J., Grossman, P. L., Lankford, H., Loeb, S., & Wyckoff, J. (2009). Teacher Preparation and Student Achievement. *Educational Evaluation & Policy Analysis*, *31*(4), 416–440. doi:10.3102/0162373709353129

Boyd, Donald, Grossman, P., Lankford, H., Loeb, S., & Wyckoff, J. (2008). Who Leaves? Teacher Attrition and Student Achievement. *National Bureau of Economic Research Working Paper Series*, *No. 14022*(september 2008), 1–39. Retrieved from http://www.nber.org/papers/w14022.

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(2), 636-654.

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*(3), 399–438. Retrieved from http://www.mitpressjournals.org/efp

Clotfelter, C. T., Ladd, H. F., Vigdor, J. L., & Diaz, R. A. (2004). Do school accountability systems make it more difficult for low-performing schools to attract and retain high-quality teachers? *Journal of Policy Analysis and Management*, *23*(2), 251–271. doi:10.1002/pam.20003

Corcoran, S.P., Evans, W.N., and Schwab, R.M. (2004). Women, the labor market, and the declining relative quality of teachers. *Journal of Policy Analysis and Management*, 23(3), 449-470.

Dee, T. and Wyckoff, J. (2013). Incentives, selection, and teacher performance: Evidence from IMPACT. NBER Working Paper 19529. Cambridge, MA.

Eckel, C. & Grossman, P. (2002). Sex Differences and Statistical Stereotyping in Attitudes Toward Financial Risk. *Evolution and Human Behavior,* 23(4), 281-295.

Encinosa, W. E., I.I.I., Gaynor, M., & Rebitzer, J. B. (2007). The sociology of groups and the economics of incentives: Theory and evidence on compensation systems. *Journal of Economic Behavior and Organization, 62*(2), 187-214.

Feng, L. (2010). Hire Today, Gone Tomorrow: New Teacher Classroom Assignments and Teacher Mobility. *Education Finance and Policy*, *5*(3), 1–39. doi:10.1162/EDFP_a_00002

Feng, L., Figlio, D. N., & Sass, T. (2010). School accountability and teacher mobility. *National Bureau of Economic Research Working Paper Series*. Working Paper 47, CALDER. Retrieved from http://www.nber.org/papers/w16070.

Freeman, R. B., & Gelber, A. M. (2006). Optimal Inequality/Optimal incentives: Evidence from a tournament. National Bureau of Economic Research, Inc, NBER Working Papers). *NBER Working Paper Series, w12588*.

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

Goldhaber, D., De Armond, M., & DeBurgomaster, S. (2007). Teacher attitudes about compensation reform: Implications for reform implementation. SFRP Working Paper 20.

Greenberg, D., & McCall, J. (1974). Teacher Mobility and Allocation. *The Journal of Human Resources*, *9*(4), 480–502. Retrieved from http://links.jstor.org/sici?sici=0022-166X%28197423%299%3A4%3C480%3ATMAA%3E2.0.CO%3B2-A.

Han, J., Todd, P., and Van der Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica,* 69(1), 201-209.

Hanushek, E.A., and Rivkin, S. (2004). How to improve the supply of high-quality teachers. Brookings Papers on Education Policy, 7-25. Brookings Institution, Washington, DC.

Hanushek, E., Kain, J., O'Brien, D., and Rivkin, S. (2005). The market for teacher quality. NBER Working Paper 11154. National Bureau for Economic Research, Cambridge, MA.

Hanushek, E.A. and Pace, R.R. (1995). Who chooses to teach (and why)? *Economics of Education Review*, 30(3), 466-479.

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

Henry, G T, Bastian, K. C., & Fortner, C. K. (2011). Stayers and Leavers: Early-Career Teacher Effectiveness and Attrition. *Educational Researcher*, *40*(6), 271–280. doi:10.3102/0013189X11419042

Henry, Gary T, Fortner, C. K., & Thompson, C. L. (2010). Targeted Funding for Educationally Disadvantaged Students. *Educational Evaluation & Policy Analysis*, *32*(2), 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. Center for Education Policy Analysis. Palo Alto, CA.

Hoxby, C.M. and Leigh, A. (2004). Pulled away or pushed out? Explaining the decline in teacher aptitude in the United States. *American Economic Review*, 94(2), 236-240.

Iaterola, P. and Stiefel, L. (2003). Intradistrict equity of public education resources and performance. *Economics of education review*, 22, 69-78.

Imbens, G. and Lemieux (2008). Regression discontinuity design: A guide to practice. *Journal of Econometrics*, 142(2), 615-635.

Imbens, G. and Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies*, 1-27.

Jacob, B. and Springer, M.G. (2008). Teacher attitudes on pay for performance: A pilot study. National Center on Performance Incentives Working Paper. Nashville, TN.

Johnson, L.L. (2014). Tennessee teachers push back on evaluation process. The Washington Times. Sunday, February 9, 2014.

Kane, T., Rockoff, J., and Staiger, D. (2008). What does certification tell us about teacher effectiveness? Evidence from New York City. *Economics of Education Review*, 27(6), 615-631.

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

Kolbe, T. and Strunk, K.O. (2012). Economic incentives as a strategy for responding to teacher staffing problems: A typology of policies and practices. *Educational administration quarterly*, 1-35.

Kukla-Acevedo, S. (2009). Leavers, Movers, and Stayers: The Role of Workplace Conditions in Teacher Mobility Decisions. *The Journal of Educational Research*, *102*(6), 443–452. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=qeh&AN=BEDI09118563&site=ehost-live&scope=site

Lankford, H., Loeb, S., & Wyckoff, J. (2002). Teacher Sorting and the Plight of Urban Schools: A Descriptive Analysis. *Educational Evaluation & Policy Analysis*, *24*(1), 37–62. doi:10.3102/01623737024001037.

Lee, D.S. and Lemieux, T. (2009). Regression discontinuity design in economics. NBER working paper 14723. Cambridge, MA.

Loeb, S., Darling-Hammond, L., & Luczak, J. (2005). How Teaching Conditions Predict Teacher Turnover in California Schools. *Peabody Journal of Education*, *80*(3), 44–70. doi:10.1207/s15327930pje8003_4

Manski, C.F. (1985). Academic ability, earnings, and the decision to become a teacher: Evidence from the national longitudinal study of high school class of 1972, Working Paper #1539. Cambridge, MA: National Bureau of Economic Research.

Murnane, R.J., Singer, J.D., Willett, J.B., Kemple, J.J., Olson, R.J. (1991). Who will teach? Cambridge, MA: Harvard University Press.

Murnane, R. J. (1981). Teacher Mobility Revisited. *The Journal of Human Resources*, *16*(1), 3–19. Retrieved from http://www.jstor.org/stable/145216.

Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics, 122*(3), 1067-1101.

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

Podgursky, M., Monroe, R., and Watson, D. (2010). Teacher mobility, pay, and academic quality. *Economics of Education Review.*

Podgursky, M. and Springer, M.G. (2011). Teacher compensation systems in the United States K-12 public school system. *National Tax Journal*, 61(4).

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

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

Rockoff, J. E. (2004). The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *American Economic Review*, *94*(2), 247–252. doi:10.1257/0002828041302244.

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

Scafidi, B., Sjoquist, D. L., & Stinebrickner, T. R. (2007). Race, poverty, and teacher mobility. *Economics of Education Review*, *26*(2), 145–159. doi:10.1016/j.econedurev.2005.08.006.

Steele, J.L., Murnane, R.J., and Willett, J.B. (2009). Do financial incentives help low-performing schools attract and keep academically talented teachers? Evidence from California. NBER Working Paper 14780. Cambridge, MA.

Vegas, E., Murnane, R.J., and Willett, J.B. (2001). From high school to teaching: Many steps, who makes it. *Teachers College Press*, 103(June), 427-449.

Winerip, M. (2012, January 15). Study on Teacher Value Uses Data From Before Teach-to-Test Era. *The New York Time,* A13.

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	528.74	529.74	525.77	-4.14
n	82	56	26	

Notes: * significant at the 10% level; ** 5% level; *** 1% level

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				
Percent Female	48.22%	48.51%	47.49%	1.03%
Percent White	2.60%	1.98%	4.16%	-2.18%*
Percent Black	93.23%	93.52%	92.50%	1.02%
Percent Asian	0.25%	0.23%	0.29%	-0.06%
Percent Hispanic	3.76%	4.10%	2.90%	1.20%
Percent Other	0.16%	0.16%	0.14%	0.02%
Percent Free and Reduced Price Lunch	89.38%	90.46%	86.42%	4.04%*
Percent Special Education	16.96%	16.51%	18.09%	-1.57%
Percent English Language Learners	2.26%	2.40%	1.92%	0.48%
Teacher Characteristics				
Percent Female	77.21%	79.41%	72.28%	7.13%**
Percent White	31.09%	30.14%	33.21%	-3.07%
Percent Black	68.10%	69.10%	65.86%	3.24%
Percent Asian	0.52%	0.54%	0.48%	0.06%
Percent Hispanic	0.26%	0.18%	0.43%	-0.25%
Percent Other	0.04%	0.04%	0.03%	0.01%
Percent with Bachelors (B.A.)	33.67%	34.56%	31.66%	2.90%
Percent with Masters (M.A.)	36.46%	36.07%	37.32%	-1.25%
Percent with More than Masters	17.14%	17.12%	17.17%	-0.05%
Percent with Specialist in Education (Ed.S.)	11.02%	10.60%	11.97%	-1.37%
Percent with Doctor of Philosophy (Ph.D.)	1.71%	1.64%	1.87%	-0.23%
Percent 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	52414.40	52811.91	51523.97	1287.95
n	81	56	26	

Notes: * significant at the 10% level; ** 5% level; *** 1% level

Table 3. Equivalence of teacher observables between treatment (c≥420) and control (c<420) groups

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Full Sample	Left of c	Just left of c	At and just right of c	Right of c	(2)-(5)	(3)-(4)
Gender							
Female	80.62%	78.18%	83.98%	84.86%	85.59%	-7.41%***	-0.88%
Race							
White	27.27%	29.37%	27.78%	23.32%	23.01%	6.36%**	4.46%
Black	71.88%	69.78%	71.67%	75.34%	76.13%	-6.35%**	-3.67%
Asian	0.43%	0.32%	0.56%	0.90%	0.65%	-0.33%	-0.34%
Hispanic	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Other	0.07%	0.11%	0.00%	0.00%	0.00%	0.11%	0.00%
Degree							
Bachelors (B.A.)	34.15%	36.56%	32.24%	30.53%	29.24%	7.33%***	1.71%
Masters (M.A.)	38.34%	37.08%	37.70%	42.92%	40.89%	-3.81%	-5.22%
More than Masters	16.69%	16.46%	19.67%	15.49%	17.16%	-0.70%	4.19%
Specialist in Education (Ed.S.)	9.57%	8.54%	8.74%	10.18%	11.65%	-3.11%*	-1.43%
Doctor of Philosophy (Ph.D.)	1.26%	1.35%	1.64%	0.88%	1.06%	0.29%	0.75%
Teaching Tested Grade-Subkect	39.96%	41.18%	29.51%	37.44%	37.42%	3.76%	-7.94%*
Years of Experience	10.95	10.79	10.59	11.46	11.28	-0.49	-0.87
Salary	54716.82	53953.90	55512.04	55745.89	56263.77	-2309.88***	-233.85
n	1437	964	183	227	473		

Notes: * significant at the 10% level; ** 5% level; *** 1% level

Table 4. Relationship between school characteristics and program participation

	Model 1
Share Econ. Disadvantaged	24.02** (11.65)
Average teacher experience	0.00 (0.21)
Share Male	-21.16 (17.16)
Share Level 5 Teacher	68.59*** (20.81)
School Size	0.00 (0.00)
Elementary	-5.73 (4.27)
Middle	-2.28* (1.36)
Percent Non-White Students	-4.79 (10.64)
Intercept	-11.72 (10.12)
N	82 (56 participants; 26 non-participant)
Adj. R-squared	0.81

Notes: Robust standard error in parentheses. * significant at the 10% level; ** 5% level; *** 1% level

Table 5. Impact of Retention Bonus

	Specification			
	Local Linear Regression		Local Polynomial Regression (Cubic)	
	n		*n*	
Optimal Bandwidth	*827*	0.098 (0.068)	*1,195*	0.230** (0.101)
75% of Optimal Bandwidth	*659*	0.087 (0.079)	*1,022*	0.233** (0.110)
110% of Optimal Bandwidth	*883*	0.081 (0.064)	*1,245*	0.217** (0.099)
125% of Optimal Bandwidth	*953*	0.058 (0.061)	*1,313*	0.158 (0.097)
Full Sample	*1,437*	0.057 (0.049)	*1,437*	0.141 (0.090)

Notes: Robust standard errors in parentheses. * significant at the 10% level; ** 5% level; *** 1% level

Table 6. Impact of Retention Bonus for Tested Subject Teachers

	Specification			
	Local Linear Regression		Local Polynomial Regression (Cubic)	
	n		*n*	
Optimal Bandwidth	*339*	0.243** (0.108)	*316*	.22131 (0.222)
75% of Optimal Bandwidth	*263*	0.310** (0.127)	*246*	.21177 (0.225)
110% of Optimal Bandwidth	*356*	0.211** (0.106)	*337*	.16241 (0.205)
125% of Optimal Bandwidth	*379*	0.232** (0.100)	*361*	0.254 (0.196)
Full Sample	*574*	0.225*** (0.074)	*574*	0.371** (0.145)

Notes: Robust standard errors in parentheses. * significant at the 10% level; ** 5% level; *** 1% level

Table 7. Impact of Retention Bonus for Untested Subject Teachers

	Specification			
	Local Linear Regression		Local Polynomial Regression (Cubic)	
	n		*n*	
Optimal Bandwidth	*537*	0.003 (0.085)	*598*	.09016 (0.153)
75% of Optimal Bandwidth	*431*	-0.002 (0.099)	*496*	.08107 (0.172)
110% of Optimal Bandwidth	*566*	-0.011 (0.080)	*614*	.1238 (0.148)
125% of Optimal Bandwidth	*607*	0.002 (0.075)	*646*	0.135 (0.143)
Full Sample	*863*	-0.037 (0.065)	*863*	0.076 (0.116)

Notes: Robust standard errors in parentheses. * significant at the 10% level; ** 5% level; *** 1% level

Table 8: Estimated Costs and Benefits of Retaining Level 5 Teachers Through Bonuses

(1) Costs per Teacher Retained	(2) Benefits to Students per Teacher (Future Earnings)	(3) Benefits to State per Teacher (Tax Revenues)	(4) Net Returns to Students (2-1)	(5) Net Returns to State* (3-1)
If Replaced by Average Teacher in the State of TN 2012-13 (1.04 SD effectiveness difference)				
$26,087	$553,264	$52,283	$527,177	$26,197
If Replaced by Average Teacher at Priority School 2012-13 (1.41 SD effectiveness difference)				
$26,087	$750,561	$70,928	$724,474	$44,841
If Replaced by Average Teacher Hired by Priority School 2012-13 (1.70 SD effectiveness difference)				
$26,087	$906,671	$85,680	$880,584	$59,593

Note. All estimates presented in this table are based on the study finding that roughly of 23 percent of teacher 351 retention bonus recipients were retained as a result of the policy. Costs of the bonus are estimated at $6000 per teacher, and benefits are based on Hanushek 2011's estimated present value of student lifetime earnings returns per 1 standard deviation increase in teacher effectiveness.

*It is worth noting that teachers who leave the priority schools might also generate additional tax revenue through improved learning outcomes at the school to which they move. To properly estimate the net benefits we need to account for the fact that some of the teachers would leave the system, while other teachers would teach at schools where they would represent a smaller benefit over the status quo.

Table 9: Sensitivity of State Returns on Investment (Tax Revenue per $1 Million Spent) to Estimated Effect of Program and Estimated Effect of Teacher on Earnings

		Teachers Retained per $1 Million Spent on Bonuses				
		10	20	30	40	50
Average Class Size of 20						
Teacher Effect on Tax Revenue from Student Earnings	Low	-$61,025	-$11,025	$5,641	$13,975	$18,975
	Preferred	-$31,527	$18,473	$35,140	$43,473	$48,473
	High	$59,916	$109,916	$126,583	$134,916	$139,916
Average Class Size of 25						
Teacher Effect on Tax Revenue from Student Earnings	Low	-$51,281	-$1,281	$15,385	$23,719	$28,719
	Preferred	-$14,409	$35,591	$52,258	$60,591	$65,591
	High	$99,895	$149,895	$166,562	$174,895	$179,895
Average Class Size of 30						
Teacher Effect on Tax Revenue from Student Earnings	Low	-$41,538	$8,462	$25,129	$33,462	$38,462
	Preferred	$2,710	$52,710	$69,376	$77,710	$82,710
	High	$139,874	$189,874	$206,541	$214,874	$219,874

Note. All estimated benefits are based on %9.45 tax rate on increased earnings associated with a 1.7 standard deviation improvement in teacher quality. Low, high, and preferred (roughly $12,000, $50,000, $21,000 per student per standard deviation increase in teacher) effect on present value of tax revenue from increased earnings estimates come from Hanushek 2011 alternative specifications due to depreciation rates and labor market returns to achievement gains.   All costs are based on $6000 per teacher cost for bonuses.

Appendix A. Sensitivity Analysis of Impact of Retention Bonus Estimates (IK Optimal Bandwidth Sample)

	Full Sample	Tested Subject Teachers	Untested Subject Teachers
RDROBUST	0.230**	0.22	0.09
	(0.10)	(0.22)	(0.15)
RDROBUST (triangular weights)	0.20**	0.21	0.07
	(0.10)	(0.21)	(0.15)
2SLS (no controls)	0.23**	0.22	0.09
	(0.11)	(0.21)	(0.17)
2SLS (controls)	0.22**	0.28	0.04
	(0.11)	(0.22)	(0.16)
BIPROBIT (no controls)	2.00***	3.18***	1.07
	(0.69)	(0.24)	(1.75)
BIPROBIT (controls)	2.07***	-	1.68**
	(0.60)	-	(0.70)

Notes: Robust standard errors in parentheses. * significant at the 10% level; ** 5% level; *** 1% level

Figure 1. Consort Diagram

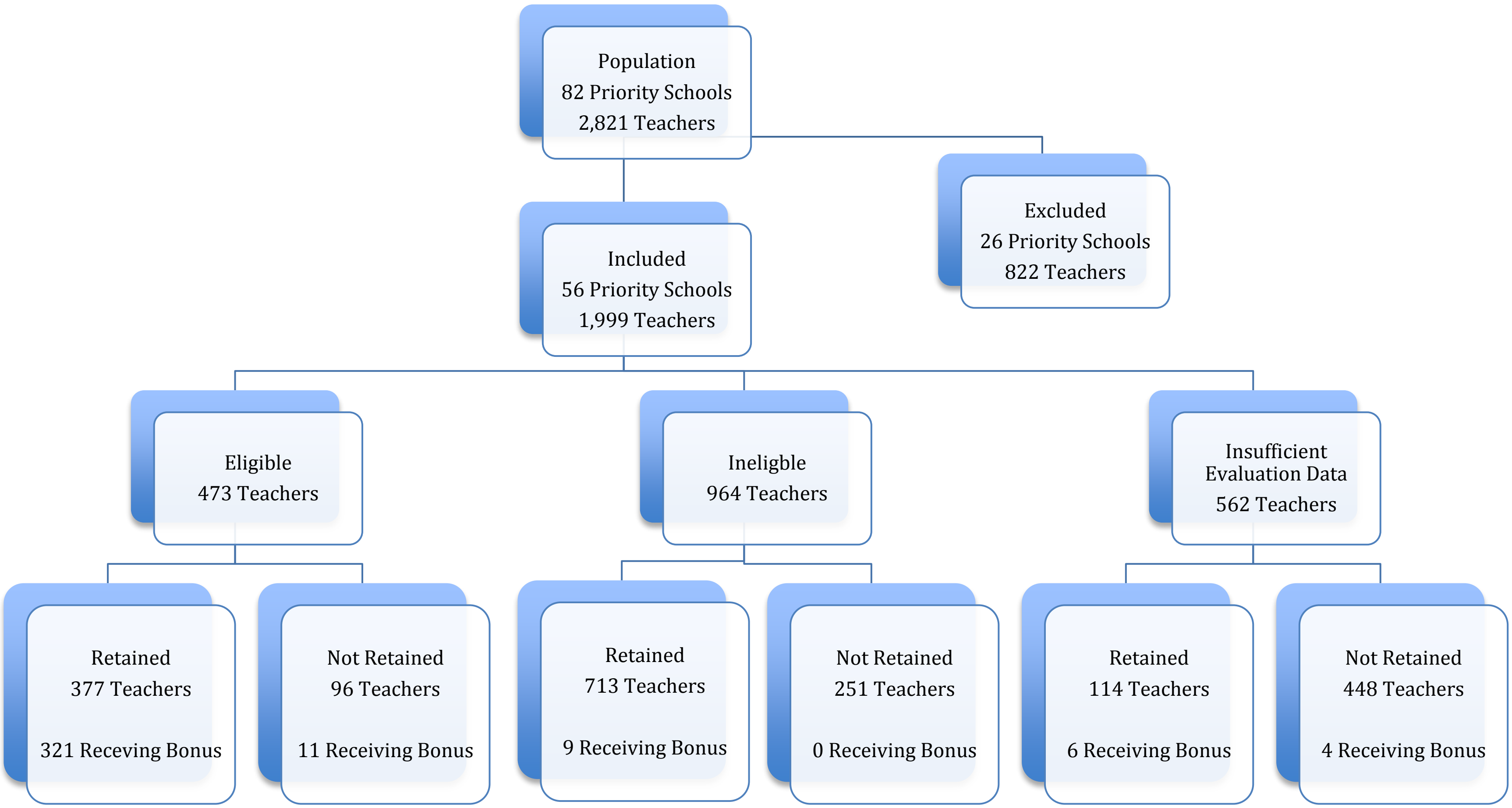

Figure 2. Noncompliance in Program Implementation

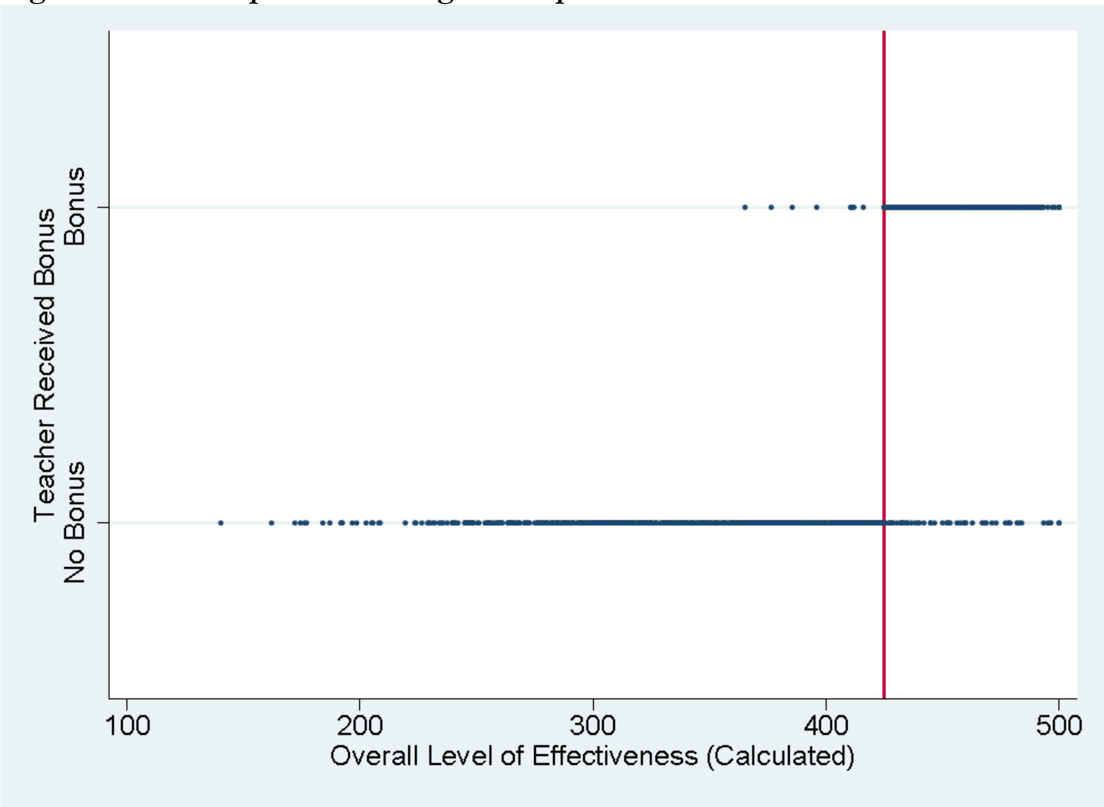

Figure 3. McCrary Test for Full Sample

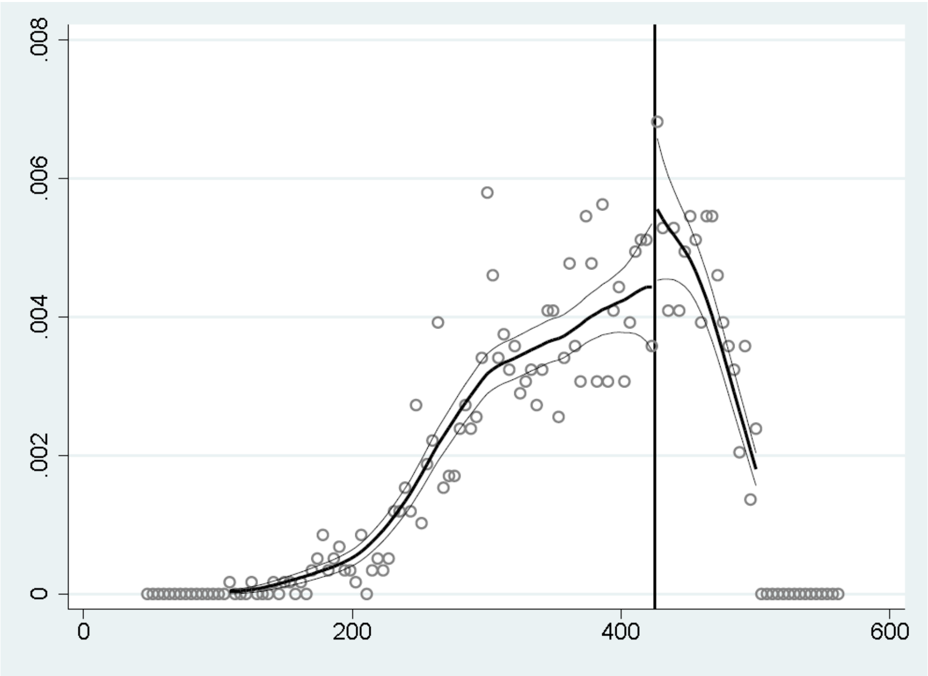

Figure 4. McCrary Test for Tested Subject Teacher Sample

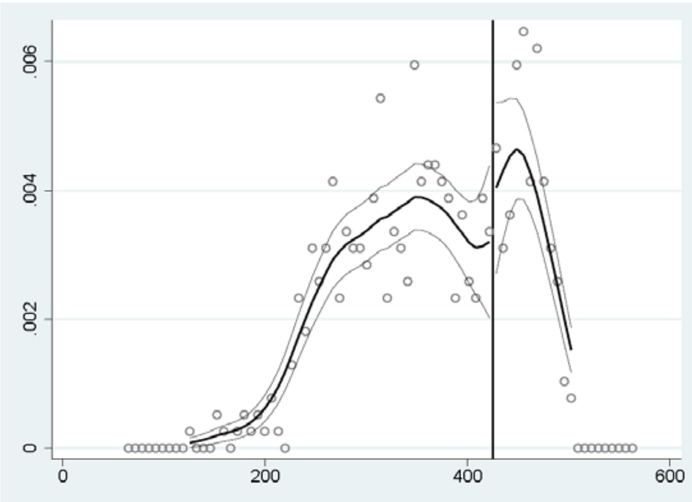

Figure 5. McCrary Test for Untested Subject Teacher Sample

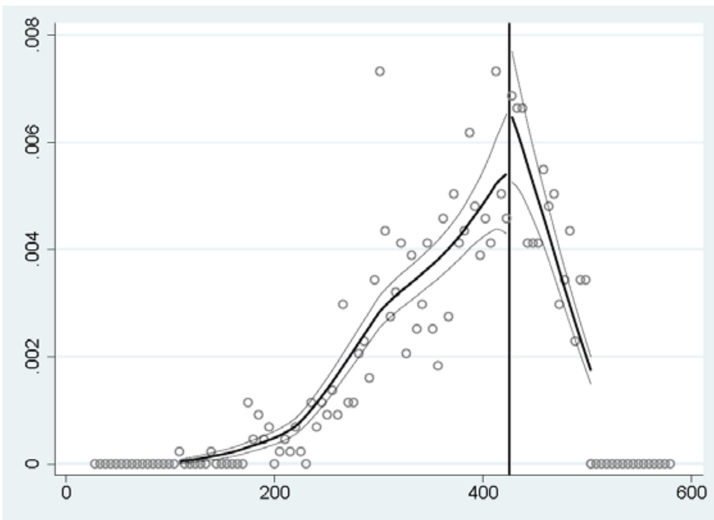

Figure 6. Impact of Retention Bonus Program, Full Sample