

Would higher salaries keep teachers in high-poverty schools? Evidence from a policy intervention in North Carolina[☆]

Charles Clotfelter^{a,b}, Elizabeth Glennie^a, Helen Ladd^a, Jacob Vigdor^{a,b,*}

^a *Sanford Institute of Public Policy, Duke University, United States*

^b *National Bureau of Economic Research, United States*

Received 22 May 2006; received in revised form 23 July 2007; accepted 26 July 2007

Available online 3 August 2007

Abstract

For a three-year time period beginning in 2001, North Carolina awarded an annual bonus of \$1800 to certified math, science and special education teachers working in public secondary schools with either high-poverty rates or low test scores. Using longitudinal data on teachers, we estimate hazard models that identify the impact of this differential pay by comparing turnover patterns before and after the program's implementation, across eligible and ineligible categories of teachers, and across eligible and barely-ineligible schools. Results suggest that this bonus payment was sufficient to reduce mean turnover rates of the targeted teachers by 17%. Experienced teachers exhibited the strongest response to the program. Finally, the effect of the program may have been at least partly undermined by the state's failure to fully educate teachers regarding the eligibility criteria. Our estimates most likely underpredict the potential outcome of a program of permanent salary differentials operating under complete information. © 2007 Elsevier B.V. All rights reserved.

Keywords: Employee turnover rates; Retention bonus; Compensating differential; Information quality; Policy design

The twin topics of teacher quality and teacher compensation have garnered considerable attention from researchers and policy makers in recent years. This attention has been motivated in part by the desire to increase the quality of individuals who select into the teaching profession, and to prevent attrition from culling the most qualified individuals from the ranks of teachers (Corcoran et al., 2004; Hoxby and Leigh, 2004). A second motivation has been concern about the uneven distribution of effective teachers across schools. Numerous studies have documented the tendency for the most qualified teachers to gravitate toward schools that serve relatively well-off children, even though salaries are often no higher in such schools (Lankford et al., 2002; Scafidi et al., 2002; Hanushek et al., 2004; Reed et al., 2006; Clotfelter et al., 2006; 2007). As states and local districts feel increasing pressure to comply with the No Child Left Behind Act's mandate that each classroom contains a "highly qualified" teacher, evidence on policy interventions that successfully recruit outstanding candidates and distribute them equitably becomes more and more valuable.

[☆] We thank Kata Mihaly, Amber Gove and Audrey Beck for outstanding research assistance, Dale Ballou, David Figlio, Joel Slemrod, participants at the 2005 AEFA and APPAM meetings, seminar participants at the University of Arkansas and Stanford University, and two anonymous referees for helpful comments on previous versions, and the Spencer Foundation for financial support. Administrative data used in this study were collected by the North Carolina Department of Public Instruction and made available by the North Carolina Education Research Data Center.

* Corresponding author. Sanford Institute of Public Policy, Duke University, United States. Tel.: +1 919 613 9226.

E-mail address: jacob.vigdor@duke.edu (J. Vigdor).

One obvious policy tool to deal with the quality and distribution of teachers would be to increase teachers' monetary compensation, perhaps in a targeted way. As straightforward as that suggestion might seem, however, a large and growing body of evidence suggests that the power of higher salaries to attract better teachers to the profession is limited (Hanushek et al., 1999; Loeb and Page, 2000), and that offering teachers pay differentials to take jobs in low-performing schools is not a cost-effective means of improving test scores, particularly when the pay differential consists of a one-time signing bonus rather than a permanent salary increase (Fowler, 2003). Although signing bonuses, absent other contractual provisions, may attract teachers to schools, they offer no incentive for teachers to stay once they have arrived.

Whether it is because of this evidence, the opposition of teacher unions, or just the sheer expense, states and districts have shown little interest in *ongoing* financial bonuses or other forms of salary differentials designed to attract and keep qualified teachers in low-performing schools. This choice is somewhat surprising given the current policy interest in market-based reforms in education; these reforms usually involve the expansion of parental choice of schools and competition among schools. Yet the use of market incentives – in the form of higher salaries or bonuses for teaching in underperforming schools – could, in theory, generate large benefits to disadvantaged students. To be sure, the benefits of higher quality teachers come with the obvious costs of paying more money for teaching services, and the question of whether the benefits exceed the costs is largely an empirical one. Moreover, the benefits of higher teaching quality are contingent on administrators' being able to pick the higher quality applicants from the enlarged pool, an ability that has been questioned by recent research (Ballou and Podgursky, 1995).

To gain insight into these questions, this paper focuses on a program implemented in North Carolina from 2001/02 until 2003/04, which awarded annual bonuses of up to \$1800 to certified teachers of math, science and special education in middle and high schools serving low-income or low-performing students (henceforth, either the North Carolina Bonus Program or, simply, the bonus program). The goal of the program was to make it easier for such schools to attract and retain qualified teachers in these fields by funding a permanent within-district salary differential. Largely because schools' initial eligibility was not determined until after the teacher labor market had cleared, most principals did not use it as a tool for recruiting new teachers. Principals also expressed concern that the program was likely to be impermanent, these concerns turned out to be justified as the state legislature canceled the program after its third year. Thus, this analysis is best thought of as an attempt to measure the short-run impact of a program designed to reduce turnover rates in low-performing or high-poverty secondary schools.

The bonus program presents an especially promising opportunity to study the impact of salary differentials on teacher retention, as it created within-school variation in teacher salaries. Most existing studies of salary differentials analyze across-district variation in salaries. Estimates derived from these studies will be biased to the extent that districts with positive or negative unobserved attributes offer higher or lower salaries. In this study, we identify the impact of the \$1800 salary differential by incorporating data before and after the program's implementation, in schools that met and barely missed the eligibility criteria, and for teachers who taught targeted subjects and other subjects. The net result is a form of difference-in-difference-in-difference analysis.

Our analysis in this paper uses longitudinal data on North Carolina public school teachers to estimate discrete-time hazard models predicting teachers' decisions to end a spell of employment at a particular school. Turnover rates are an important outcome to consider because administrators often must fill vacancies in these schools with inexperienced teachers, who have been shown to be less effective in their first years on the job than otherwise similar teachers with more experience (Clotfelter et al., 2006; Rivkin et al., 2005; Rockoff, 2004).¹ The results suggest that the sum of \$1800 per year was sufficient to reduce turnover rates of the targeted teachers by roughly 17%. Moreover, communication failures led many teachers to underestimate the probability that they would receive a bonus payment if they remained at their school, or to incorrectly infer that the bonus was portable across schools. Evidence suggests that these failures reduced the program's impact, making our result an underestimate of the potential change that would be brought about by well-implemented program of this type.

Beyond the impact of more money on turnover, another question is whether the teachers influenced by such pay are the types that principals and other administrators would most like to retain. Our tests reveal that the bonus program had

¹ Although the ultimate test of the Bonus Program would be whether it raised student achievement, our analysis of teacher retention rates suggests that any impacts would be too small to discern empirically except in samples significantly larger than the one we have. Hence, we focus here on an intermediate outcome, one which, had the program been maintained, might well have raised achievement in the long run.

the highest relative impact on experienced teachers.² As experience is one of the few observable teacher characteristics that reliably predict higher student achievement, this evidence further suggests that salary differentials may be an effective strategy for improving the quality of education in high-poverty schools. The bonus program also had stronger impacts on math teachers.

In Section 1 of this paper, we provide a brief overview of the bonus program that encapsulates information contained in Clotfelter et al. (in press). In Section 2, we describe the data set we constructed for the survival analysis and the hazard function methodology. In Section 3, we present our basic results, both on the average effects of the program and on how those effects differ across various types of teachers and schools. In Section 4, we analyze whether the implementation difficulties described in Section 1 reduced the effectiveness of the bonus program. A brief conclusion in Section 5 completes the paper.

1. The structure and implementation of the North Carolina Bonus Program

The North Carolina Bonus Program took effect in September 2001. The state, which administers a centralized payroll system covering employees in each of the state's 117 districts, began providing an annual salary supplement of \$1800 to teachers certified in math, science, and special education teaching those subjects in middle schools or high schools that met either of the following criteria:

- (1) 80% or more of students had to be eligible for free or reduced price lunch, or
- (2) 50% or more of its students had to perform below grade level in *both* Algebra 1 and Biology, as measured by the state's end-of course tests.

Part-time teachers and those who taught both targeted and non-targeted subjects received a prorated bonus amount. Uncertified teachers of math, science and special education received no bonus payment. These uncertified teachers comprised roughly one-third of all teachers of these subjects statewide, with significantly higher fractions in some schools, including many targeted by the bonus program. Over the next three years, nearly two thousand teachers working in 148 schools in 65 districts were eligible for bonus payments in at least one year. Schools were informed of their eligibility close to the beginning of the school year, and teachers received the bonus as a monthly supplement to their paycheck.

Table 1 displays conditional turnover rates for teachers working in schools that became eligible for the bonus program in its first or second year of implementation, along with rates for a set of comparable schools.³ The average turnover rates are computed as the probability that a teacher working in a school in year t ceases to be employed at that school in year $t+1$, averaged over a four-year period immediately before and after the bonus program's implementation.⁴ Both sets of schools are relatively impoverished and low-performing, but the "ever-eligible" schools are generally worse off along observable dimensions. Across subjects and school types, anywhere from 22 to 37% of teachers working in a school in a given year departed at the end of that year. Turnover rates are somewhat lower in the comparable schools, but still substantial. Interestingly, the turnover rates for math, science, and special education teachers are relatively high in ineligible-but-comparable schools, but do not particularly stand out in schools that participated in the bonus program. This is the first clue that the bonus program may have had an impact.⁵ As has been noted in prior literature, turnover rates are particularly high among inexperienced teachers, and among those with more than thirty years of experience.

In North Carolina, teacher salaries are determined largely by a statewide schedule that varies by experience level, highest degree attained, and National Board for Professional Teaching Standards certification status. Local districts have the option of adding a locally-financed supplement to this base salary, and most do. In the time period under consideration, teacher salaries ranged from just over \$25 000 for a teacher with no prior experience and no advanced degrees in a district with no local supplement to just over \$60 000 for teachers with 30 or more years of experience and

² Since experienced teachers generally have lower turnover rates overall, the bonus program's absolute impact, though still larger for experienced teachers, was closer to uniform across the experience spectrum.

³ The set of comparable schools is the control group utilized in the baseline hazard model analysis. The selection of this set is detailed in Section 2.

⁴ Note that these turnover rates include temporary departures from schools, such as those associated with maternity leaves. In the analysis below, we treat all departures from schools equally, largely because we are ultimately unable to observe whether some departures are temporary or not. So long as these temporary departures are idiosyncratic and uninfluenced by policy, they will be incorporated into baseline hazards in our analysis below and not produce bias in our estimates of the impact of the bonus program. If for some reason the bonus program increased (decreased) the propensity to take a temporary leave, we will understate (overstate) the impact on retention.

⁵ For further rudimentary analysis of the bonus program's impact, including graphical analysis, see Clotfelter et al. (in press).

Table 1
Conditional turnover rates in North Carolina secondary schools

	“Ever eligible” schools	Comparable ineligible schools
<i>Subject taught</i>		
Math	0.305	0.297
Science	0.316	0.274
English	0.306	0.263
History	0.272	0.220
Social Studies	0.295	0.262
Foreign Language	0.375	0.260
Arts	0.324	0.228
Special Education	0.320	0.265
Practical	0.275	0.224
Technical	0.224	0.199
<i>Type of school</i>		
Middle school	0.292	0.335
High school	0.308	0.229
<i>Experience</i>		
1 year	0.450	0.399
2 years	0.399	0.315
3 years	0.378	0.283
4–9 years	0.291	0.241
10–19 years	0.223	0.187
20–29 years	0.202	0.169
30–39 years	0.353	0.304
40+ years	0.524	0.438

Note: Sample consists of teacher/school/year observations for personnel working at least part time in a selected school. A turnover rate is the proportion of teachers working in a school in year t who do not teach in the same school in year $t+1$. Temporary absences from a school, such as maternity leaves, are included in the turnover rate. “Ever eligible” schools are those which participated in the North Carolina Bonus Program in either the first or second years of implementation. The procedure for selecting comparable ineligible schools is described in the text. $N=29,584$.

an advanced degree working in a district with a particularly generous local supplement. A teacher with 15 years of experience and a master’s degree, in a district with average supplement levels, earned about \$43 000.⁶ Thus the \$1800 bonus represented anywhere from 3 to 7% of a teacher’s base salary, and about 4% on average.

Although teachers needed to be employed in an eligible school to begin receiving the bonus, continued receipt did not require continued school eligibility. The only requirement was that the teacher continues to work in the same school, teaching one of the eligible subjects. This provision was intended to eliminate both uncertainty and any perverse incentives for teachers to keep test scores low so that the school would remain eligible.

For reasons described more fully in Clotfelter et al. (in press), the vast majority of teachers receiving bonus payments in the 2003/04 school year misunderstood the provisions of the bonus program, according to responses to surveys we administered in 2004.⁷ The most common misconception was that continued receipt of the bonus required that the school remain eligible. Thus, although the program was designed to assure teachers that they would continue to receive the bonus (if they did not change schools or subjects), many teachers did not understand this provision. Teachers’ misperceptions led them to expect future bonus payments lower than what the enabling legislation specified.⁸ For teachers who inaccurately believed that their continued receipt was linked to their school’s continuing eligibility, the expected value of future bonus payments would be strongly linked to the criterion variable, either subsidized lunch receipt rates or algebra and biology test

⁶ These salary figures are taken from the 2001–2002 schedule and are expressed in 2001 dollars. By 2006–2007, base salaries had increased 10–15% in nominal terms, while the average local supplement had increased by nearly 30%.

⁷ We administered surveys by mail, with telephone follow-up when necessary, to 1165 principals and teachers in eligible schools. The response rate was 83% for principals and 72% for teachers. More detail on the survey and our findings can be found in Clotfelter et al. (in press).

⁸ Ironically, events proved the doubters correct. Despite the program’s stated assurances, the state legislature canceled the bonus program after the 2003/04 school year. Thus skeptical teachers may also have – correctly – factored a lower probability of continued receipt into their calculations. The state’s decision to cancel the bonus program was made rather abruptly in the summer of 2004. It is unlikely that the foreshadowing of this event directly impacted the teacher decisions analyzed in this paper, as the latest moving decisions factoring into our analysis would have occurred by Fall 2003.

failure rates, determining that eligibility. Because the criterion variables fluctuate from year to year, a misperceiving teacher in a school with an 80.1% rate of subsidized lunch receipt, for example, would have had a lower expected value of future bonus payments than a similarly misperceiving teacher in a school with a 90% rate. In Section 4 below, we will use this fact to test whether misperceptions altered the impact of the bonus program.

Straightforward economic theory suggests that the extra pay provided through this bonus program should increase the supply of teachers willing to work in an eligible school, and by extension reduce the rate of departure for teachers currently employed at that school. The magnitude of the impact is an empirical question. The less elastic the labor supply curve for teachers in disadvantaged schools, the less an effect we should expect to observe.

The North Carolina Bonus Program was ill-suited as a tool for recruiting teachers, as principals were not informed of their school's eligibility for the upcoming year until after the recruiting process had ended (Clotfelter et al., *in press*).⁹ Thus it is reasonable to think that the program's impact on turnover closely approximates its complete impact on the distribution of teachers in public schools. Programs designed in different ways, however, could well have broader impacts on the recruitment of teachers to specific schools, or the recruitment into the teaching profession more generally.¹⁰

2. Data and methods

To estimate how the bonus affected the decisions of teachers about whether to remain in particular schools, we utilized a longitudinal dataset of teacher employment covering a period beginning two years prior to the implementation of the bonus program and continuing for the next three years. The unit of observation in our analysis is a teaching spell, defined as a period in which a teacher works continuously at a single school, with the outcome of interest an indicator for whether that spell ended during the time period under observation.¹¹ In practice, only the first four years of our sample are useful for this analysis, since for any year t we must refer to data for year $t+1$ to determine whether a spell has ended. It is also important to note that many of our observed spells are left-censored — that is, they begin during or before our first year of data, implying that we cannot directly observe the length of the spell. For the most part, we will treat these censored spells as though they had begun in the first year of our dataset. In alternative specifications, we drop these left-censored spells from the analysis sample.

To analyze the factors determining whether teachers remain employed at the same school from one year to the next, we estimated a series of discrete-time hazard models. Such models predict the probability that a teaching “spell” will end after year t , conditional on that spell having lasted that long in the first place. As a baseline specification, we adopt the Cox proportional hazard model, a non-parametric specification that allows us to sidestep the estimation of the baseline hazard function. This approach is advantageous in our setting because of the high prevalence of spells in which the true duration is unknown. The principal disadvantage of the Cox model is its difficulty in handling spells of exactly equal length, of which there are many in our dataset. We thus report alternative specifications employing the Weibull proportional hazard model, with and without censored spells in the sample.

Our identification strategy has three basic components: we compare hazard rates (1) before and after the implementation of the bonus program, (2) across eligible and ineligible teachers working in the same schools, and (3) across teachers in eligible schools and in schools that narrowly missed the eligibility criteria. Ours is thus a difference-in-difference-in-difference strategy, with the third difference resembling a hybrid between a randomized experiment and a regression discontinuity design. A representative estimating equation is thus:

$$\text{logit}[\lambda(t_{ijs})] = \alpha + \beta_1 E_j + \beta_2 E_s + \beta_3 Y + \beta_4 E_j * E_s + \beta_5 E_j * Y + \beta_6 * E_s * Y + \beta_7 * E_s * E_j * Y + X_{ijs} \gamma \quad (1)$$

⁹ The main reason for this delay was the program's reliance on end-of-course test scores to determine eligibility. These tests are generally taken at the end of a school year, and the scores reported months thereafter. In many cases, principals could have inferred eligibility with a low degree of uncertainty. Our interviews with principals at eligible schools, however, indicate that these individuals were sufficiently risk-averse to avoid promoting a probable but not guaranteed benefit.

¹⁰ Results below point to two elements of program design that could influence the impact of a bonus program, beyond the dollar amount of the payment. First, a program perceived as permanent appears to be more effective than a program perceived as temporary. From this perspective, the North Carolina legislature's decision to cancel a program that had been promoted as permanent introduces a potential reputation problem for future policy initiatives. Second, we will argue below that certain results point to the conclusion that bonus payments are more effective at influencing decisions regarding where to teach relative to decisions regarding whether to teach. Thus, the cost of redistributing the existing stock of teachers would most likely be less than the cost of achieving wholesale improvements in the set of individuals selecting into the teaching profession.

¹¹ In practice, there are many instances of “interrupted” spells, where teachers leave a school after year t and return in year $t+2$ or a later date. Maternity leaves often result in these interrupted spells. For our purposes, we treat any departure from a school, whether permanent or temporary, equivalently, largely because we are unable to distinguish among these types of departures in the last year of our data.

where i indexes individuals, j indexes subjects, and s indexes schools. The dependent variable $\lambda(t_{ijs})$ is the probability that a spell ends at the close of period t , conditional on that spell lasting through period t .¹² The variable E_j indicates whether the teacher in question is certified in an eligible subject – math, science, or special education – and teaching that subject in the school in question. The variable E_s indicates that the teaching spell takes place in an “ever eligible” school, one that meets the eligibility criteria in one or more of the years when the program was actually in place. The variable Y indicates that the spell took place in a year when the bonus program actually existed. These three variables can be interpreted as main effects in a difference-in-difference-in-difference framework.

The three two-way interaction terms in Eq. (1) represent the differential impact of teaching an eligible subject in an eligible school at any point in time, the impact of teaching an eligible subject in a year when the bonus program was in existence in any school, and the impact of teaching any subject in an ever-eligible school in a year in which the bonus program existed. Finally, the three-way interaction term represents the differential hazard rate of teachers who taught eligible subjects in ever-eligible schools in years when the bonus program actually existed. The coefficient β_7 thus represents the difference-in-difference-in-difference estimate of the bonus program’s impact on the exit hazard. Finally, the matrix X_{ijs} contains assorted characteristics of the teacher and school.

Eq. (1) is a slightly stylized version of the models we actually estimate, for two reasons. Several covariates are time-varying, and Eq. (1) omits subscripts for time. Perhaps more importantly, our difference-in-difference-in-difference estimate β_7 is not actually determined by using a three-way interaction term. Instead, we use an indicator for whether a teacher actually received a bonus payment at the specified point in time. This strategy is necessary for two reasons. First, there is not a single clean break into the “pre” and “post” program periods, as some schools became eligible subsequent to the first year of estimation, and others became ineligible after a period of eligibility. Second, our process of inferring eligibility is in practice imperfect. One source of imperfection is actual errors in the administration of the program, which we have confirmed in conversations with state and school personnel.¹³ A second is the presence of ambiguities and coding errors in the three administrative datasets required to infer eligibility: a certification database, a payroll database indicating the school(s) at which a teacher worked in a given year, and a third database of actual teaching assignments.¹⁴

One concern with the estimation of Eq. (1) is that there may be important, time and subject-varying differences between ever-eligible schools and the remainder of secondary schools in North Carolina. To address this concern, most specifications reported below will restrict the sample to ever-eligible schools along with a set of comparable “control” schools. To identify the set of control schools, we ranked schools by the average over five years of the criterion variables used to determine eligibility for the bonus program (percent of students receiving subsidized lunch for middle schools and percent of test-takers failing Algebra I and Biology exams for high schools). The five years include the three used to determine eligibility in the bonus program and the two previous years, 1998/99 through 2002/03.

Figs. 1 and 2 illustrate the control school selection process for middle and high schools, respectively. For middle schools, the illustration is a histogram, as only the free or reduced lunch rate could have qualified a school for the program. For high schools, the illustration is a scatterplot, as schools could have qualified by either criterion. In practice, all but one high school qualified according to the failure rate criterion. In both cases, there is at least one example of a school that became eligible even though their five-year average criterion variables were below the eligibility threshold. This outcome occurs because eligibility is based on a single year’s value for the criterion variable, and these figures present five-year averages.¹⁵ In both cases, there are also examples of schools with criterion variables above the eligibility threshold who nonetheless never participate in the program. In both middle and high schools, this

¹² In parametric hazard models, this conditional probability is modeled as a function of t as well. In the Cox proportional hazard model, this so-called baseline hazard is not directly estimated.

¹³ We discuss some of these errors in more detail in Clotfelter et al. (in press). Our analysis operates under the presumption that any impact of the bonus program was conditional on actually receiving a payment, rather than merely being eligible for one.

¹⁴ Many of these ambiguities concern special education teachers. For example, a certified special education teacher offering instruction in an academic subject may be coded as teaching that subject, even if the course section serves exclusively special education students. One potential concern arising from our inability to perfectly replicate the state’s list of bonus recipients is endogenous take-up. If teachers were required to take some action in order to receive a bonus, for example, one might expect that teachers who considered themselves more likely to remain at a school would be more likely to take up the bonus. In practice, however, teachers who met eligibility criteria were required to take no further action to receive the bonus. In fact, our survey of teachers revealed that one in twelve were not even aware that they were receiving a bonus, in spite of the fact that a line item for the bonus appeared on their pay stub (Clotfelter et al., in press).

¹⁵ Interestingly, the “treatment” schools in our analysis tend to be smaller than the “control” schools, which is to be expected when eligibility is based on a noisy indicator. Smaller schools are more likely to have criterion variable values that depart markedly from the mean in any particular year.

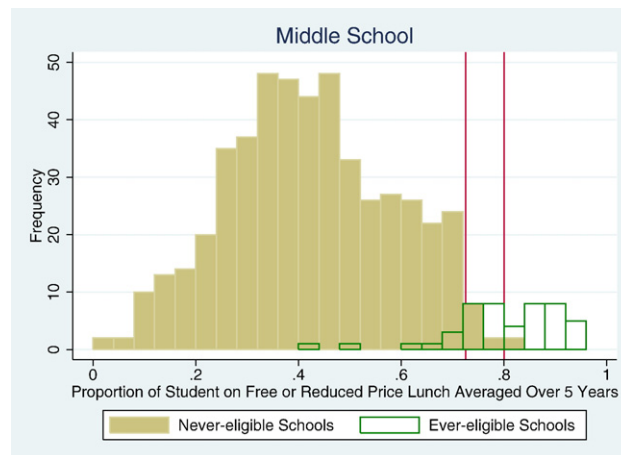


Fig. 1. Histogram of eligibility criterion variable for middle schools. The rightmost vertical line indicates the threshold for bonus program eligibility, and the leftmost vertical line indicates the threshold for inclusion in our baseline estimation sample. Schools with criterion variables below the eligibility threshold can become eligible because eligibility is based on a single years' value for that variable, whereas the proportions plotted here are averaged over five years.

situation can occur when the criterion variable is trending downward over time.¹⁶ In high schools, it can also occur when failure rates on one of the two criterion examinations is high, but the other falls below the 50% threshold.

To select control schools, we subjectively set a second threshold value for middle schools and high schools. These threshold values are displayed on the histogram in Fig. 1 and the scatterplots in Fig. 2. We excluded middle schools with five-year average free and reduced price lunch rates below 72.5%.¹⁷ We exclude high schools where the sum of failure rates on Algebra I and biology exams was less than 83%, unless their free and reduced lunch eligibility rate was above 72.5%. We set these thresholds at local break points in the distribution of each criterion variable, to include a reasonable number of control schools in the analysis without expanding the set so far as to include schools that look dramatically different in terms of the criterion variable. In both cases, a few eligible schools have at least one criterion variable below its respective threshold for inclusion in the control group.

As this selection of control schools is inherently subjective, we present results below to test the sensitivity of the results. In one specification, we include all ineligible schools in the control group, while introducing a few school-level control variables into the analysis. In a second specification, we omit all ineligible schools and identify the impact of the bonus program in a difference-in-difference framework, comparing patterns for eligible and ineligible teachers before and after program implementation. These specifications show results that are statistically indistinguishable from our baseline results.

The results of the selection process are summarized in Table 2. Treated middle schools had average criterion variables – subsidized lunch receipt rates – that ranged from 0.633 to 0.932. Our selected group of control middle schools had average criterion variables in the range from 0.725 to 0.841. Average criterion variables tended to be much lower in middle schools excluded from the baseline analysis.

Among eligible high schools, failure rates on biology and algebra test scores tend to be quite a bit higher than the 50% eligibility threshold, particularly for biology exams. Note, however, that there are eligible schools where the failure rate averaged over several years is below the threshold. Selected control schools have average failure rates that generally fall close to the eligibility threshold. Because high schools had to have failure rates in excess of 50% on both exams to be eligible, it was possible to have very high failure rates on one exam without attaining eligibility. High schools excluded from the analysis tend to have average failure rates well below the eligibility threshold.

¹⁶ The control group also includes schools that became eligible in the final year of the bonus program, which is not incorporated in our analysis sample. Excluding these schools does not influence the results.

¹⁷ We also excluded schools which came into existence or ceased to exist at some point during the sample period.

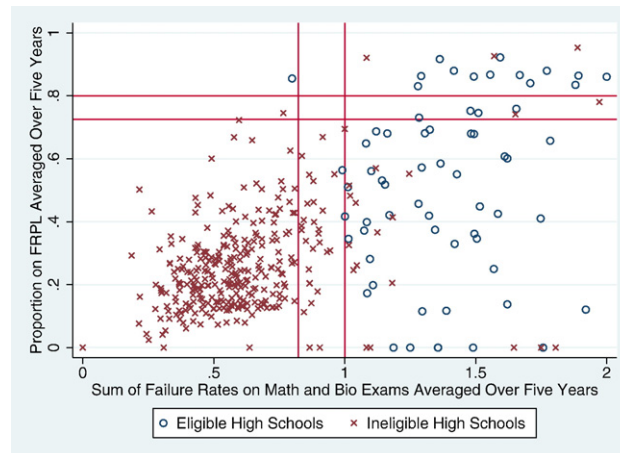


Fig. 2. Scatterplot of eligibility criterion variables for high schools. The rightmost vertical/topmost horizontal lines indicate the thresholds for bonus program eligibility, and the leftmost vertical/bottom horizontal line indicates the thresholds for inclusion in our baseline estimation sample. Schools with criterion variables below the eligibility threshold can become eligible because eligibility is based on a single years' value for that variable, whereas the proportions plotted here are averaged over five years.

Finally, note that well over half of our treatment and control groups are high schools, even though the majority of secondary schools in the state of North Carolina are middle schools. This pattern reflects the fact that, given the eligibility criterion, it was easier for a high school to qualify than a middle school.

Our longitudinal dataset on teacher workplace locations begins with the 1999/2000 school year, two years before the bonus program was implemented. Since the bonus program was not passed into law until the 2001/02 school year had already started, we will presume that any effects on teacher decisions commenced at the close of that year. The latest year for which administrative data are currently available is 2003/04. Hence we constructed a four-year panel dataset, where the outcome of interest is whether a teacher employed in a particular school in year t remains employed at the same school in year $t + 1$. Summary statistics for the teacher/year variables included in the analysis appear in Table 3. The teachers in our sample are roughly 65% female, 34% black, and 4% Hispanic or other nonwhite. These racial proportions differ in important respects from the overall population of secondary school teachers in North Carolina; in the excluded schools, only about 12% of the teacher/year observations are for black teachers.

Within the treatment and control schools, about 7% of the teacher/year observations represent individuals in their first year of the profession, the median teacher has between 10 and 19 years of experience, and 29% of the teachers in these schools had some form of graduate degree. Roughly 9% of our observations correspond to teachers who received a bonus payment in the relevant year.¹⁸

3. Results

3.1. Simple difference-in-difference-in-difference results

Fig. 3 expands on the basic information in Table 1 by plotting turnover rates by year, individual eligibility, and school eligibility — a graphical form of difference-in-difference-in-difference analysis, albeit one that does not incorporate spell duration dependence, as our hazard models do. Eligible teachers are labeled MSSE, for math, science, and special education. “Other” teachers include uncertified teachers of those same subjects, plus any teacher of any other subject. Schools are classified as “ever eligible” and “never eligible.” This simple graph does not permit a thorough analysis of the impact of the bonus program, especially as not all “ever eligible” schools are eligible in a

¹⁸ Under some circumstances, such as teaching an eligible subject part time, some teachers received prorated bonuses. In some alternative models, we replaced the indicator for whether the teacher received any bonus with the amount of the bonus received. Because the decision to teach an eligible subject part time may be correlated with other unmeasured individual characteristics, however, we relegate those results to footnotes.

Table 2
School level summary statistics

	Ever eligible schools	Control schools	Excluded schools
High schools as a proportion of sample schools	0.631 ($n=103$)	0.625 ($n=56$)	0.382 ($n=791$)
Proportion eligible for free or reduced price lunch (middle school criterion variable)	0.821 [.633, .932] ($n=38$)	0.767 [.725, .841] ($n=21$)	0.421 [0, .995] ($n=488$)
Failure rate on Biology Exam (high school criterion variable)	0.743 [.479, 1.000] ($n=65$)	0.501 [.310, .938] ($n=35$)	0.348 [0, 1] ($n=284$)
Failure rate on Algebra Exam (high school criterion variable)	0.648 [.233, 1.000] ($n=65$)	0.474 [.273, .708] ($n=35$)	0.238 [0, 1] ($n=284$)

Note: Unit of observation is the school. A school is defined to be in the “ever eligible” group if it met the eligibility criteria for inclusion in the bonus program in the first or second year of implementation. “Control” group schools were not eligible in either of the first two years, but approached the thresholds for eligibility. Five of the 56 control group schools became eligible for the bonus program in its third year. The range is in brackets, and the number of observations are in parentheses.

particular year, but it does show very basic evidence pertinent to the analysis below. The vertical line in this graph separates the pre- and post-implementation periods.

The graph shows that turnover rates, regardless of subject, tend to be higher in “ever eligible” schools relative to others. This is a pattern first observed in Table 1 above. Interestingly, though, turnover rates appear to be significantly higher in eligible schools in subjects other than math, science, and special education. This is also consistent with evidence in Table 1. Not apparent in that earlier table, however, is the time trend apparent here: the implementation of the bonus program is associated with a reduction in turnover rates in eligible schools that is not apparent in comparison

Table 3
Teacher-year level summary statistics

	Mean
<i>Subject taught (no omitted category)</i>	
Math	0.225
Science	0.184
English	0.300
History	0.073
Social Studies	0.185
Foreign Language	0.044
Arts	0.070
Special Education	0.068
Practical	0.093
Technical	0.038
Female	0.646
<i>Race (White omitted)</i>	
Black	0.339
Hispanic	0.013
Asian	0.005
Native American	0.022
Multiracial	0.001
<i>Education level (Bachelors omitted)</i>	
Masters	0.278
Other graduate degree	0.016
<i>Experience (10–19 years omitted)</i>	
1 year	0.067
2 years	0.054
3 years	0.049
4–9 years	0.216
20–29 years	0.225
30–39 years	0.070
40+ years	0.002
Paid bonus	0.085

Note: Sample consists of teacher/school/year observations for personnel working at least part time in a treatment or control group school. $N=29,584$.

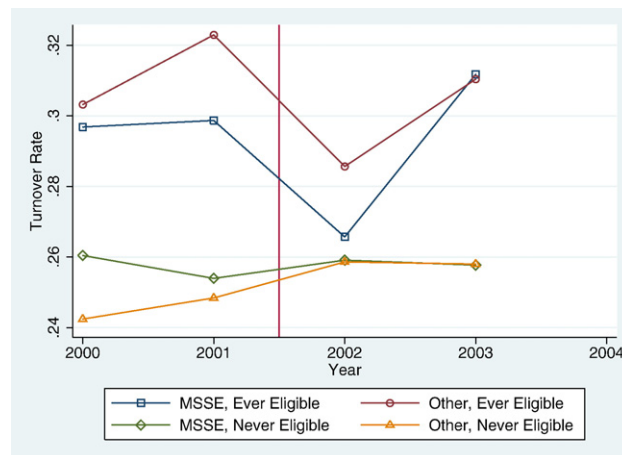


Fig. 3. Turnover rates, by individual and school-level eligibility, pre- and post-bonus program implementation.

schools. This reduction is not confined to teachers of eligible subjects, however. Moreover, the reduction in turnover rates is confined to the first year of implementation. In the second year of implementation, turnover rates return to levels at or above those exhibited in the pre-implementation period. Overall, this simple look at the data suggests that there may have been some impact associated with the program, but also begs a more thorough analysis to determine whether some confounding variables or trends might truly explain this pattern.

Table 4 presents a basic Cox proportional hazard model implementing a difference-in-difference-in-difference estimate of the bonus program's impact on teacher turnover rates. It closely approximates Eq. (1), without incorporating any vector of teacher covariates, but including a full set of year fixed effects. Table entries are hazard ratios, which indicate the multiplicative impact of a unit change in an independent variable on the conditional probability of a spell ending at the end of a particular year.

The main effects show several significant patterns.¹⁹ Turnover rates in ever-eligible schools were significantly higher than in comparable ineligible schools, consistent with the evidence in Table 1 and Fig. 3, and an indication that the match between the treatment and control groups is imperfect. Certified math, science, and special education teachers were neither more nor less likely to depart relative to other teachers — again, a pattern consistent with basic evidence shown to this point. This insignificant pattern may incorporate two distinct patterns: teachers of these subjects may be more likely to depart overall, but teachers who have gone through the process of certification may also be generally less likely to leave teaching.

The interaction between teacher eligibility and the post-program indicator, β_5 in Eq. (1), is greater than one and marginally significant, indicating that the turnover rates for eligible teachers were trending up over time. The effect of being in a currently eligible school, which approximates the interaction β_6 in Eq. (1), is significantly less than one, which suggests that turnover rates for all types of teachers decreased in ever-eligible schools, relative to other schools, once they actually began participating in the program. Again, this is consistent with the basic evidence in Fig. 3. This could reflect a spillover impact of the bonus program on ineligible teachers, who may have been encouraged to remain in their positions by teachers who received the bonus. The coefficient β_4 from Eq. (1), on the interaction between teacher eligibility and school ever-eligibility, is not significantly different from one.

¹⁹ The year effects in this specification, and later specifications, should not necessarily be interpreted as evidence of increasing hazard rates over time. Cox proportional hazard models incorporate a non-parametric baseline hazard function, which indicates how the conditional probability of exit from a teaching spell changes over time. Given the relatively short overall duration of our panel dataset, we face what amounts to a collinearity problem between the hazard function and the year fixed effects. In later specifications, we introduce controls for teacher experience, which potentially exacerbate this problem. Year fixed effects are incorporated in these specifications to introduce the flexibility traditionally associated with difference-in-difference-in-difference models. The operationalization of hazard models, however, implies that coefficients on these fixed effects should not be taken as evidence of trends in turnover rates. In fact, predicted survival probabilities derived from these models increase over time, consistent with declining duration dependence and inconsistent with large increases in hazard rates. The collinearity concern does not apply to estimates of the bonus program's impact, as much of the variation used to estimate that impact is cross-sectional in nature.

Table 4

Basic estimate of the program's impact

Teacher receives a bonus payment	0.848** (0.057)
Teacher is certified in Math, Science or Special Education and is employed by an ever-eligible school	1.005 (0.062)
Teacher is employed by a currently eligible school	0.802** (0.034)
Teacher is certified in Math, Science or Special Education in a post-program year	1.114* (0.070)
Teacher is certified in Math, Science or Special Education	0.996 (0.054)
Teacher is employed by an ever-eligible school	1.286** (0.041)
Year is 2000	1.141** (0.050)
Year is 2001	1.560** (0.070)
Year is 2002	1.907** (0.076)
<i>N</i>	29,562
Log likelihood	−59,123.93

Note: Table entries are hazard ratios, with standard errors in parentheses. The hazard refers to the probability of exiting a school after period t , conditional on remaining in that school until period t . Unit of observation is the teacher/school/year. ** denotes a hazard ratio significantly different from 1 at the 5% level; * the 10% level.

The top row of the table presents the difference-in-difference-in-difference estimate of the bonus program's impact on turnover. Teachers who received a bonus were 15% less likely to end their teaching spell at the end of a year. This effect is statistically significant at the 5% level.²⁰

3.2. Refined estimates of the average program impact

The specifications reported in Table 5 incorporate additional covariates. To begin, we introduce controls for the subject(s) taught by each teacher in each year.²¹ Although coefficients associated with these controls are not included in the table, results show significant variation across subject areas. Math, special education, and foreign language teachers are significantly more likely to depart, while history, practical arts, and technical/vocational teachers are significantly less likely to leave. Controlling for subject area leads to a modest increase in the estimated impact of the bonus program: recipients are now estimated to be 17% less likely to depart, other things equal. The hazard ratio continues to be significantly less than one at the 5% level.

In the second column of Table 5, we introduce controls for teacher gender and race. While there is no significant difference in the departure rates of male and female teachers, Hispanic and multiracial teachers display higher hazard rates, and Native American teachers lower rates. Once again, these controls lead to a very modest change in the estimated impact of the bonus program, with no change in its statistical significance.

Controls for teacher experience and education levels, introduced in the third column, reveal patterns consistent with the basic statistics presented in Table 1. Teachers in their first year in the profession are 48% more likely to depart than those with 10 to 19 years of experience; departure rates are somewhat lower but still elevated for teachers in their second or third years. Departure rates rise significantly once teachers reach 30 years of experience. Teachers with graduate degrees are significantly more likely to depart schools in this sample, perhaps because these teachers have access to a wider array of opportunities in other schools and districts. Controlling for these factors produces a slight increase in the estimated impact of the bonus program, to an 18% reduction in turnover rates.

The final specification permits hazard rates to differ for middle and high schools. A substantial difference emerges across the two school types. The average likelihood of departing in any year is 9% lower for teachers in high schools than for those in middle schools, other things equal. Controlling for this factor reduces our estimate of the program impact slightly, back to a 17% reduction in the probability of departure, conditional on remaining in the school until the

²⁰ Using the dollar amount of the bonus in place of a binary indicator for receipt, we find a statistically significant pattern suggesting that teachers receiving the full \$1800 were 15% less likely to depart, while those receiving a prorated bonus of \$900 were 8% less likely to depart.

²¹ Note that these variables are not mutually exclusive since teachers can teach multiple subjects in the same year. Hazard ratios associated with these variables should be interpreted as the marginal impact of teaching a particular subject in addition to the set of subjects one already teaches.

Table 5
Refined estimates of the program's impact

	(1)	(2)	(3)	(4)
Teacher receives a bonus payment	0.828** (0.057)	0.826** (0.057)	0.815** (0.055)	0.827** (0.056)
Teacher is certified in Math, Science or Special Education and is employed by an ever-eligible school	1.014 (0.063)	1.016 (0.063)	1.036 (0.064)	1.030 (0.064)
Teacher is employed by a currently eligible school	0.805** (0.034)	0.805** (0.034)	0.804** (0.034)	0.806** (0.034)
Teacher is certified in Math, Science or Special Education in a post-program year	1.124* (0.070)	0.126* (0.071)	1.127* (0.071)	1.120* (0.070)
Teacher is certified in Math, Science or Special Education	0.918 (0.056)	0.916 (0.056)	0.892* (0.055)	0.907 (0.056)
Teacher is employed by a potentially eligible school	1.276** (0.041)	1.291** (0.042)	1.283** (0.042)	1.276** (0.041)
Year is 2000	1.151** (0.052)	1.155** (0.052)	1.221** (0.055)	1.224** (0.055)
Year is 2001	1.560** (0.070)	1.559** (0.070)	1.660** (0.075)	1.644** (0.075)
Year is 2002	1.901** (0.076)	1.092** (0.076)	1.937** (0.077)	1.934** (0.077)
Female	–	0.981 (0.027)	0.995 (0.027)	0.991 (0.027)
Black	–	0.973 (0.027)	0.971 (0.027)	0.963 (0.027)
Hispanic	–	1.085 (0.115)	1.095 (0.116)	1.089 (0.115)
Asian	–	1.015 (0.171)	0.976 (0.164)	0.986 (0.116)
Native American	–	0.721** (0.071)	0.733** (0.072)	0.704** (0.070)
Multicultural	–	1.704** (0.428)	1.697** (0.426)	1.688** (0.424)
Masters Degree	–	–	1.129** (0.032)	1.135** (0.033)
Other graduate degree	–	–	1.210** (0.110)	1.224** (0.112)
1 year of experience	–	–	1.475** (0.067)	1.470** (0.067)
2 years of experience	–	–	1.268** (0.067)	1.263** (0.067)
3 years of experience	–	–	1.120** (0.063)	1.118** (0.063)
4–9 years of experience	–	–	0.965 (0.034)	0.966 (0.034)
20–29 years of experience	–	–	0.725** (0.029)	0.726** (0.029)
30–39 years of experience	–	–	1.265** (0.063)	1.268** (0.063)
40+ years of experience	–	–	1.695** (0.402)	1.715** (0.407)
High school	–	–	–	0.908** (0.027)
Subject area controls	Y	Y	Y	Y
N	29,562	29,562	29,562	29,562
Log likelihood	–59,106.71	–59,097.59	–58,969.25	–58,963.80

Note: Table entries are hazard ratios, with standard errors in parentheses. Unit of observation is the teacher/school/year. ** denotes a hazard ratio significantly different from 1 at the 5% level; * the 10% level.

year in question. This hazard ratio of 0.83, which is significantly different from one at the 5% level, represents our best estimate of the overall mean impact of the bonus program on the retention of eligible teachers.²² Across these varying specifications, the magnitude of the coefficient on the three-way interaction term that represents our estimate of the bonus program's impact does not change much; coefficients on main effects and two-way interaction terms are relatively unaffected as well.

As reported above, base salaries for teachers in North Carolina vary from just under \$30 000 to over \$60 000. On average, an \$1800 bonus represents a 4 to 5% increase in pay. The estimated reduction in turnover rates of 17% implies that the elasticity of turnover with respect to salary is on the order of –3 to –4. This estimate is larger than others in the existing literature. Notably, Hanushek, Kain, and Rivkin (2004) report that a 10% increase in teacher pay reduces the probability of departure from a school by one to four percentage points. Converted to an elasticity, this effect would be more on the order of –1. Dolton and van der Klaauw (1995), using British data and a measure of relative rather than absolute earnings, also arrive at an elasticity estimate on the order of –1. Murnane and Olsen (1989) present estimates that are not directly comparable, but appear to be consistent with an elasticity of similar magnitude.²³ Several factors

²² Estimates employing our alternative proxy measure of teacher eligibility, rather than official records indicating bonus receipt, continue to be statistically indistinguishable from one. Estimates using the dollar value of the bonus, rather than a binary measure of receipt, imply that the effect of receiving the full \$1800 is a 17% reduction in the hazard rate, while the impact of a prorated \$900 reduction would be 9%. This in turn suggests a rough rule of thumb that a \$100 increase in the bonus reduces the probability of departure by approximately 1% — not one percentage point, but 1%.

²³ Murnane and Olsen report that a 15% increase in earnings from the mean is associated with an increase in median spell duration of roughly 4 years. Given summary statistics on turnover rates provided in the paper, this increase in median duration could be accomplished with a 15% reduction in turnover rates.

Table 6
Robustness checks

	All schools	Ever eligible schools only	Using only left-censored spells	Omitting left-censored spells	Weibull model: all spells	Weibull model: omitting left-censored spells
Teacher receives a bonus payment	0.867** (0.054)	0.779** (0.060)	0.880 (0.108)	0.809** (0.067)	0.783** (0.055)	0.746** (0.064)
Teacher is certified in Math, Science or Special Education and is employed by an ever eligible school	1.067 (0.049)	–	1.020 (0.084)	1.069 (0.103)	1.049 (0.066)	1.137 (0.111)
Teacher is employed by a currently eligible school	0.781** (0.029)	0.918 (0.049)	0.872* (0.066)	0.793** (0.043)	0.795** (0.043)	0.740** (0.040)
Teacher is certified in Math, Science or Special Education in a post-program year	1.036 (0.028)	1.279** (0.120)	1.088 (0.102)	1.171 (0.130)	1.072 (0.069)	1.168 (0.131)
Teacher is certified in Math, Science or Special Education	0.899** (0.019)	0.892* (0.060)	0.916 (0.070)	0.852 (0.100)	0.806** (0.050)	0.765** (0.091)
Teacher is employed by an ever eligible school	1.179** (0.034)	–	1.228** (0.053)	1.320** (0.066)	1.319** (0.043)	1.340** (0.067)
Year is 2000	1.337** (0.025)	1.292** (0.086)	–	–	7.136** (0.393)	–
Year is 2001	1.834** (0.034)	1.694** (0.111)	–	1.808** (0.110)	3.128** (0.138)	6.762** (0.466)
Year is 2002	1.967** (0.034)	1.830** (0.100)	–	1.958** (0.088)	2.413** (0.086)	3.807** (0.176)
Demographic controls and district fixed effects	Yes	No	No	No	No	No
Duration dependence parameter	–	–	–	–	1.991 (0.025)	2.273 (0.035)
<i>N</i>	211,330	14,890	17,456	12,106	29,562	12,106
Log likelihood	–403,813.39	–30,216.29	–29,655.55	–25,211.12	–11,969.98	–5160.48

Note: Table entries are hazard ratios, with standard errors in parentheses. Models are estimated using the Cox proportional hazard model, except where indicated. The hazard refers to the probability of exiting a school after period t , conditional on remaining in that school until period t . Unit of observation is the teacher/school/year. ** denotes a hazard ratio significantly different from 1 at the 5% level; * the 10% level.

can potentially explain our larger elasticity estimate. First, we exploit within-school variation in teacher salaries, breaking the correlation between salaries and unobserved working conditions that plagues many existing studies. Second, Hanushek, Kain, and Rivkin exclude within-district moves from their definition of turnover, and within-district salary differentials may be particularly effective in reducing the likelihood of this type of move. Dolton and van der Klaauw consider only exits from the teaching profession, as do Murnane and Olsen. Finally, the Hanushek, Kain and Rivkin estimate is derived from a linear probability model, rather than a hazard model.

3.3. Robustness checks

Table 6 presents the results of a number of alternative specifications which assess the sensitivity of our estimate to variations in empirical strategy. The first two columns present estimates that alter the set of schools included in the analysis. First, we expand the set of ineligible schools incorporated in the analysis to include all secondary schools in North Carolina. To control for potentially important differences between ever-eligible schools and others, we add a set of district fixed effects to the analysis, as well as demographic controls for percent black and percent receiving free or reduced lunch in the school, as well as interactions of these variables with an indicator for whether the school is a high school.

In this specification, the estimated impact of the bonus program is statistically significant, though of slightly smaller magnitude than our final baseline estimate: 13% rather than 18%. From a statistical perspective, the difference in point estimates across models is small enough to be indistinguishable.

In the second column of Table 6, we restrict the sample of schools to the “ever-eligible” group, identifying the impact of the bonus program through a simpler difference-in-difference estimator exploiting the program’s discrete (and time-varying) implementation and the fact that not all teachers received the bonus in each school. In this case, the program’s estimated impact is of a slightly larger magnitude: teachers receiving bonus payments are 22% less likely to depart at the end of the school year. Overall, then, the issue of control group selection appears to have some impact on the magnitude of the bonus program’s estimated impact, but does not alter the fundamental conclusion that teachers receiving the bonus payment were more likely to remain employed at their current school.

Table 7
Testing for differential effects by type of teacher or school

	(1)	(2)	(3)
Paid bonus, teach Math	0.809** (0.068)	—	—
Paid bonus, teach Science	0.925 (0.083)	—	—
Paid bonus, teach Special Education	1.140 (0.153)	—	—
Paid bonus, teach 1 year	—	0.873 (0.128)	—
Paid bonus, teach 2 years	—	0.794 (0.159)	—
Paid bonus, teach 3 years	—	0.854 (0.174)	—
Paid bonus, teach 4–9 years	—	0.848 (0.098)	—
Paid bonus, teach 10–19 years	—	0.690** (0.093)	—
Paid bonus, teach 20–29 years	—	0.794 (0.124)	—
Paid bonus, teach 30–39 years	—	0.864 (0.150)	—
Paid bonus, teach 40 years or more	—	0.267 (0.275)	—
Paid bonus, teach in a middle school	—	—	0.812* (0.088)
Paid bonus, teach in a high school	—	—	0.833** (0.062)
<i>N</i>	29,562	29,562	29,562
Log likelihood	–58,963.83	–58,960.72	–58,963.77

Notes: Table entries are hazard ratios, with standard errors in parentheses. Unit of observation is the teacher/school/year. All specifications include controls for being certified in Math, Science or Special Education, teaching at an eligible school, subject taught, gender, race, experience, education and high school (with the same omissions for categorical variables as listed in Table 4).

** denotes a hazard ratio significantly different from 1 at the 5% level; * the 10% level.

The third and fourth specifications in Table 6 examine the potential importance of spell left-censoring in our data. In theory, the use of the Cox proportional hazard model assuages many concerns regarding left-censoring, since the impact of covariates is presumed to be independent of time. In practice, we are concerned that the bonus program may have had different proportional impacts on spells of different durations. We provide further evidence to this point in Table 7 below. These estimates separate the sample into two groups: spells that are left-censored because they began during or before the first year of observation, as well as uncensored spells. Contrasts between results derived using these two samples may reveal information regarding the overall impact of left-censoring on our estimates of the bonus program's impact.²⁴

Among the set of teachers with spells underway in the initial year, the estimated impact of the bonus program is of relatively small magnitude (12%) and imprecisely estimated.²⁵ Restricting our attention to uncensored spells, by contrast, reveals a more precisely estimated hazard ratio indicating a 19% reduction in conditional departure rates for teachers who received a bonus. At face value, the contrast between these estimates suggests two potential interpretations. One is that the net impact of spell left-censoring is to reduce the magnitude of our coefficient estimates. The second is that the bonus program had a stronger impact among those teachers recruited to work in eligible schools either during or immediately before the program's implementation. In Table 7 below, we provide evidence that seems to contradict this second implication — the bonus program actually appears to have the strongest proportional impact on the turnover behavior of more experienced teachers.

As stated in Section 2 above, the Cox proportional hazard is problematic in its treatment of spells of identical duration, of which there are many in our analysis sample.²⁶ To assess whether this problem unduly influences our results, the final columns in Table 6 report estimates of Weibull proportional hazard models, which avoid the problem of tied durations at the cost of requiring parametric specification of the hazard function. The fifth column presents an estimate including all spells, treating left-censored observations as though they began in the first year of the dataset. The final column excludes these left-censored spells. In both specifications, the estimated impact of the bonus program is statistically significant and larger in magnitude relative to the estimates utilizing the Cox model. When left-censored

²⁴ There are other reasons to expect potentially differential bonus program impacts in these two samples. For example, the bonus program may have had some impact on the set of teachers recruited into eligible positions in eligible schools. As noted above, the structure of the program made it a poor tool for recruitment, a fact confirmed in our interviews with principals and other administrative personnel.

²⁵ In the Cox proportional hazard model, year effects can only be identified by comparing individuals with spells of varying observed length in a specific year. We are thus unable to identify year effects in the model including only those individuals with spells underway as of 1999/2000.

²⁶ All of the Cox proportional hazard models estimated in this paper utilize the Breslow method for ties.

spells are excluded, teachers receiving bonus payments are estimated to be 25% less likely to depart at the end of a school year.²⁷

3.4. Variation in program impacts

Although the mean impact of the program is an important measure, it is also helpful to know whether the program's effects were selective in the sense of decreasing the likelihood of departure more for some types of teachers than for others. We focus our attention here on the potential differential impacts of the bonus program among teachers who differ along easily observable dimensions, as well as differential impacts by type of school. In principle, it would also be desirable to differentiate the bonus program's impact on teachers who exhibited varying amounts of success in raising student test scores. Such an estimation is unfortunately infeasible at the current time, primarily because a large portion of teachers in the sample cannot be reliably linked to a complete set of test scores for students they instructed.²⁸

We estimate the differential effects by interacting the various characteristics of interest with whether the teacher received a bonus in the particular year. The results are shown in Table 7. The first specification examines how the response to the program differed by subject taught. A clear significant pattern of selective impacts appears. Math teachers receiving the bonus were 19% less likely to depart the following year when they received the bonus.²⁹ Point estimates suggest a weaker, statistically insignificant response among teachers of science, and no impact on teachers of special education. This finding is consistent with basic evidence on conditional turnover rates shown in Table 1. Averaged across two pre-program and two post-program years, conditional turnover rates for math teachers were nearly equal in treatment and control schools, differing by only one percentage point. The treatment–control differences in turnover rates for science and special education teachers was higher, at 4 and 5.5 percentage points, respectively.³⁰

The second column in Table 7 presents an intriguing pattern of differences in the program impact by the experience level of the teacher. Point estimates suggest that the bonus program had the largest relative impacts on relatively experienced teachers. Among teachers with 10 to 19 years of experience, the estimated reduction in hazard rates is a statistically significant 31%.³¹ Large but imprecise effect estimates are also associated with other high experience categories.³² The larger relative effects on experienced teachers may reflect the fact that the bonus program was not effective at preventing teachers from leaving the profession altogether. Younger teachers are much more likely to exit a job because they are exiting the profession. Experienced teachers, by comparison, are very likely to take another teaching job when they exit a position, and the bonus program may have been most effective in forestalling this sort of mobility.³³

²⁷ The duration dependence parameters reported at the bottom of Table 6 indicate positive duration dependence, which is somewhat surprising given evidence of generally declining conditional turnover rates in Table 1. Note, however, that these estimates incorporate year fixed effects, which complicate the interpretation of the duration dependence parameters. In all specifications, each spell is of duration four or less, which further complicates inference regarding the specific functional form of the underlying hazard. For purposes of this analysis, specification of the underlying hazard is less important than identifying the regression parameters of interest.

²⁸ Our inability to reliably estimate the effectiveness of secondary math, science, and special education teachers has five root causes. First, in middle school, students take end-of-grade tests which are frequently administered by home room teachers, rather than teachers who actually taught the subject being tested. The test score data include information only on the teacher who administered the test, implying that we cannot estimate any effectiveness models whatsoever using middle school teachers. Second, in high school, where students take end-of-course (EOC) tests in a selected set of subjects, we have been able to match these test score records to teachers in only 77% of all cases. Third, not all math or science courses are tested (for example, there are no EOC tests in calculus or statistics), and there are no standardized tests in special education. Thus many of our teachers can be linked to no test score results at all. Fourth, to estimate teacher effectiveness using a value-added type specification, it is necessary to control for a prior test score, and roughly 10% of all high school students lack a prior score in the database. Finally, there are multidirectional concerns regarding selection into test-taking and test-administering. The bonus program may have influenced the set of students who choose to enroll in tested courses, and may have also influenced administrative decisions regarding which teachers to assign to tested versus untested courses.

²⁹ Estimates from an alternative specification relating the amount of the bonus received to the conditional probability of departure suggest that math teachers receiving the full \$1800 bonus were 24% less likely to leave.

³⁰ A possible explanation for this pattern is that math teachers perceived the eligibility criteria for the bonus program more accurately, and thus considered it to be more valuable than their counterparts in other subjects. We present evidence on the role of information quality as a moderator of program impact below.

³¹ Estimates from an alternative specification relating the amount of the bonus received to the conditional probability of departure suggest that teachers with 10 to 19 years of experience receiving the full \$1800 bonus were 43% less likely to leave.

³² Note that confidence intervals for hazard ratios are asymmetric; the non-significance on the last coefficient reported in this specification is not a typographical error.

³³ A complete test of this hypothesis would require the estimation of competing risk models, which we leave to further research.

It should be noted that these large relative impacts on hazard rates translate into somewhat more modest absolute impacts on turnover, since experienced teachers generally exhibit lower baseline hazards. Table 1 shows that teachers with 10–19 years of experience had average turnover rates of 22% in treatment schools; a 30% reduction in this value would be a six percentage point decline in turnover. The bonus program appears to have had a much smaller relative impact on inexperienced teachers, but with absolute turnover rates as much as twice the magnitude of their experienced counterparts, the absolute impacts appear to be quite similar.³⁴ Nonetheless, given previous findings indicating greater classroom effectiveness of experienced teachers relative to novice teachers (Clotfelter et al., 2006; Rivkin et al., 2005), and the greater propensity for novice teachers to serve disadvantaged students (Clotfelter et al., 2005), this pattern shows the importance of further investigation into the potential role of monetary rewards as a tool for retaining experienced teachers in disadvantaged schools.³⁵

The final set of interaction results sheds light on the differential impact of the bonus program by type of school.³⁶ The results suggest that the bonus program was slightly more effective in middle schools relative to high schools, but the difference in coefficients is statistically indistinguishable from zero.

4. The potential role of information

Following the discussion in Section 1 above, we conducted additional hazard models that introduced the bonus eligibility criterion variable – the same factor used to select control group schools in our baseline analysis – and interacted that variable with the teacher-level indicator of bonus eligibility. Eligible schools with lower levels of the criterion variable were more likely to revert to ineligible status in a subsequent year. Teachers who fully understood the structure of the bonus program would have realized that this was of no concern to them. If all teachers had understood the operation of the bonus program, therefore, there would be no reason to expect the program to have been less effective at marginally eligible schools. If we consider working conditions in a school to vary inversely with the criterion variable, then it actually would be reasonable to expect a larger impact in marginally eligible schools. For example, suppose working conditions in a marginally eligible school can be described as a scalar value a , and in unambiguously eligible schools as a value $b < a$. Given an equal bonus amount in the two schools, it is reasonable to think that a larger number of teachers would consider the bonus adequate compensation for accepting conditions a rather than b .

If, on the other hand, teachers widely misperceived the eligibility criteria, then the program's impact on retention rates might well be lower at marginally eligible schools. If this were the case, estimates of the program's impact based on schools that were markedly above the eligibility threshold might be better indicators of what a fully comprehended bonus program might accomplish.

Column (1) in Table 8 presents a hazard model that introduces the criterion variable alone. The criterion variable has been normalized so that a value of zero corresponds to a school where the long-run average value equals the eligibility threshold. For middle schools, this is a subsidized lunch receipt rate of 85%. For high schools, it is a combined failure rate on biology and algebra exams of 100% (on a scale from 0 to 200%). This is a relatively crude method of normalization, but provides a relatively straightforward interpretation.

In this specification, the estimated mean impact of the bonus program is very similar to previous estimates, a 14% reduction in turnover rates. The criterion variable enters significantly and positively, consistent with the presumption that higher values of the criterion variable correspond with less favorable working conditions at schools. The point estimate indicates that middle schools with 90% subsidized lunch rate exhibits turnover rates 2 percentage points higher

³⁴ This pattern can potentially explain why estimates of the bonus program's impact are larger in models that exclude left-censored observations. In models that include these spells, the bonus program appears to have a more modest relative impact because of its smaller absolute impact on highly experienced teachers, who are in many cases indistinguishable from teachers with as few as four years of experience.

³⁵ An alternative explanation for this pattern is that experienced teachers may have been more likely to understand the eligibility criteria, either because of their greater experience with public school bureaucracy or because they tended to teach in schools where the information flow was more reliable. Our survey evidence suggests that experienced teachers were more likely to be aware of the program (Clotfelter et al., in press).

³⁶ In additional unreported specifications, we tested for differential impacts of the bonus program by teacher race, gender, and education level. Point estimates suggest that the program had its largest impact on white teachers, female teachers, and teachers with masters degrees. In no case are we able to reject the null hypothesis of equal impact across categories. We also attempted to test whether the bonus program had a stronger impact in its first or second year of implementation. The results suggest that there was a stronger impact in the first year of implementation, consistent with the basic evidence in Fig. 3. The program's reduced impact in subsequent years could reflect teacher misunderstandings: teachers in schools that lost their eligibility may have incorrectly thought that their bonus payments had ceased. Teachers may have also correctly anticipated that the bonus program would be rescinded.

Table 8
The role of information quality

	(1)	(2)
Teacher receives a bonus payment	0.838** (0.057)	0.858* (0.068)
Teacher is certified in Math, Science or Special Education and is employed by a potentially eligible school	1.025 (0.064)	1.025 (0.064)
Teacher is employed by a currently eligible school	0.769** (0.033)	0.769** (0.033)
Teacher is certified in Math, Science or Special Education in a post-program year	1.113* (0.070)	1.112* (0.070)
Teacher is certified in Math, Science or Special Education	0.921 (0.057)	0.921 (0.057)
Teacher is employed by a potentially eligible school	1.167** (0.042)	1.164** (0.042)
Year is 2000	1.210** (0.055)	1.210** (0.055)
Year is 2001	1.640** (0.074)	1.640** (0.074)
Year is 2002	1.942** (0.077)	1.943** (0.077)
Masters	1.133** (0.033)	1.134** (0.033)
Other graduate degree	1.198** (0.110)	1.197** (0.110)
1 year of experience	1.466** (0.067)	1.466** (0.067)
2 years of experience	1.260** (0.067)	1.260** (0.067)
3 years of experience	1.118** (0.063)	1.118** (0.068)
4–9 years of experience	0.965 (0.034)	0.945 (0.034)
20–29 years of experience	0.729** (0.029)	0.729** (0.029)
30–39 years of experience	1.260** (0.055)	1.260** (0.062)
40+ years of experience	1.678** (0.398)	1.678** (0.398)
High school	0.826** (0.028)	0.826** (0.028)
Criterion variable	1.579** (0.121)	1.579** (0.121)
Paid bonus* criterion variable	–	0.884 (0.198)
Subject area controls	Y	Y
Gender and race controls	Y	Y
N	29,562	29,562
Log likelihood	–58,946.42	–58,946.27

Note: Table entries are hazard ratios, with standard errors in parentheses. Unit of observation is the teacher/school/year. ** denotes a hazard ratio significantly different from 1 at the 5% level; * the 10% level.

than schools with 85% subsidized lunch rates, or that high schools where the failure rates on algebra and biology exams are 55% each have turnover rates 4 percentage points higher than in schools where the failure rates are 50% each.

In column (2), we introduce an interaction between bonus receipt and the criterion variable. The main effect of bonus receipt, which now represents the predicted impact in schools where the average criterion variable equals the eligibility threshold, is 14%, and statistically significant at the 5% level. The interacted effect is estimated imprecisely, with a point estimate clearly but insignificantly below one. Taken at face value, this point estimate implies that the impact of the bonus program on turnover rates increased by six-tenths of 1% for every five percentage point increase in the criterion variable.

How large might the bonus program's impact have been if all teachers had correctly perceived the eligibility criteria? Without formally modeling the process of teacher's expectation formation, any estimate we provide would be little more than an educated guess, particularly given the imprecision of the estimates in Table 8. But it is reasonable to presume that our basic estimates of the bonus program's impact understate the potential impact of an alternative program that convincingly promised teachers permanent salary differentials in exchange for working at high-poverty or low-performing schools.

5. Concluding discussion

From one perspective, the effectiveness of the bonus program can be measured in terms of reductions in teacher turnover. We associate the bonus program with a one-sixth reduction in turnover rates, or roughly from 30% to 25%. This suggests that the program spent \$36,000 for every teacher departure averted or delayed.

The ultimate goal of the N.C. Bonus Program, however, was to improve the quality of math, science, and special education instruction for students in disadvantaged schools. The fact that the program appears to have reduced departure rates of teachers from the schools serving disadvantaged and low-performing students means that the program could potentially have raised student achievement; in practice the program's brief duration hampers any

efforts to directly estimate its impact.³⁷ To estimate the potential impact on student achievement, and provide a rough estimate as to the program's overall cost-effectiveness, we conducted a simple simulation. We simulated the process of teacher departure and arrival for two samples of 1000 hypothetical teachers, selecting parameters to match the observed experience distribution shown in Table 3 and the conditional turnover rates reported in Table 1. In the second sample, we multiplied the turnover rates by a factor equal to our preferred estimate of the bonus program's impact, which is the value of 0.827 in the last column of Table 5.

The repeated simulation suggests that the bonus program would have led to a steady-state increase in the average experience level at targeted schools by one-half of one year.³⁸ The translation of this impact on experience levels into an impact on test scores is complicated by the fact that evidence – most of it derived from studies on primary rather than secondary schools – shows nonlinear returns to experience. If we presume that one extra year of experience raises student test scores by 1% of a standard deviation, which is not unreasonable given estimates in the literature, and that the typical secondary school teacher instructs 100 students over the course of a year, then the bonus program spent \$36 to raise one student's test score in one subject by 1% of a standard deviation.³⁹ Of course, unless the bonus program had an impact on the overall distribution of teacher quality and experience, this benefit to a student in a low-performing school would be offset by a cost to a student in a different school. Effects on entry into teaching will in all likelihood increase with the overall magnitude of a salary differential program. Even in the absence of such effects, though, policy makers may very well attach greater value to improve the test scores of disadvantaged students.

Despite its premature demise, the North Carolina Bonus Program provides an important example of a reform that deserves more attention. Though we were able to examine only one outcome of the program in this study, the fact that the program generated positive effects, despite the flaws in its implementation which quite possibly diluted its impact, suggests that a policy of salary differentials for teachers in low-performing or high-poverty schools, particularly differentials that were credibly advertised as permanent, could be a cost-effective means of improving the academic achievement of disadvantaged students.

References

- Ballou, Dale, Podgursky, Michael, 1995. Recruiting smarter teachers. *The Journal of Human Resources* 30, 326–338 (Spring).
- Clotfelter, Charles T., Ladd, Helen F., Vigdor, Jacob L., 2005. Who teaches whom? Race and the distribution of novice teachers. *Economics of Education Review* 24, 377–392.
- Clotfelter, Charles T., Glennie, Elizabeth, Ladd, Helen F., Vigdor, Jacob L., in press. "Teacher Bonuses and Teacher Retention in Low Performing Schools: Evidence from the North Carolina \$1,800 Teacher Bonus Program." *Public Finance Review*.
- Clotfelter, Charles T., Ladd, Helen F., Vigdor, Jacob L., Wheeler, Justin, 2007. High Poverty Schools and the Distribution of Teachers and Principals. *North Carolina Law Review* 85, 1345–1379.
- Clotfelter, Charles T., Ladd, Helen F., Vigdor, Jacob L., 2006. Teacher Sorting, Teacher Shopping, and the Assessment of Teacher Effectiveness. *Journal of Human Resources* 41, 778–820.
- Cocoran, Sean P., Evans, William N., Schwab, Robert M., 2004. Changing labor-market opportunities for women and the quality of teachers, 1957–2000. *American Economic Review Papers and Proceedings* 230–235 (May).
- Dolton, P., van der Klaauw, W., 1995. Leaving teaching in the UK: a duration analysis. *The Economic Journal* 105, 431–444.
- Fowler, R. Clarke, 2003. "The Massachusetts Signing Bonus Program for New Teachers: A Model of Teacher Preparation Worth Copying?" *Education Policy Analysis Archives* 11 (April 22). Retrieved 10/4/04 from <http://epaa.asu.edu/epaa/v11n13/>.
- Hanushek, Eric A., 1999. Some findings from an independent investigation of the Tennessee STAR experiment and from other investigations of class size effects. *Educational Evaluation and Policy Analysis* 21, 143–163 (Summer).
- Hanushek, Eric A., Kain, John F., Rivkin, Steven G., 1999. Do Higher Salaries Buy Better Teachers? NBER Working Paper #7082.
- Hanushek, Eric A., Kain, John F., Rivkin, Steven G., 2004. Why public schools lose teachers. *Journal of Human Resources* 39, 326–354 (Spring).
- Hoxby, Caroline M., Leigh, Andrew, 2004. Pulled away or pushed out? Explaining the decline of teacher aptitude in the United States. *American Economic Review* 94, 236–240 (May).

³⁷ The simulation described below suggests that the complete steady-state impact of the bonus program on the distribution of teacher experience would fully appear after a period of five or more years.

³⁸ Over a ten-year period beginning in the 4th year after implementation, the simulation suggests that the bonus program would raise mean years of experience to 13.71 from 13.20.

³⁹ For purposes of comparison, Krueger (1999), in an analysis of the Tennessee STAR experiment, reports that student assignment to smaller class sizes in Kindergarten raises test scores by roughly 20% of a standard deviation. Accepting this estimate (which not all observers do — see Hanushek, 1999), and using \$35,000 as an estimate of the cost of hiring an additional teacher, assigning 45 students to 3 rather than 2 classrooms would result in the expenditure of \$39 for every 1% of a standard deviation in a student's test score. While in theory this gain comes with no concomitant decrease in the test scores of other students, in practice a large increase in demand for teachers may well lead to reductions in the effectiveness of the marginal teacher (Reed and Rueben, 2003).

- Krueger, Alan B., 1999. Experimental estimates of education production functions. *Quarterly Journal of Economics* 114, 497–532.
- Lankford, Hamilton, Leub, Susanna, Wyckoff, James, 2002. Teacher sorting and the plight of urban schools: a descriptive analysis. *Educational Evaluation and Policy Analysis* 24, 37–62 (Spring).
- Loeb, Susanna, Page, Marianne E., 2000. Examining the link between teacher wages and student outcomes: the importance of alternative labor market opportunities and non-pecuniary variation. *The Review of Economics and Statistics* 82, 393–408 (August).
- Murnane, R.J., Olsen, R.J., 1989. The effect of salaries and opportunity costs on duration in teaching: evidence from Michigan. *Review of Economics and Statistics* 71, 347–352.
- Reed, Deborah, Rueben, Kim S., 2003. Teacher recruitment and retention following class size reduction in California. Urban Institute manuscript.
- Reed, Deborah, Rueben, Kim S., Barbour, Elisa, 2006. Retention of New Teachers in California. Public Policy Institute of California, San Francisco. Retrieved 4/14/06 from www.ppic.org/content/pubs/R_206DRR.pdf.
- Rivkin, Steven G., Hanushek, Eric A., Kain, John F., 2005. Teachers, schools, and academic achievement. *Econometrica* 79, 417–458 (March).
- Rockoff, Jonah E., 2004. The impact of individual teachers on student achievement: evidence from panel data. *American Economic Review Papers and Proceedings* 247–252 (May).
- Scafidi, Benjamin, Sjoquist, David, Stinebrickner, Todd R. “The Impact of Wages and School Characteristics on Teacher Mobility and Retention,” Unpublished paper, September 2002.