



Working Paper:

Performance Screens for School Improvement: The Case of Teacher Tenure Reform in New York City

Susanna Loeb¹, Luke C. Miller² & James Wyckoff²

Tenure is intended to protect teachers with demonstrated teaching skills against arbitrary or capricious dismissal. Critics of typical tenure processes argue that tenure assessments are superficial and rarely discern whether teachers in fact have the requisite teaching skills. A recent reform of the tenure process in New York City provides an unusual opportunity to learn about the role of tenure in teachers' career outcomes. We find the reform led to many fewer teachers receiving tenure. Those not receiving tenure typically had their probationary periods extended to allow them an opportunity to demonstrate teaching effectiveness. These "extended" teachers were much more likely to leave their schools and be replaced by a teacher who was judged to be more effective.

¹Stanford University

²University of Virginia

Updated September 2014

EdPolicyWorks
University of Virginia
PO Box 400879
Charlottesville, VA 22904

EdPolicyWorks working papers are available for comment and discussion only. They have not been peer-reviewed.

Do not cite or quote without author permission. Working paper retrieved from:

http://curry.virginia.edu/uploads/resourceLibrary/31_Teacher_Tenure_Reform_in_NYC.pdf

Acknowledgements: Thanks to Joanna Cannon, Anne-Marie Hoxie and Keely Alexander at the New York City Department of Education for providing the data employed in this paper and for answering questions about the NYCDOE tenure policy. We appreciate financial support from the National Center for the Analysis of Longitudinal Data in Education Research (CALDER). CALDER is supported by IES Grant R305A060018. The views expressed in the paper are solely those of the authors. Any errors are attributable to the authors.

EdPolicyWorks Working Paper Series No. 31. September 2014.

Available at <http://curry.virginia.edu/edpolicyworks/wp>

Curry School of Education | Frank Batten School of Leadership and Public Policy | University of Virginia

**PERFORMANCE SCREENS FOR SCHOOL IMPROVEMENT:
THE CASE OF TEACHER TENURE REFORM IN NEW YORK CITY**

Susanna Loeb, Luke C. Miller & James Wyckoff

Introduction

Teacher tenure has been controversial since the first tenure provisions were enacted over a century ago. Proponents typically argue that tenure prevents teacher dismissal for political purposes or due to capricious decisions by administrators or politicians. Tenure could guard against dismissal of more experienced, higher paid teachers during periods of tight budgets when school leaders may be more focused on reducing costs while meeting class size requirements than they are on student learning. Tenure does not require schools or districts to retain ineffective teachers but instead provides a due process mechanism to dismiss tenured teachers for cause. Critics, however, argue that the cost of due process does, in practice, lead districts to retain ineffective teachers and as a result tenure not only allows poor teachers to stay in the classroom but also reduces the incentive for teachers to be as effective as they could be. They argue that the due process mechanisms for removing teachers with tenure are so burdensome that they rarely are pursued.

With the availability of large-scale student performance measures linked over time has come clear evidence that teachers vary substantially in their effectiveness at improving student test performance and that these differences can have meaningful effects on students in both the short run and the long run (Chetty, Friedman, & Rockoff, 2012; Rivkin, Hanushek and Kain, 2005; Rockoff, 2004). At least partially as a result, education reforms in the US recently are focusing on improving the quality of teaching through human resource policies such as improved evaluation systems and differentiated pay. Given the controversial nature of teacher tenure, it is not surprising that interest also has increased in changing teacher tenure provisions so that the due process is less onerous and so that school leaders have greater control over their workforce. Yet, the evidence on which to base reform decisions is scarce. We know little about what types of tenure provisions improve the quality of teaching and what types do not. Similarly, we know little about how long the probationary period prior to tenure should be, if there is tenure, in order for school systems to accurately assess teachers' effectiveness so that they can make well informed decisions about tenure.

Part of the reason that we have little evidence on the effects of tenure is that until recently tenure laws have been relatively stable over time and similar, though not the same, across states.

New Jersey passed the nation's first teacher tenure law in 1909. Over the next several decades other states adopted similar laws: New York in 1917, California in 1921, and Michigan, Pennsylvania, and Wisconsin in 1937. The state statutes used a variety of synonyms for tenure: continuing contract or service, permanent status, career status, and post-probationary status. Regardless of the terminology, these laws have three main components: tenure requirements, reasons for dismissal, and process for appeals. The first specifies the length of the probationary period after which teachers are eligible for tenure. Employers can dismiss a non-tenured teacher at any time for any reason so long as the decision is neither arbitrary or capricious nor discriminatory, but tenured teachers can only be dismissed for the reasons provided in the law. The third component details the appeals process a dismissed tenure teacher can pursue in an effort to be reinstated. Of the 48 states in which public elementary and secondary teachers are awarded tenure, the minimum probationary period exceeds three years in 11 states (National Council on Teacher Quality, 2012). In most states it is three years, although in a few states, such as California, teachers typically receive tenure with fewer than three years of experience.

The last decade has seen substantial change in tenure laws in the US. In 2000, Georgia eliminated due process rights for teachers hired after 1 July 2000, but reinstated these rights three years later. Florida eliminated teacher tenure in 2011. That same year Idaho enacted a law that would have eliminated teacher tenure had it not been repealed by voters the following year. Voters in South Dakota turned back an effort to repeal a 2012 law thereby allowing a law eliminating tenure after 1 July 2016 to take effect. North Carolina's governor signed a bill into law last year that eliminates teacher tenure by 2018. Most recently, the court ruling in the Vergara case has set California on a similar course pending any appeal. Though almost all states currently grant tenure, more than half now require meaningful evaluation during the tenure process. As an example, in 2009 only four states used student test performance as a criterion for tenure; by 2012, 20 states did and 25 states require multiple categories for teachers in their evaluation, not just satisfactory and unsatisfactory (National Council on Teacher Quality, 2012).

A recent reform by the NYCDOE provides an unusual opportunity to learn about the role of tenure in teachers' career outcomes including both strategic retention on the district side and choice-based retention stemming from teachers' decisions. In what follows, we start by describing the reform. We then use data from NYCDOE and the New York State Education Department (NYSED) to provide initial evidence on the magnitude of responses to the reform, concluding with

a discussion of the results. We find that the NYCDOE tenure reforms extended the probationary periods of teachers judged to be less effective and that these “extended” teachers were substantially more likely than other teachers to leave their schools. Their likely replacements were typically more effective as judged by their principals and as measured by value-added. These changes disproportionately benefited schools with larger percentages of black students.

The Teacher Tenure Process in New York City

The criterion for tenure in New York City is that a teacher possesses “significant professional skill and a meaningful, positive impact on student learning.” This criterion is not new. However, prior to AY 2009-10 the tenure process in New York City was similar to that in many other large urban districts. The receipt of tenure had become an expectation for nearly all teachers and frequently was based on little evidence of accomplishment. In 2007-08 and 2008-09, well into the period of accountability reforms, 94 percent of all eligible teachers were approved for tenure.

Beginning in 2009-10, New York City changed the tenure review process, infusing more information and increasing the responsibility and accountability of principals to insure that teachers met challenging performance standards. Tenure decisions in 2009-10 were informed by sources of information that had been available previously: classroom observations, evaluations of teacher work products including lesson plans, and the annual rating sheet that principals completed giving teachers a Satisfactory, Doubtful, or Unsatisfactory rating. In addition, tenure decisions in 2009-10 included new student learning measures from the Teacher Data Reports (which included teacher value-added), in-class assessments aligned with the New York State standards, and other evidence of student progress (NYCDOE, 2009).

As in previous years, principals sent recommendations to the superintendent about whether a teacher should be denied tenure, have their probationary period extended or be granted tenure, but starting in 2009-10 principals had to provide a rationale for this decision if the evidence available at the district level suggested either a strong case to approve or deny tenure and this information ran counter to the principal’s recommendation. The district provided principals with tenure guidance for teachers for whom there was evidence that performance was particularly strong or weak. For a teacher whose value-added results had been in the lowest 50 percent over the past two years (with a 95 percent confidence interval), who had previously received an Unsatisfactory annual rating, or whose tenure decision had previously been extended, the principal received guidance from the

district that the teacher should be considered to have “tenure in doubt”. A principal recommendation to extend or approve tenure for these teachers required a supporting rationale for the superintendent to consider in his or her review. The principal received guidance of “tenure likely” for a teacher whose value-added results had been in the highest 50 percent over the past two years (with a 95 percent confidence interval). Principals recommending denying tenure or extending the probationary period for these teachers similarly needed to provide supporting evidence to the superintendent (NYCDOE, 2009).

The process introduced in 2009-10 remained in place in 2010-11 with some notable changes (NYCDOE, 2010). New in 2010-11, principals were asked to evaluate all teachers up for a tenure decision based a four-point effectiveness rating scale (Highly Effective, Effective, Developing and Ineffective) as described in the district-developed Effectiveness Framework.¹ As in the prior year, the evidence for these ratings came from measures of the teacher’s impact on student learning such as value-added measures from the Teacher Data Reports, student work products and tests aligned to the New York State standards. Principals also could use evidence from measures of instructional practice coming from their own classroom observations, teacher work products, and the annual rating sheet that principals complete for each teacher.² In addition to these sources of information, which were available in the prior year as well, principals in 2010-11 gained information about professional contributions from surveys of students and parents, from measures of attendance, from colleague feedback, and from work products related to the Comprehensive Educational Plan for each school. In contrast to 2009-10, principals in 2010-11 no longer received “tenure likely” or “tenure in doubt” guidance from the district but rather were given flags indicating a “low value add” teacher as an “Area of Concern” and a “high value add” teacher as a “Notable Performance”. Low and high value-added scores were defined as in the previous year. Other problematic teacher behaviors flagged as Areas of Concern included: low attendance (defined as exceeding 20 days in the previous two fiscal years), an Unsatisfactory or Doubtful rating on a prior Annual Review Sheet, having been previously extended, having been previously excessed or currently in the Absent Teacher Reserve pool.

¹ These effectiveness ratings are distinct from the ratings built into the new statewide teacher evaluation system which was not implemented until 3 years later in 2013-14. Although they use the same ratings scale, both the evidence synthesized and the relative weight assigned to the evidence differs between the two.

² These sources of evidence were employed in 2009-10 tenure decisions but they were not aggregated in the effectiveness ratings.

The tenure review process for 2011-12 was very similar to that in 2010-11, but with two important changes. As before, teachers were evaluated on their impact on student learning, instructional practice and professional contributions. Principals were provided guidance as to the expected (though not required) alignment between the effectiveness ratings they determined using the Effectiveness Framework and their tenure recommendations: Highly Effective and Effective ratings were evidence in favor of granting tenure; a Developing rating, evidence for an extension; and an Ineffective rating, evidence for denying tenure. Additionally, responsibility for producing teacher value-added estimates shifted from the district to the NYSED beginning with 2010-11 and no measures were available for principals to incorporate them into their 2011-12 tenure decisions (NYCDOE, 2011).

The state-provided value-added estimates did inform principals' 2012-13 recommendations. Teachers received a growth score (0-20) that corresponded to a HEDI rating (Highly Effective, Effective, Developing, and Ineffective). No explicit guidance was provided to principals as to how to incorporate these growth ratings into their tenure recommendations. They were only told these ratings are a source of evidence for a teacher's impact on student learning.

Screening Employees to Improve Quality

How might we expect principals and teachers to respond to the New York City tenure policy changes? Consider a principal who aims to improve the quality of instruction in her school in the short run. Principals are likely to have this goal for numerous reasons, including the high rate of principal mobility in urban school districts (Beteille, Kalogrides, and Loeb, 2012), the substantial accountability pressure on principals, and the principal training and selection mechanisms in place in NYC.³ To be more specific, we assume that quality of instruction, Y_{st} , in school s , in time t , is a function of the underlying quality of the teachers, T_{st} , of some form of investment, I_{st} , that teachers make to improve their performance and the performance of their peers, as well as other aspects of the school, O_{st} .

$$Y_{st} = q(T_{st}, I_{st}, O_{st})$$

³ See Corcoran, Schwartz and Weinstein (2009) as well as descriptions of accountability in NYC at <http://schools.nyc.gov/Accountability/tools/accountability/default.htm>

A principal will choose to deny, extend, or grant tenure to teachers in order to maximize school quality. However, *a priori* it is not clear how their choice of tenure for a given teacher will affect who teaches in the school, nor how it will affect teachers' investment in their work. These dynamics depend upon how teachers respond to the decisions both about themselves and about other teachers in the school. If teachers do not change their choices in response to tenure decisions then the principal's decision is straightforward – deny teachers whom they want to leave the school and extend all other teachers. There would be no reason to recommend tenure for any teacher since this decision limits the principal's future actions and provides no benefits in the present.

However, there are reasons to believe that teachers do care about tenure. As an example, Brunner and Imazeki (2010) find that districts with longer probationary periods compensate teachers more, other things equal, indicating that teachers value shorter probationary periods and are willing to accept somewhat reduced compensation in return. This finding is consistent with evidence from other occupations that jobs with greater risk – greater physical risk, or greater risk of unemployment, or less predictable salaries – have to pay more in order to attract the same quality workers (Feinberg, 1981). These results lead us to assume that in deciding whether or not to stay in a job, teachers value job stability and thus care about whether or not they have tenure.

Teachers are also likely to value being recognized for strong performance as well as simply gaining utility from strong performance itself (Kalleberg, 1977). Evaluation systems that differentiate among employees are likely to increase the satisfaction of teachers who perform well and may increase their willingness to stay in the school. For those, on the border line of a positive review, this evaluation may also increase their interest in investing in their own improvement and the improvement of their peers. The same differentiation is likely to have negative consequences for the satisfaction of teachers who perform worse on the evaluations, perhaps decreasing their interest in staying and investing.

Finally, teachers care about the pecuniary and non-pecuniary job attributes that informed their rationale for becoming a teacher (Flyer and Rosen, 1997). Some of these characteristics – in particular, the climate of the school such as the attitudes of teachers and their willingness to support other teachers in the school – may change as a result of the implementation of tenure reform in each school, thereby serving as a mediator for the reform's impact on instructional quality. For some teachers, the benefits of being a teacher far outweigh those of alternative occupations, while for others the choice was less clear.

In shaping the quality of teaching in their schools, principals' tenure recommendations are likely informed by their perceptions of how teachers will respond to tenure outcomes. Are teachers more likely to improve or leave their school in response to being extended? Given the dramatic effect of NYCDOE tenure reform on the percentage of teachers extended, in this paper we focus on the retention decisions of extended teachers. However, the reform could well influence other teacher decisions. For example, does the reform affect whether individuals are attracted to teaching in New York or how much investment teachers make in their teaching effectiveness? Does it influence the retention decisions of teachers who are approved for tenure? Conceptually, tenure reform could influence each of these directly or through changes in the school learning environment. These decisions are potentially important but in this paper we focus solely on the effects of tenure reform on teacher retention due to data limitations.

Consider those teachers who would have received tenure pre-reform but are extended post-reform. These teachers now receive negative signals about their teaching with an increased potential of being denied tenure. Each of these increases their likelihood of leave teaching, particularly in their current school. However, for some teachers who are extended the appeal of teaching will be high enough to lead them to choose to stay in teaching. For this subgroup, an increase in investment may increase their probability of future recognition and receipt of tenure. This group may increase investment as a result of reforms if they assess that this increase will likely lead to tenure in the future. Understanding the potential effects of tenure reform on teacher behavior, principals will be more likely to extend teachers who are not performing well if they believe they can hire more effective teachers. Because school principals play a central role in the process and because the teacher workforce differs across schools, we might expect the changes to differ across schools. In keeping with these potential effects, we address the following three research question in this paper:

1. Tenure Decisions – How did tenure rates change following reform?
2. Workforce Composition – Of teachers who become eligible for tenure, how did the composition of those continuing to teach in NYC change following reform?
3. School Differences – How have schools varied in their tenure decisions and the subsequent behaviors of their teachers?

Data

In order to assess the effects of NYCDOE tenure reforms, we must accurately identify teachers eligible for tenure, as well as other teachers potentially affected by the changes. The Tenure Notification System (TNS) tracked the tenure review process for all probationary teachers in New York City public schools between 2007-08 and 2012-13. Each school year, the district made tenure decisions for teachers whose probationary period was scheduled to conclude between November 1st of the current school year and October 31st of the following school year. The probationary period for the 2009-10 cohort, for example, concluded between November 1, 2009 and October 31, 2010. The TNS provided principals with a list of teachers at their school eligible for tenure as well as all official guidance concerning each teacher's job performance prior to the current year (e.g., prior Unsatisfactory annual performance ratings, low attendance, value-added classification, etc.). Principals enter their preliminary and final ratings and recommendations into the TNS and district superintendents make and record final tenure decisions in the system.

We assembled additional information on all teachers, not just those in the TNS, from a variety of sources. NYCDOE provides basic teacher demographic characteristics, the value-added calculations for 2008-09 and 2009-10, the state's value-added calculations for 2011-12 and annual performance ratings used in the tenure review process. We identify teachers' pathways into the teaching profession from state certification records and rosters for the New York City Teaching Fellows program and Teach for America corps members in the New York City region. State certification files provide scores on certification exams. From the College Board we obtain teachers' SAT scores for those teachers who attended a New York public school from 1980 to 2008 or a New York private school from 1980 to 2001. Characteristics of the schools in which teachers teach (e.g., race/ethnicity, free/reduced-price lunch eligibility, AYP status, etc.) come from the annual state-level School Report Cards database and Institutional Master Files and the federal Common Core of Data.

Finally, leveraging data from the NYCDOE Teacher Data Initiative, we observe characteristics of the students taught by specific teachers of grades 4 through 8 mathematics and English language arts (ELA) including demographic and achievement information. We use these data to estimate our own value-added measures of teacher effectiveness using a residuals-based approach controlling for individual student, classroom, and school characteristics and described in

the technical appendix. Currently, 2010-11 is the final year for which we can calculate these estimates.

Just over three quarters of the teachers in our post-reform sample are female, approximately 18 percent are black and 17 percent are Hispanic. They have average math and verbal SAT scores of approximately 500 points each. Approximately half of the teachers entered teaching through traditional teacher preparation programs that recommended certification, while 22 percent came through the Teaching Fellows Program, the largest early-entry route serving New York City. These teachers work at schools where 44 percent of students are Hispanic students and 31 percent are black, with 67 percent eligible for subsidized lunch (see appendix Table 1).

Recall that principals complete an Annual Rating Sheet for each teacher. Just 2.3 percent of teachers in our sample received an Unsatisfactory rating and one tenth of one percent of teachers received a Doubtful rating, with the remaining 97.6 percent receiving a Satisfactory rating. On the four-point effectiveness rating scale assigned by their principals, most teachers received either a Developing (29 percent) or an Effective (41 percent) rating, while 17 percent received Highly Effective and two percent received Ineffective ratings. Principals provided no effectiveness rating for 11 percent of teachers. Eight percent of teachers had what the district classified as low attendance (more than 20 absences over prior two years), and 12 percent had low value-added.

Results

Tenure Decisions

As described in Figure 1, 94 percent of teachers were approved for tenure during AY 2007-08 and 2008-09, the two years prior to the introduction of the policy. The approval rate dropped to 89 percent in the first year of the policy (2009-10) and averaged 56 percent in the three subsequent years. Virtually all of the decrease in the tenure approval rate resulted in an increase in the percentage of teachers whose probationary periods were extended, which averaged less than 4 percent prior to the policy, but 41 percent in 2010-11 through 2012-13. The percentage of teachers denied tenure increased marginally following the introduction of the program from an average of two percent pre-policy to three percent post-policy.

Principals have played an important role in the determination of tenure decisions. Principal effectiveness ratings of teachers using the Effectiveness Framework are highly predictive of tenure outcomes under the new policy. Ninety-four percent of teachers rated Highly Effective and 83

percent of those rated Effective were approved for tenure. In contrast, less than two percent of those rated Developing and less than one percent of those rated Ineffective were approved. The vast majority (97 percent) of teachers rated Developing were extended, while the vast majority (81 percent) of those rated Ineffective were denied tenure. Given that almost all teachers were approved for tenure prior to the reform, many teachers who would have been approved prior to the reform received a different outcome under the new system.

Tenure decisions also correspond with other teacher performance measures. For teachers in tested grades and subjects, value-added estimates track tenure decisions. Teachers denied tenure have math value-added estimates that are a full standard deviation in teacher effectiveness lower than those approved for tenure. On average, extended teachers are 13 percent of a standard deviation in student achievement less effective than the average teacher and 38 percent of a standard deviation less effective than those who are approved. Value-added differences in ELA are smaller but demonstrate the same pattern. Similarly, extended teachers are far more likely to have had prior Unsatisfactory or Doubtful annual performance ratings and to have had Low Attendance than are teachers approved for tenure (see appendix Table 2).

Overall, the reforms dramatically reduced the percentage of teachers who received tenure, but because most teachers who became eligible for tenure were extended and not dismissed it is unclear a priori whether the reform meaningfully altered the workforce.

Workforce Composition

Changes in the tenure process can affect the quality of teaching by denying tenure to less effective teachers. As discussed, denied teachers had lower value-added in both math and ELA than teachers who were extended or approved. However, even under the new policies, few teachers are dismissed. Larger changes in the workforce instead may come from changes in voluntary turnover, particularly of teachers who are extended.

Extended teachers may voluntarily exit from New York City, creating vacancies which can be filled by more effective teachers. We find some evidence of this phenomenon. Extended teachers were more likely to transfer to other New York City schools and exit from New York City in the year following their decision than teachers who were approved for tenure. Ninety percent of approved teachers return to their schools, while only 75 percent of extended teachers did so.

Being extended appears to meaningfully increase the likelihood of transfers and exits as extended teachers are more likely to leave even after controlling for teacher and school

characteristics. Table 1 shows regressions with controls for the final principal effectiveness rating of the teachers and includes school fixed effects, allowing us to examine the effect of being extended on teacher transfer and exit behavior for teachers in the same schools with equivalent principal ratings. The probability of transferring increases by 9 percentage points if the teacher had been extended rather than approved. This represents a *50 percent increase* in the probability of transferring following a tenure decision. Similarly, extended teachers exit NYC at a rate that is 4 percentage points higher than approved teachers, holding other factors constant. This represents a *66 percent increase* in the probability of exiting. These results provide evidence that the new tenure process is having a substantial effect on the composition of the teaching workforce even without substantially increasing the percentage of teachers directly denied tenure.

Among extended teachers, those who remain in the same school have somewhat different measured attributes than those who transfer or exit the system. Teachers with higher academic qualifications, such as teacher certification exam scores, are less likely to stay in the same school than to exit. Extended teachers entering through alternative routes such as the New York City Teaching Fellows program or Teach for America are less likely to remain in the same school than teachers entering through college recommended programs. In contrast, the average value-added estimates of extended teachers who remain in the same school are higher than those who do not, but the sample sizes are smaller for these measures and the differences are not statistically significant at traditional levels (see appendix Table 3).

The voluntary attrition of these less effective teachers would benefit students only if they are replaced by relatively more effective teachers. We explore this question by comparing the effectiveness of teachers who were extended and left schools in 2010-11 or 2011-12 with teachers hired at these schools.⁴ Teacher effectiveness measures for teachers hired at these schools in 2011-12 and 2012-13 (actual replacements) are unavailable. Rather we employ the effectiveness of teachers hired at these schools in 2008-09 and 2009-10.⁵ For each school with an extended leaver, we compare the average effectiveness of extended leavers with that of their proxy replacements, and

⁴ Teachers who were hired include both those new to teaching and teachers who transferred from other schools.

⁵ The vast majority of teachers with tenure decisions in 2010-11 and 2011-12 began their probationary periods in 2008-09 or 2009-10. We therefore are comparing the extended leavers to other teachers hired under similar circumstances to themselves. We are making the assumption that the teachers hired in 2009 and 2010 at the schools where an extended teacher left in 2010-11 or 2011-12 have measured effectiveness similar to those teachers who hired at these schools in 2011-12 and 2012-13. We have also created a replacement comparison group of teachers by examining teachers who were hired at these schools from 2006-07 through 2009-10.

then average these within school differences across all such schools. In this way we examine the difference in teacher effectiveness between extended leavers and proxy replacements in the typical school.

As shown in Table 2, there are substantial differences in the effectiveness of extended leavers and their proxy replacements. For example, there are 45 percentage points fewer teachers rated as highly effective or effective among all extended leavers than their proxy replacements (14 percentage points Highly Effective and 31 percentage points Effective). Estimated value-added in ELA is 20 percent of a standard deviation higher among the proxy replacements than the extended leavers.⁶ Although proxy replacement teachers are estimated to outperform extended leavers in math value-added, this difference is not statistically significant at traditional significance levels, due primarily to relatively few observations (N=158).

From a principal's perspective, these are large effects relative to almost any other intervention they might contemplate. For example, many principals rightly privilege experience when hiring teachers as the value-added of a teacher with six years of experience is estimated to be up to 15 percent of a standard deviation higher than a novice teacher (Atteberry, Loeb and Wyckoff, 2013). Extending the probationary period of teachers with insufficient skills to be approved for tenure and thereby nudging some teachers to leave the school who are then replaced with a new teacher has an effect on teacher effectiveness about the same as the gains of hiring a teacher with six years of experience rather than a novice.

School Differences

While implementation of the policy may have varied across schools, most schools experienced a substantial change in the percentage of teachers who were approved for tenure under the new policy. More than 70 percent of schools granted tenure to fewer than 80 percent of their teachers following the introduction of the policy as shown in Figure 2. While a cluster of schools approved 100 percent of eligible teachers, most schools approved far less, with another large cluster of schools with between 50 and 70 percent approval.

The variation in approval rates seen in Figure 2 corresponds to some school characteristics, particularly average student attributes, as shown in Table 3. On average, teachers approved for

⁶ Employing the sample of teachers entering schools between 2006-07 and 2009-10 as the proxy replacement comparison group, we estimate the percentage of teachers rated highly effective or effective is 44 percentage points higher for the proxy replacements than the extended leavers. Estimated value-added is 13 percent of a standard deviation higher in ELA and 14 percent of standard deviation higher in math, which are both significant at the 0.06 level.

tenure work in schools in which the percentage of white students is more than 50 percent greater than schools where teachers' probation is extended. Black students experience the reverse. In schools where teachers are approved for tenure, black students comprise 27 percent of all students, but they comprise 35 percent of students in schools where teachers' probation is extended. The achievement of students in schools where teachers receive tenure is 15 percent of a standard deviation better in math and 13 percent of a standard deviation better in ELA than the average achievement in schools where teachers are extended.

Given the strong link between principal effectiveness ratings and tenure decisions shown above, it is not surprising that the pattern of differences in school attributes across principal effectiveness ratings mirror the differences across tenure outcomes as shown in Table 3. For example, the average highly effective teacher works in schools where the percentage of white students is twice as large as it is for the average ineffective teacher. The average ineffective teacher is located in a school with 65 percent more black students than their average highly effective colleague. As is also shown in Table 3, the average ineffective teacher is located in a school where the ELA performance of students is more than a quarter of standard deviation lower and more than 30 percent of a standard deviation lower in math than that of the average highly effective teacher. This suggests that replacing ineffective and developing teachers with a teacher whose performance is closer to the average would disproportionately improve the quality of teaching in schools with higher percentages of black students.

To examine the types of schools most affected by the tenure reforms, we estimated the relationship between school characteristics and tenure decisions in a multivariate framework controlling for teacher performance measures. When we estimate a model that includes only the attributes of the students in the school, the percentage of students who are black is the only measure that is associated with the likelihood of being extended. However, when teacher attributes are added to the model, they dominate the determination of whether a teacher is extended. The estimate for the percent of black students drops substantially in magnitude such that a 1 standard deviation increase in the percentage of black students (26.4 percentage points) is estimated to increase the likelihood of a teacher being extended by just over 1 percent (see appendix Table 4). Thus, the likelihood of teachers being extended is much greater in schools with larger percentages of black students, but the differential primarily results from these schools having more teachers judged to be Ineffective and Developing.

Discussion

Teacher tenure has been a hotly debated issue for decades, but there is surprisingly little research that documents the effects of various tenure policies. This paper examines an unusual change in the tenure policy in New York City as a step toward providing evidence to support the design of teacher workforce policies.

Our analysis documents substantial changes in tenure decisions following the NYC reforms. While almost all eligible teachers received tenure prior to the change, after the reforms a large share of teachers instead had their probationary periods extended to provide more opportunity for them to demonstrate the skills necessary for effective teaching and for district decision makers to better assess teachers' performance. Not surprisingly, low-performing and less qualified teachers were more likely to be extended. Teachers in schools with disproportionate shares of black and low-performing students also were more likely to be extended. Our analyses provide some evidence that this differential reflects a uneven distribution of less effective teachers, which is consistent with recent research (Isenberg et al., 2013; Sass et al., 2012), although we cannot rule out differential application of tenure rules. Finally, we found evidence that the tenure policy resulted in additional voluntary attrition of teachers who were extended, as well as additional involuntary dismissal of the small share of teachers who were denied tenure. Among extended teachers, those with lower effectiveness, as measured by principals' ratings, but higher qualifications (e.g. SAT scores) were more likely to leave, potentially further strengthening the teacher workforce. Extended teachers who leave their schools are less effective as measured by principal ratings and value-added estimates than are those likely to replace them. Because teachers with poor effectiveness ratings are more likely to be in schools with higher percentages of black students, these schools are most affected by the policy change and most likely to see attrition of these less effective teachers as a result of the reforms. These schools on average were able to hire more effective teachers to fill these vacancies.

New York City's reforms to the tenure process are still in their early stages. Our results suggest large effects but provide only preliminary evidence because we have not fully ruled out the effects of other factors that may have been at play in the district simultaneously. With additional data a causal analysis will be more feasible and we can address additional questions. While the direct effects of the tenure reforms are felt by teachers facing tenure decisions, the labeling of teachers and increased likelihood of receiving an extension may induce other teachers in the same school, subject,

and/or grade to reassess their positions. These processes may encourage principals to reassign teachers across grades and subjects or to reallocate responsibilities in other ways.

Changes in human resource practices including new hiring and evaluation policies have been hallmarks of many recent reforms. While the tenure process has been the subject of continual debate, reforms have been slower and less sustained in this area. In part as a result, research on tenure policies and variety of possible approaches to probationary periods and screening is sparse. Nearly all districts grant some form of tenure based at least in theory on teachers demonstrating proficiency. Yet many districts do only cursory evaluation during the tenure process. As such, adopting tenure reform similar to that presented here may be comparatively easy relative to other much discussed human resource policies that require more controversial policy changes.

References

- Atteberry, Allison, Loeb, Susanna, & Wyckoff, James (2013). *Do First Impressions Matter? Improvement in Early Career Effectiveness*, CALDER Working Paper No. 90. Washington, D.C.: CALDER.
- Beteille, Tara, Kalogrides, Demetra, & Loeb, Susanna (2012). Stepping stones: Principal career paths and school outcomes. *Social Science Research*, 41(4), 904–919.
- Brown, Steven P., & Leigh, Thomas W. (1996). A new look at psychological climate and its relationship to job involvement, effort, and performance. *Journal of Applied Psychology*, 81(4), 358-368.
- Brunner, Eric J., & Imazeki, Jennifer (2012). Probation length and teacher salaries: Does waiting pay off? *Industrial and Labor Relations*, 64(1), 164-180.
- Chetty, Raj, Friedman, John N., & Rockoff, Jonah E. (2012). *The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood*. Working Paper 17699. Cambridge, MA: National Bureau of Economic Research.
- Corcoran, Sean P., Schwartz, Amy Ellen, & Weinstein, Meryle (2009). *The New York City Aspiring Principals Program: A School-Level Evaluation*. New York: IESP.
- Feinberg, Robert M. (1981). Earnings-Risk as a Compensating Differential. *Southern Economic Journal*, 48(1), 156-163.
- Flyer, Fredrick, & Sherwin Rosen (1997). The New Economics of Teachers and Education. *Journal of Labor Economics*, 15(1), S104-S139.
- Isenberg, Eric, Max, Jeffrey, Gleason, Philip, Potamites, Liz, Santillano, Robert, Hock, Heinrich, & Hansen, Michael (2013). *Access to Effective Teaching for Disadvantaged Students*. Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Kalleberg, Arne L. (1977). Work Values and Job Rewards: A Theory of Job Satisfaction. *American Sociological Review*, 42(1), 124-143.
- Lee, Raymond T., & Ashforth, Blake E. (1996). A meta-analytic examination of the correlates of the three dimensions of job burnout. *Journal of Applied Psychology*, Vol 81(2), 123-133
- National Council on Teacher Quality (2012). *State of the States 2012: Teacher Effectiveness Policies – Area 3: Identifying Effective Teachers*. Washington, D.C.: National Council on Teacher Quality.
- New York City Department of Education (2009). *The Tenure Toolkit, 2009-10*. New York: New York City Department of Education.
- New York City Department of Education (2010). *The Tenure Toolkit, 2010-11*. New York: New York City Department of Education.
- New York City Department of Education (2011). *The Tenure Toolkit, 2011-12*. New York: New York City Department of Education.
- Rivkin, Steven G., Hanushek, Eric A., and Kain., John F. (2005). “Teachers, Schools, and Academic Achievement.” *Econometrica*, 73(2), 417-458.
- Rockoff, Jonah E. (2004). The impact of individual teachers on student achievement: evidence from panel data. *American Economic Review Papers and Proceedings*, 94(2), 247–252.

Sass, Tim R., Hannaway, Jane, Xu, Zeyu, Figlio, David N., & Feng, Li (2012). Value added of teachers in high-poverty schools and lower poverty schools. *Journal of Urban Economics*, 72(2), 104-122.

Figure 1. Percentage of Teacher Tenure Cases by Tenure Outcome 2007-08 to 2012-13

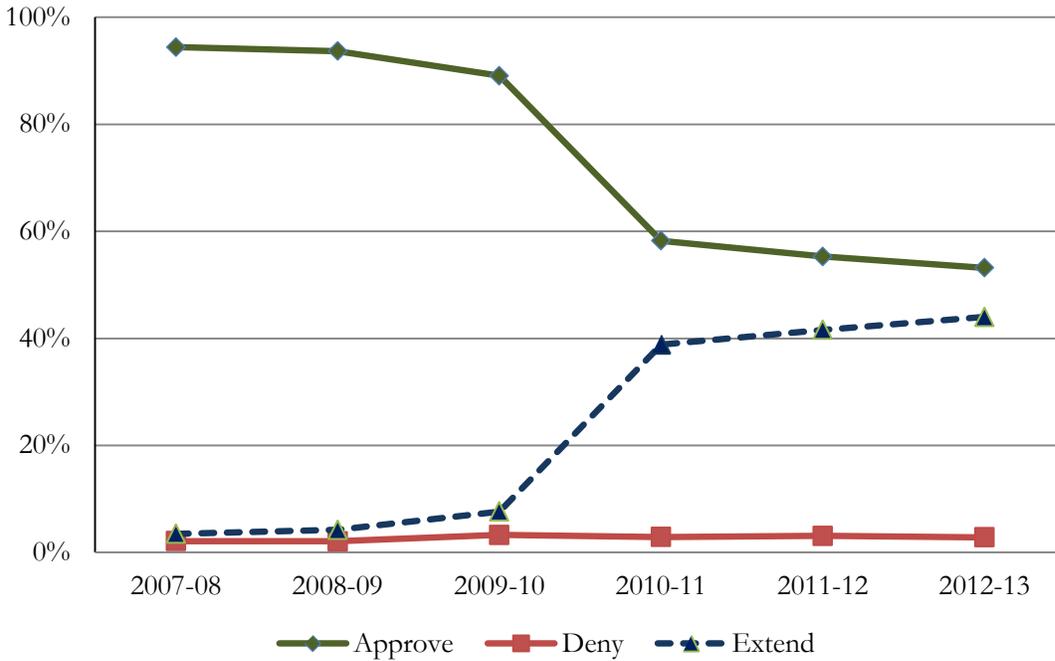
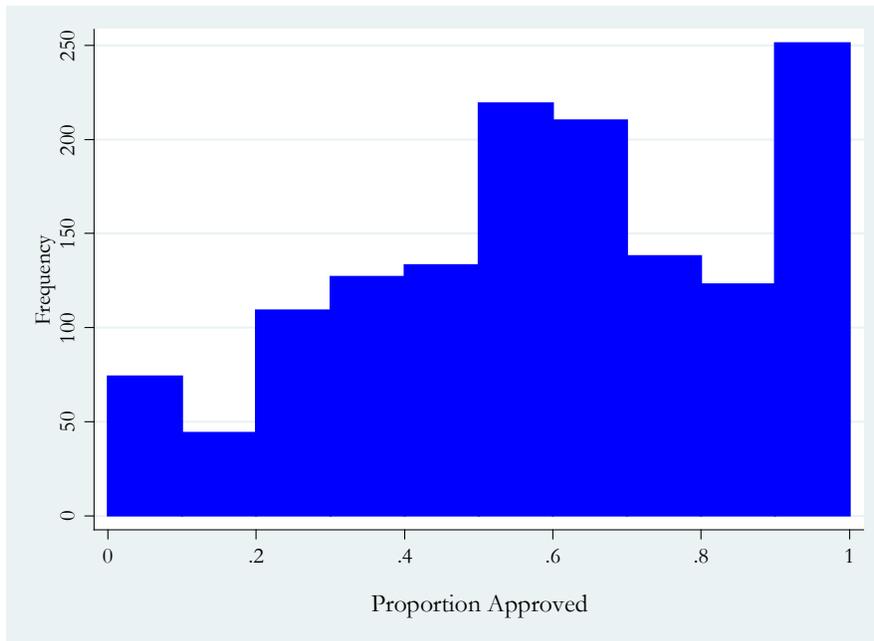


Figure 2. Distribution of School Proportion of Tenure Cases Approved 2009-10 through 2012-13



Notes: Includes only schools with at least four tenure decisions over the period (81 percent of all schools).

Table 1. Determinants of Teacher Disposition in Year Following Tenure Decision, 2010-11 and 2011-12

Variables	(1) Transfer	(2) Transfer	(3) Transfer	(4) Exit	(5) Exit	(6) Exit
Extend	0.145** (15.21)	0.124** (13.04)	0.087** (6.06)	0.057** (9.38)	0.055** (9.07)	0.040** (4.32)
Student Attributes						
Mean Math score	-0.024 (-0.68)			0.016 (0.68)		
Mean ELA score	-0.024 (-0.64)			-0.019 (-0.82)		
Black (%)	0.113* (4.24)			0.042* (2.46)		
Hispanic (%)	0.066~ (2.35)			0.075** (4.21)		
Free lunch (%)	-0.099** (-3.12)			-0.085** (-4.23)		
Reduced lunch (%)	-0.307* (-2.33)			-0.187* (-2.23)		
Principal Final Effectiveness Rating						
Ineffective			0.285* (4.24)			0.110* (2.54)
Developing			0.071** (3.58)			0.026* (2.02)
Effective			0.030* (2.13)			0.007 (0.74)
Missing			0.045* (2.45)			0.013 (1.11)
Constant	0.142** (4.88)	0.135** (24.56)	0.111** (9.60)	0.064** (3.48)	0.037** (10.52)	0.031** (4.22)
School Fixed Effect		X	X		X	X
Observations	6,351	8,855	8,855	6,351	8,855	8,855

Notes: T-statistics in parentheses. ** p<0.01, * p<0.05, ~p<0.1

Table 2. Mean School Difference in Teacher Effectiveness Measures between Proxy Replacement and Extended Leavers in Schools with Extended Leavers, 2010-11 and 2011-12^a

Extended Leaver Status	Principal Final Effectiveness Rating (%)				Value-Added	
	<i>Highly Effective</i>	<i>Effective</i>	<i>Developing</i>	<i>Ineffective</i>	<i>ELA</i>	<i>Math</i>
All Extended leavers	14.34***	30.7***	-36.45***	1.37*	0.197**	0.119
Extended transfers	11.97***	30.16***	-34.53***	1.14	0.127	0.181*
Extended exiters	16.15***	27.55***	-33.24***	1.72	0.298*	0.037

Notes: ^a Proxy replacement teachers are all teachers hired at the school in 2009 and 2010. Only schools with an extended leaving teacher in 2011 or 2012 included in all comparisons. Positive values indicate on average within schools average value for replacement pool exceeds that for the Extended leavers. Comparing extended leavers to proxy replacements *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

Table 3. Attributes of the Students in Teacher's School by Tenure Decision and Principal Effectiveness Rating, 2010-11 and 2011-12

	White (%)	Hispanic (%)	Black (%)	Home Lang Eng (%)	Free Lunch (%)	Reduced Lunch (%)	Math Achieve (z-score)	ELA Achieve (z-score)
Tenure Decision ^a								
Approve	13.8	44.4	27.4	56.6	72.3	4.4	0.081	0.086
Extend	8.9	44.6	35.1	60.3	77.3	4.1	-0.066	-0.042
Deny	7.1	43.5	39.6	63.3	77.8	4.2	-0.152	-0.093
Principal Effectiveness Rating ^b								
Highly Effective	16.4	42.8	24.1	56.5	69.2	4.8	0.184	0.181
Effective	12.1	45.3	29.9	57.2	74.6	4.2	0.007	0.019
Developing	8.4	45.0	35.3	60.8	78.1	4.1	-0.068	-0.046
Ineffective	7.2	42.4	39.9	62.7	77.7	4.6	-0.161	-0.102
No rating	12.3	42.7	31.0	57.4	71.3	4.1	0.055	0.073
Total	11.7	44.5	30.8	58.3	74.4	4.2	0.015	0.029

Notes:^a Extended teachers work in schools with different student attributes than approved teachers (p-value less than 0.01 for all attributes except the percentage of Hispanic students). Teachers denied tenure work in schools with different attributes than teachers who are extended with respect to the percentage of students who are black, the percentage whose home language is not English and mean student math scores (p-value less than 0.05). Differences in other student attributes are not significantly different from zero.

^b Teachers rated ineffective work in schools with different student attributes than teachers rated effective or highly effective (p-value less than 0.01 for all attributes except the percentage of Hispanic students and the percentage eligible for reduced-price lunch). Teachers rated developing work in schools with different student attributes than teachers rated effective or highly effective (p-value less than 0.01 for all attributes except the percentage of Hispanic students).

Technical Appendix

Estimation of Teacher Value-Added

We estimate teacher-by-year value-added employing a multi-step residual-based method similar to that employed by the University of Wisconsin's Value-Added Research Center (VARC). VARC estimates value-added for several school districts, including until quite recently New York City.

We initially estimate equation 1, which regresses achievement (Y_{icsjt}) for student i in class c at school s taught by teacher j in time t as a function of prior achievement ($Y_{icsjt-1}$), student attributes (X_{icsjt}), and class fixed effects (α_{csjt}). In this model, the class fixed effects subsumes both the teacher-by-year fixed effect (τ_{jt}) and any other class (Z_{csjt}) or school-level (S_{st}) predictors of student achievement.

$$Y_{icsjt} = \lambda Y_{icsjt-1} + \beta' X_{icsjt} + \alpha_{csjt} + \varepsilon_{icsjt} \quad (1)$$

$$\text{where } \alpha_{csjt} = \gamma' Z_{csjt} + \varphi' S_{st} + \tau_{jt}$$

Employing these estimates, we calculate the residuals (r_{icsjt}) from this regression without accounting for α_{csjt} and then estimate equation 2 which regresses this residual on class and school characteristics as well as a class random effect (ζ_{jt}) to reflect the grouping of students into classrooms.

$$r_{icsjt} = \alpha_{csjt} + \varepsilon_{icsjt} = \gamma' Z_{csjt} + \varphi' S_{st} + \zeta_{jt} + \omega_{icsjt} \quad (2)$$

Employing these estimates, we calculate the residuals (q_{icsjt}) from this model and calculate teacher-by-year value-added by averaging across the student-level residuals within a teacher and year.

$$\hat{\tau}_{jt} = \bar{q}_{icsjt} \quad (3)$$

The standard errors of the teacher-by-year value-added estimates are estimated as shown in equation 4 using the student-level errors ($e_{icsjt} = q_{icsjt} - \hat{\tau}_{jt}$) from equation 3 and number of observations for each teacher-by-year group.

$$SE(\hat{\tau}_{jt}) = \sqrt{\text{var}(e_{icsjt})/N_{jt}} \quad (4)$$

We then employ a standard Empirical Bayes shrinkage method to account for the varying uncertainty associated with each teacher-by-year value-added estimate.

Appendix Table 1. Descriptive Statistics for the Analytic Sample, 2009-10 to 2012-13

Variable	Obs.	Mean	Std. Dev.
Tenure Outcome (%)			
Approve	19,372	66.6	47.2
Extend	19,372	30.3	46.0
Deny	19,372	3.1	17.2
Teacher Attributes (% unless otherwise noted)			
Female	19,335	75.7	
Black	17,113	18.5	
Hispanic	17,113	16.6	
SAT math	9,104	498	102.0
SAT verbal	9,104	501	98.0
Preparation Path (%)			
College recommended	19,237	51.8	
Teaching Fellow	19,237	21.9	
TFA	19,237	2.9	
Individual evaluation	19,237	8.5	
Temporary license	19,237	5.1	
Student Attributes (aggregated to school)			
Hispanic (%)	15,124	43.8	25.1
Black (%)	15,124	31.0	26.3
Free lunch (%)	14,057	63.4	27.5
Reduced lunch (%)	14,057	3.9	4.1
Mean ELA score (z-score)	11,010	4.9	44.7
Mean Math score (z-score)	11,022	3.0	49.8
Teacher Performance Measures (%)			
U rated	19,372	2.33	
D rated	19,372	0.14	
Principal Final Effectiveness Ratings			
Ineffective	13,080	2.3	15.0
Developing	13,080	29.2	45.5
Effective	13,080	42.6	49.4
Highly Effective	13,080	15.3	36.0
No Rating	13,080	10.6	30.8
Low attendance	13,080	8.0	27.0
VAM ELA	1,498	0.00	1.1
VAM Math	1,538	0.08	1.2
NYC VAM low	2,410	9.3	
NYC VAM high	2,408	8.0	

Appendix Table 2. Attributes of Teachers by Tenure Outcomes, 2010-11 through 2012-13^a

Tenure Decision	Value Added		U Rated (%)	D Rated (%)	Low Attd (%)	SAT		LAST Exam	Preparation Route (%) ^b			
	ELA	Math				Math	Verb		Coll Rec	NYCTF	TFA	Ind Eval
Approve	0.081	0.248	5.7	22.2	37.1	505	505	257	59.9	49.5	60.2	55.0
Extend	-0.138	-0.129	52.1	66.7	56.2	490	494	254	37.8	47.2	38.9	40.7
Deny	-0.115	-0.740	42.2	11.1	6.7	469	490	248	2.4	3.2	0.1	4.3
Total	-0.009	0.070	100.0	100.0	100.0	498	500	255	100.0	100.0	100.0	100.0

^a Means of teachers approved exceed those of teachers extended at a p-value of 0.05 or lower for all attributes. The means of teachers extended exceed those of teachers denied at a p-value of 0.05 or lower for all variables except ELA value-added and verbal SAT.

^b The tenure approval rate is lower for teachers prepared through the NYCTF and IE preparation routes than those from CR programs at p-values of .01 or lower. There is no statistical difference between CR and TFA.

Appendix Table 3. Attributes of 2011 and 2012 Extended Teachers by Disposition in the Following Year

Attrition Status	Value Added		U Rated (%)	D Rated (%)	Low Attd (%)	SAT		LAST Exam	Preparation Route (%)			
	ELA	Math				Math	Verb		Coll Rec	NYCTF	TFA	Ind Eval
Same School	-0.091~	-0.090	4.0~	0.2**	10.7	491	495	253**	77.5	70.9**	53.3**	78.8*
Transfer	-0.355	-0.421	2.7	0.2	11.2	482	486	253	16.3	15.6	9.0	17.7
Exit	-0.332	-0.145	2.9	0.0	9.1	530	539	267	6.2	13.6	37.7	3.5

Notes: ** p<0.01, * p<0.05, ~ p<0.10. For Value-Added, U Rated, D Rated, Low Attendance, SAT and LAST Exam, significance levels denote significant differences between the values of these variables for Extended teachers who remain in same school and those who either transfer or exit. For Preparation Routes, significance levels denote differences between designated route and College Recommended.

Appendix Table 4. Determinants of Whether Teacher is Extended Relative to being Approved, 2010-11 and 2011-12

	(1) Extended (=1)	(2) Extended (=1)
Student Attributes		
Mean Math score	-0.096 (-1.41)	-0.073~ (-1.77)
Mean ELA score	-0.010 (-0.14)	0.021 (0.49)
Black (%)	0.211** (-4.41)	0.048~ (1.80)
Hispanic (%)	0.032 (-0.62)	-0.008 (-0.27)
Free lunch (%)	0.012 (-0.20)	-0.041 (-1.13)
Reduced lunch (%)	-0.066 (-0.26)	-0.043 (-0.28)
Teacher Attributes		
Low Attendance		0.066** (3.84)
Unsatisfactory Rated		0.101** (2.85)
Doubtful Rated		-0.125 (-0.75)
Principal Final Rating		
Ineffective		0.867** (25.61)
Developing		0.906** (95.62)
Effective		0.100** (8.72)
No rating		0.334** (12.95)
Constant	0.340** (-6.12)	0.081* (2.45)
Observations	6,351	6,351
R-squared	0.033	0.613

Notes: T-statistics in parentheses. ** p<0.01, * p<0.05, ~ p<0.1