IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 10547

Promotion Incentives in the Public Sector: Evidence from Chinese Schools

Naureen Karachiwalla
Albert Park

FEBRUARY 2017

DISCUSSION PAPER SERIES

IZA DP No. 10547

Promotion Incentives in the Public Sector: Evidence from Chinese Schools

Naureen Karachiwalla
IFPRI

Albert Park
HKUST, IZA and CEPR

FEBRUARY 2017

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Promotion Incentives in the Public Sector: Evidence from Chinese Schools

We provide evidence that promotion incentives influence the effort of public employees by studying China's system of promotions for teachers. Predictions from a tournament model of promotion are tested using retrospective panel data on primary and middle school teachers. Consistent with theory, high wage increases for promotion are associated with better performance, teachers increase effort in years leading up to promotion eligibility, and reduce effort if they are repeatedly passed over for promotion. Evaluation scores are positively associated with teacher time use and with student test scores, diminishing concerns that evaluations are manipulated.

JEL Classification: J31, J33, J45, M51

Keywords: teacher incentives, promotions, China

Corresponding author:
Albert Park
Hong Kong University of Science and Technology
Lo Ka Chung Building
Clear Water Bay
Kowloon
Hong Kong
E-mail: albertpark@ust.hk

1 Introduction

Public servants comprise a large proportion of formal employment in low-income countries, but service delivery is often poor (World Bank, 2004). In many developing countries, where teachers account for a large share of public sector workers, there are widespread complaints about high rates of teacher absenteeism and poor learning outcomes (Glewwe, 2002). Similar concerns are voiced in developed countries about systems in which teachers enjoy high job security and pay is determined by education levels and seniority rather than by performance. In such settings, teacher characteristics are only weakly correlated with student learning (Podgursky and Springer, 2007). As a result, there has been great interest in reforms that incentivise teachers, in particular through pay for performance schemes that link bonuses to student test scores (Muralidharan and Sundararaman, 2011; Glewwe and Muralidharan, 2015). The United States' *No Child Left Behind Act* included linking school sanctions to student performance to improve accountability in many states, with varying success on the types of effort exerted by teachers, and on student outcomes (Barlevy and Neal, 2012; Reback et al., 2014).

China has taken a much different approach to incentivising teachers, and civil servants more generally, by establishing a sophisticated system of annual performance evaluations that inform promotion decisions (Ding and Lehrer, 2007b). In Chinese public schools, teacher absence is very rare, and student achievement is sometimes spectacular, as evidenced by students in Shanghai scoring top in the world in all three tests on the Programme for International Student Assessment (PISA) exams (OECD, 2014). However, the design and performance of China's public service promotion system has received little attention or systematic study.

In this paper, we provide empirical evidence on how China's promotion system influences the performance of teachers in public schools. We test theoretical predictions of a tournament model of promotions using retrospective panel data on primary and middle school teachers in Gansu province in western China, and find evidence that effort responds to promotion incentives in a manner remarkably consistent with theory. Specifically, we show that in Chinese schools, promotions reward outstanding performance and are associated with large pay increases, performance is greater when wage increases are larger, performance improves as the teacher gets closer to the year in which she is eligible for a promotion, teachers who are repeatedly passed over for promotions exhibit deteriorating performance, and the larger the number of teachers competing for a promotion, the worse is the performance of teachers at the extremes of the skill distribution. Although we focus attention on teachers, key features of the evaluation and promotion system are the same for all public employees in China. Thus our results may help shed light on the puzzle of how China has grown so rapidly despite having weak institutions for protecting property rights and guarding against corruption, which some argue threaten the sustainability of China's growth (Acemoglu et al., 2012).

Nearly all previous studies of promotion systems focus on private sector employees, often testing whether pay within companies increases substantially with promotion rather than reflecting current marginal productivity (Medoff and Abraham, 1980; Baker et al., 1994). Only a few link promotion incentives to direct measures of effort or performance (Kwon, 2006; Campbell, 2008).

The performance of promotion systems may differ in the public sector in comparison to the private sector. Work output or productivity may be more difficult to quantify (Dixit, 2002), and public service jobs may involve multiple tasks or multiple principals to a greater degree than private sector jobs, making optimal incentives for specific tasks weaker (Holmstrom and Milgrom, 1991). Teacher performance pay schemes have generally produced positive results on the outcomes they incentivise (e.g., test scores) (McEwan, 2015); however, they have also been subject to teachers gaming the system (Cowgill, 2015; Glewwe et al., 2010; Jacob and Levitt, 2003), and may also have unintended dynamic effects on teacher effort (Macartney, 2014).

Unfortunately, there are few rigorous empirical studies of promotion systems in the public sector. Ashraf et al. (2014) find that career incentives induce the selection of more career-oriented individuals into the health services sector in Zambia. Dal Bó et al. (2013) and Ferraz and Finan (2009) find that higher pay in Brazil attracts political leaders who are more effective at delivering public services, while Enikolopov (2011) finds that US mayors receive higher pay if they promote economic growth. These studies do not focus directly on promotion decisions. Previous research on promotion outcomes has found that provincial and city officials in China are more likely to be promoted if they achieve high rates of economic growth (Li and Zhou, 2005; Landry, 2008). A study by Haeck and Verboven (2012) also examines promotion incentives in the education sector, showing that the wages of professors of a European university do not respond to the external labour market, and that those with better research and teaching performance are more likely to be promoted. However, none of the studies described above distinguish clearly whether the link between performance and promotions or higher pay reflects selection or incentive effects. This study is the first of which we are aware to provide empirical evidence that performance in the public sector responds directly to incentives. Thus, in addition to informing the current debate on teacher incentives and accountability in both low-income countries and in developed countries such as the United States about reforming teacher pay structures and evaluations, this paper also contributes to the literature on promotions and effort incentives more generally.

The rest of the paper is organized as follows. The next section describes the promotion system for teachers and public servants in China. Section 3 presents theoretical predictions to be tested empirically, based on intuition from a tournament model of promotions. Section 4 introduces the data and key variables. Section 5 presents the empirical specifications and results, and Section 6 concludes.

2 China's Promotion System

In China, all civil servants compete with their colleagues for promotions based on extensive annual performance evaluations. In the education sector, teachers in primary schools can be assigned to four different ranks: intern, primary level 2, primary level 1, and primary high level. In middle school, teachers can be assigned to one of five ranks: intern, middle level 3, middle level 2, middle level 1, and middle high level. Primary level 2 is equivalent to middle level 3, primary level 1 is equivalent to middle level 2, and primary high level is equivalent to middle level 1 in terms of administrative status. Teachers are all hired as interns and then promoted. Promotion to a higher rank is accompanied by a substantial increase in salary, as we show below.

There are specific rules regarding the years of service required before a teacher is eligible to apply for promotion to each rank. Once promoted, the years of service required for promotion to the next rank apply. All teachers begin as interns in their first year, and are promoted up the rank levels successively. Appendix Table 12 provides the number of years teachers must wait until promotion eligibility for each rank, based on their level of education.

Most teachers in the sample completed vocational college or regular college, with nearly all of the rest completing vocational high school or regular high school. Almost all middle school teachers are college graduates. In China, after completing middle school, the best students usually attend regular high school which requires passing a competitive exam, and others enrol in vocational high schools or exit schooling. Entrance to regular and vocational colleges is also based on competitive exams taken by regular high school graduates. The required educational credentials for teachers has increased significantly over time.

Apart from needing to wait the requisite number of years before applying to the next rank, promotion applications also require the attainment of minimum scores on annual performance evaluations. Annual evaluations are conducted on a four point scale: excellent, good, pass, and fail. To be eligible for promotion, teachers must obtain either two 'good' annual performance evaluation scores or one 'excellent' evaluation score *in the past five years*. Other promotion requirements may vary by county or education district (typically comprising all of the schools in a township). For example, in some education districts teachers must publish an article in a teaching journal.[1] There are also teaching awards for teachers at the county, education district, or school levels that can influence promotion eligibility.

Each year the Education Bureau of each county determines the specific number of promotions available for each rank in each education district, and teachers are informed of the quota numbers by the district education officials. The timing of the announcement is uncertain and will not generally coincide with the beginning of the school year, as it depends on budget and other considerations. All teachers who have fulfilled the minimum conditions are then considered for rank promotions after annual evaluations at the end of each school year.

Annual teacher performance evaluations assess multiple dimensions of teachers' performance, grouped into four categories: 1) student test scores, 2) work ethic and attitude, 3) preparation (lesson plans, homework, reporting etc.) and 4) teacher attendance. A point system is used, and although there is variation across districts, student test scores generally have the highest weight (34.2% on average), followed by preparation (29.5%), attitude (23.2%), and attendance

[1] The requirements for this also vary, with some counties simply requiring a summary of the teacher's teaching experience, and some counties ranking publications by the quality of the journal.

(13.2%).[2] Weights for each category and the specific indicators and points used to evaluate performance within each category are determined by a district committee typically comprised of principals, vice-principals, and outstanding teachers from schools in the district. Interviews with teachers made clear that most do not have a precise idea of what the specific weights are, but they do know the components. The committee draws upon information from assessments based on direct classroom observation,[3] assessments by other teachers in the school (peer review), questionnaires from students, data on student test scores, and principal reports of the teacher's attendance, preparation, and attitude. Once final scores are agreed upon within the committee, there are fixed percentages of teachers who receive 'excellent' and 'good' evaluation scores. These percentages are set at higher levels of government (provincial and national) and did not change during the period of our study. The names of the recipients of excellent and good evaluation scores are announced at an annual meeting of all teachers at the school. Since there is an explicit quantitative formula and the results are announced publicly, the ability of committees to manipulate evaluation scores is limited; however, this practice will vary by district. The district committee in charge of evaluations also makes recommendations on who should be promoted each year subject to final approval by the County Education Bureau, based on another point system that incorporates the history of annual evaluation scores, publications, awards, and other district-specific criteria. Note that the decision regarding quotas for the number of promotions available for each district is set at the county level, and teachers compete with other teachers in their district for promotions.

Principals also are evaluated annually using a point system and receive promotions and salary increases based on the performance of their school. A committee appointed by the County Education Bureau visits each school every year and observes the school's facilities and grounds, collects test score data, surveys teachers about the principal's management of the school, looks through the school accounts, etc. This process helps make the system of promotions incentive compatible, avoiding to a great degree the problems of influence activities for which promotions are often criticised (see Fairburn and Malcomson, 2001). Unlike in many countries, teacher

[2]In some districts, mean test scores are used, and in other districts, value added test scores are used. Test scores from different subjects count equally in the performance evaluation.

[3]Teachers are typically observed once per semester. Senior teachers on the evaluation committee may assist in such evaluations. Teachers observing each other in public lectures for purposes of training and improvement also is common practice.

unions do not exist in China, and so play no role in the design or implementation of the promotion and compensation systems.

Wage rates reflect availability of fiscal revenue (Ding and Lehrer, 2007b), and can also reflect the priorities of current leaders. In addition to increasing substantially with each higher rank, wages also increase over time, often in the form of inflation adjustments or periodic national or provincial policies to increase the salaries of public employees. We note that because County Education Bureaus have a local monopoly (there are almost no private schools in these areas) and determine assignments of teachers to schools, there is very limited mobility of teachers across schools, and even less across counties because County Education Bureaus can administratively block teachers from making such a move. Teachers are only moved for specific reasons, such as a change in work location of a spouse, or because they have subject-specific skills that are needed in another school or county. Salaries are very similar across public schools within the same counties, providing little incentive to switch schools within the county to pursue higher pay. Han (2013) notes that beginning in 2001, Gansu began to centralize teacher allocation across schools. She finds, however, that differences in wages and other benefits are not significantly different across counties that did and did not centralize early.[4] China's residential registration system links access to public goods and services to one's place of official registration, which also inhibits labour mobility (Chan et al., 2008).

3 Theoretical Predictions

In this section, we present a set of theoretical predictions on how the promotion system should function and influence teacher performance, which we later take to the data. To fix ideas, we first sketch out basic features of a one-period tournament model of promotions and provide intuition for the main predictions, referring readers to Gibbs (1989) and Gibbs (1995) for more details and proofs. We also extend the Gibbs model and present predictions about dynamic incentives when the tournament occurs over multiple periods.

In a one-period tournament model, n teachers in a district compete for a fixed number of

[4] In related work, Han et al. (2014) find that schools in Gansu that centralized teacher deployment faster from 2001 to 2003 performed more poorly in terms of test scores and teacher work hours. Our data are from 2007 by which time the centralization of teacher deployment was long completed.

promotion slots, k, or with a promotion rate $\bar{p} = k/n$. Wage increases associated with promotion are taken to be exogenous. Teachers differ in their own skill, s, which is not directly observable by principals. Teachers have informed beliefs about their own skill. Teachers exert effort, e, in their job (capturing both hours and intensity of effort), which is costly and gives them disutility $C(e)$ that increases convexly with e. Performance of teacher i is measured by $q_i = s_i + e_i + v_i$, where q_i is output and v_i is an error term.

Each teacher's probability of promotion depends on her skill and on her effort, and on the skill and effort of his or her competitors. Thus, in each tournament a teacher's probability of promotion can be written as $p = p(e, s, \underline{E}, \underline{S}, \bar{p}, n)$ where $\underline{E}$ and $\underline{S}$ are vectors of other teachers' efforts and skills, respectively, and $\bar{p}$ and n are the promotion rate and number of teachers in the tournament, respectively. The county government offers a reward for promotion, ΔEV, which is the change in expected lifetime utility gained from winning a promotion relative to not being promoted. This change mainly reflects the greater lifetime income associated with the wage increase that comes with promotion but could also reflect non-pecuniary benefits and the option value of possible future promotions.

The teacher chooses effort to maximize the expected reward minus the disutility of effort:

$$\max_e \{p(e, s, \underline{E}, \underline{S}, \bar{p}, n)\Delta EV - C(e)\} \tag{1}$$

The first order condition for optimal effort sets the marginal cost of effort equal to the gain in utility times the marginal effect of effort on the probability of promotion:

$$C'(e) = \frac{dp(e, s, \underline{E}, \underline{S}, \bar{p}, n)}{de}\Delta EV \tag{2}$$

With appropriate assumptions, we can solve for the Nash equilibrium effort levels of this game. We note that with imperfect information, effort will depend on beliefs about one's own skill (s) and that of others, which may not be fully accurate. In deriving predictions for the static model, we assume away such error, but later we discuss how beliefs may be updated when we extend the model to multiple periods. From the FOC (equation (2)) we see that effort depends on the gain in expected utility from promotion, in particular the wage gain that a teacher would receive if promoted, and the marginal effect of effort on promotion probability

(MPE). The MPE depends on the number of competitors (n), the promotion rate ($\bar{p}$), and skill (s). Thus, we can express the reduced form equation for the determinants of each teacher's effort as follows:

$$\begin{aligned} e &= f(\Delta EV, MPE), MPE = h(\bar{p}, n, s) \\ e &= e(\Delta EV, \bar{p}, n, s) \end{aligned} \tag{3}$$

The assumptions and results of this tournament model deliver a number of testable predictions. First, we can test for evidence of two important features of promotion systems; these two conditions must hold in order for the model to predict behaviour:

Condition 1. Wages increase substantially after promotion, independent of changes in education or experience.

Condition 2. Greater effort (and skill) increases the probability of promotion.

Next, we present a number of predictions related to the determinants of effort predicted by the model:

Prediction 1. Effort incentives are greater the higher the wage increase if promoted. In the model, this corresponds to a higher ΔEV in equation (2).

Prediction 2. Average effort is greatest when the promotion probability is close to one half.

This will be true in models in which individuals have no knowledge about their skill level, or if the distribution of skills is concentrated around the median (for example, the normal distribution). This is because the chance that greater effort will influence promotion outcomes falls as promotion probability moves closer to one or zero. For example, if the promotion probability is close to zero, it is less likely for effort to matter, because only a few persons will be promoted, so effort will matter only for those with very high random draws of luck (when priors about skill are homogeneous) or very high skill levels (if skills are known). Similarly, when the promotion probability is close to one, effort will matter only for those with very low draws of luck (when priors about skill are homogeneous) or very low skill levels (if skills are known).

Prediction 3. Given promotion rate $\bar{p}$, incentives decrease monotonically with distance from the marginal skill percentile.

Define a marginal skill teacher as one whose skill percentile is equal to $1 - \bar{p}$. This is the teacher for whom effort incentives are highest. For example, if 1/3 of teachers are promoted, the teacher with the highest incentives is the one who believes she is at the 2/3 skill percentile (from the bottom). Such a teacher is on the margin between being promoted and not being promoted based on his or her skill level, so that effort will make the most difference to her probability of promotion.

Prediction 4. Given heterogeneity in skills, and holding $\bar{p}$ constant, as the number of competitors increases (higher n) effort increases for those with skill near the marginal skill percentile and decreases for those far from the marginal skill percentile.

This is because with more competitors, the contest becomes more precise in that the variance of the threshold score required to win is reduced. With only a few competitors, even with relatively low skill, individual effort might still matter if there is a chance that competitors also have low draws of skill or luck. However, with many competitors, there is less uncertainty about one's relative position, so for those whose skills are very low (or very high) relative to the marginal skill percentile, effort is much less likely to influence outcomes.

In practice, promotion tournaments for teachers in China are not completed in a single period. Rather, evaluations occur every year while promotions occur after multiple years of teaching, and failure to be promoted in one year does not preclude promotion in future years. Thinking about the model in terms of multiple periods leads to further predictions about how effort responds to the timing of promotion incentives.

We focus first on how a specific rule in China's promotion system influences effort incentives over time. As described earlier, to be eligible for promotion, a teacher must have completed X years of service (depending on her level of education) and must have obtained one 'excellent' or two 'good' evaluations within the past five years to become eligible for promotion. Our next prediction concerns the timing of incentives in relation to the year of promotion eligibility.

Prediction 5. As long as $t \leq X - 5$, there is little incentive to exert effort because in all of the subsequent five years, the teacher is ineligible for promotion so the performance evaluation has no bearing on promotion. In the years from $t = X - 4$ to $t = X$, the marginal probability of effort increases because with each additional year of service the number of future eligible years that can be influenced by current effort increases by one year. In year $t = X - 4$ for example, there is one year in which current effort affects probability of promotion and the teacher is eligible for promotion ($t = X + 1$). In year $t = X - 3$ there are two such years ($t = X + 1$ and $t = X + 2$), etc. In all subsequent years, the teacher is eligible for promotion in each of the subsequent five years (there are always five 'eligible years') so there is no further increase in effort incentives. For a more detailed proof of Prediction 5, see Appendix D.

Another dynamic aspect of incentives is how a teacher's effort responds to updated beliefs about his or her relative skill. Our next prediction regards the way in which effort responds to one important type of information, the promotion outcome itself.

Prediction 6. A teacher's effort declines as she continues to be passed over for promotion year after year.

The simple intuition for this prediction is that continued failure reveals negative information that should discourage teachers, leading to lower effort. Assume that in year 1, teachers begin with a prior belief (s_1) about their skill, while their true skill is s. Their beliefs can change over time as teachers update their beliefs based on new information. Because the probability of promotion is a positive function of both skill and effort in the previous year, the lack of promotion in a year in which the teacher is eligible for promotion provides an unambiguously negative signal to the teacher about her relative skill level. Using Bayesian updating, if a teacher is passed over for promotion in period t, the teacher will adjust downward her estimate of her relative skill level. In other words, $s_t < s_{t-1}$ if the teacher is not promoted in period $t > X$.

A number of factors influence the extent to which the teacher believes the failure to be promoted is informative about her skill level. For example, if very few teachers can be promoted because of few quota positions available for higher ranks (low $\bar{p}$) or if the teacher knows that some other teachers are better performers and are likely to be promoted before her (meaning that she believes her skill percentile is further from $1 - \bar{p}$ than that of her competitors), then it may take repeated promotion failures before she substantially revises downward her assessment of her relative skill level. However, as additional years pass without promotion, the teacher's estimate of her relative skill level should eventually fall, further increasing the gap between estimated relative skill s_t and $1 - \bar{p}$.[5] From earlier, we know that the larger this gap, the lower

[5]We may wonder whether, once the better teachers are promoted, those remaining in the rank may believe they have higher relative skill. However, there still will be many teachers in the rank who are not yet eligible for promotion (including new entrants to the rank) as well as older, likely less skilled teachers who are retiring. We

the marginal probability of effort.

Another potentially important source of information that could influence assessments of one's relative skill level is the annual evaluation score, which can be considered a type of intermediate feedback. However, it is difficult to predict how such feedback will influence effort incentives or to estimate the relationship empirically. There is an emerging literature that examines how feedback influences the outcomes of dynamic tournaments. Only a few studies consider heterogeneous contestants. Ederer (2010) shows that in a model with two contestants with differing ability, interim feedback has the effect of lowering effort incentives for both high and low ability types. The effect is more pronounced and equilibrium effort is lower when differences in ability are greater. When differences in ability are small, however, equilibrium effort is higher since both contestants are close to the margin. Thus, interim information may help teachers learn about their skill, enhancing the validity of the model's predictions assuming that individuals are informed about their own skill level. However, in our context the informational content of annual evaluation scores remains limited in important respects.[6]

Our final prediction follows from our assumption that the possibility of promotion is what provides effort incentives.

Prediction 7. After being promoted to the highest rank, effort incentives no longer exist.

4 The Gansu Survey of Children and Families

The Gansu Survey of Children and Families (GSCF) provides a unique data source that enables us to test the conditions and predictions of the model. The GSCF is a longitudinal study of rural

thus expect that not being promoted for multiple years will cause a downwards revision to a teacher's beliefs on their relative skill.

[6]Despite receiving interim feedback from annual evaluations, teachers are still unlikely to have very complete information about their skill level relative to their competitors. They often do not know the full history of evaluation scores and other credentials of other teachers, especially those who teach in other schools. They also do not know the record of other teachers on specific requirements for promotion other than annual evaluation scores (i.e., publication in teaching journals, teaching awards, etc.). Even if a teacher has herself received either two good evaluation scores or one excellent score, she will still be unsure of her promotion probability, since that is a necessary but not sufficient condition for promotion. Unfortunately, it is beyond the scope of this paper to fully model or test the dynamic impact of interim feedback in the form of annual evaluation scores on subsequent effort. Given the promotion rules used in Chinese schools, it would be important to know the full history of evaluation scores of all teachers in a given rank in an education district to understand how they would influence subsequent effort. These data are not available. A practical concern is that if we were to regress annual evaluation scores on lagged scores, it would be difficult to disentangle incentive effects from mean reversion or serial correlation, especially given the short period of data available.

children in Gansu province in northwest China. The main survey sampled 2,000 children aged 9 to 12 in the year 2000 living in 100 villages in 42 districts in 20 randomly sampled counties. In order to understand the supply side of the educational system, all principals and teachers in the main primary and middle schools attended by the sample children in each village were also surveyed, providing detailed information on the background of principals and teachers, their time allocation, and their remuneration. Generally speaking, in rural counties in Gansu there is one primary school in each village and one middle school in each district. Each sample child can (often) be linked to his or her homeroom teacher. A second and third wave of the GSCF followed the same children in 2004 (at age 13 to 16) and 2007 (at age 16 to 19). In 2007 (at the end of the 2006/07 school year), individual and household questionnaires, as well as achievement tests, were also administered to a new cohort of 1,500 children aged 9 to 15.

This study uses a retrospective panel dataset constructed for teachers spanning the years 2003 to 2006, based on surveys of 2,100 teachers collected during the third wave of the GSCF in 2007. It was only in the third wave that retrospective data was collected on past teacher evaluation scores (data was collected for the years 2003-2006), as well as full promotion and education histories. For example, the panel of evaluation scores is constructed by taking the reported performance evaluation score in each of 2003-2006 and creating a separate variable for each year, and then transforming the data to long format. This procedure is also followed for other variables that change over time (for example, the teacher's rank, experience, etc.). The average sampling rate of teachers in the schools is 76%, 40% of schools have a 100% sampling rate, and only 27% of the schools have a sampling rate lower than 50%.[7]

Table 1 provides summary statistics on teachers in the sample, including breakdowns by teacher rank in 2006/07. Among the teachers, 36% are female. Teachers in higher ranks tend to be male, older, more experienced, better educated, and more likely to have received teaching awards or published papers in teaching journals.[8] Teachers in higher ranks also performed slightly better on their teacher placement tests (on the subjects of education, psychology, and language) taken when teachers apply for their first teaching job before they begin teaching,

[7]The sampling rate is calculated from the total number of teachers in the school as reported in the Principal's questionnaire. The rate is: number of teachers interviewed / total number of teachers reported by the principal. The principal reports the number of teachers in the school in the current year, and the number of teachers who have joined and left the school in the past three years (including the current year).

[8]These characteristics are consistent with the findings of Ding and Lehrer (2007b) describing teachers in much wealthier Jiangsu Province.

which we take as an indicator of ability prior to having any teaching experience. Of course, some of these differences in performance across ranks may be related to differences in years of experience.

To construct a measure of teacher ability, we define an 'ability index' using the teachers' scores on the three teacher placement tests. Scores are first standardized by the year in which the tests were taken. Then, for each year of the data (2003-2006) the following measure of relative skill is calculated: $[S - ave(S)]/ave(S)$ where S is the standardized score on the tests for a teacher, and $ave(S)$ is the average of the scores of the teachers in the same rank in the same education district.[9] An index of normalized scores for the three tests was constructed using Principal Components Analysis (PCA).[10] To study whether being at the extremes of the skill distribution affect incentives (for example, Prediction 3), we define dummy variables for whether a teacher falls into the top or bottom 10% of this skill measure.

As expected, average salaries increase substantially with rank. The biggest jumps are from primary 2 to primary 1, followed by middle 2 to middle 1, and then middle 1 to middle high. As the penultimate row of Table 1 shows, the average salary increases are 22.2%, 19.6%, and 17.8%, respectively.

Table 2 describes the teacher time use data. We have data on how teachers use their time in 2007. The 2007 GSCF survey asks each teacher "in a typical week, how many hours do you spend on...". Time use data is collected for the following teaching-related activities: teaching classes, grading homework, preparing lesson plans, participating in research activities, coaching students outside class, organising extracurricular activities, home visits, and disciplining students. We also sum up the total number of hours spent on these activities and report the total number of hours per week. Teachers at lower rank levels spend more hours per week on grading, lesson plans, and with students, which could reflect greater motivation or incentives, or a learning effect if more experienced teachers use their time more efficiently. Teachers in higher ranks also spend slightly less time teaching classes. Interestingly, teachers in the highest rank (middle high) work substantially fewer hours. So, promotions do not only entail higher pay, but also

[9] We standardize the test scores by the year in which the test was taken to account for the fact that the test changes each year and could change systematically in difficulty over time.

[10] Another index was also constructed that standardized each of the variables and then summed them. This did not affect the results.

less work.[11]

Table 3 provides a breakdown of evaluation scores by rank for the years 2003-2006. For most teachers, four years of data are recorded, so the total number of teacher-year observations of evaluation scores is 6,421. Not many teachers receive an evaluation score of 'fail', but the majority of the evaluation scores are a 'pass'. There is evidence that strict limits are enforced on the share of teachers who can obtain good evaluations, as the proportions of teachers receiving each type of evaluation score are quite similar across ranks. The data show that approximately 10-15% of teachers receive 'excellent' evaluation scores and 30-40% receive 'good' evaluation scores, and this is consistent across ranks.

We note that only 20 teachers (0.85%) in the sample reported switching counties between 2003 and 2007.[12] Further, the average proportion of teachers in a school who leave the school (for reasons not related to retirement) is only 3.7% per year.[13]

Figure 1 shows how long it takes to be promoted to each of the rank levels for all teachers for whom we have data from 2003 to 2006. Figure 1 shows, by rank level, the cumulative distribution of how long it took to be promoted from one rank level to the next for those who have been promoted (the cumulative percentage of teachers who were promoted within a particular number of years). From intern to primary 2, promotion is very fast, usually within two years.[14] Most teachers who get promoted from primary level 2 to primary level 1 do so within 4 to 5 years, so promotion at this level also appears to be routine. For promotions from primary level 1 to primary high level, the distribution of the number of years until promotion is far more spread out. There tends to be a steady stream of promotions, and it takes almost ten years on average to attain this promotion. Promotion from intern to middle 3 almost always takes one year or less. At middle level 3 rank, promotion also is almost inevitable, with most teachers promoted within 4 years. The promotions from middle level 2 to middle level 1 look similar to those from primary level 1 to primary high. They are more spread out, but the majority occur between 5 and 10 years. Very few people (less than 25%) are promoted from middle level 1 to middle high level in this sample, and the promotions that do occur take considerably longer

[11]Only 7 teachers in the data report having other sources of income during school terms, indicating that teachers spending less time in teaching are taking more leisure rather than pursuing other types of work.

[12]All the results are robust to dropping these 20 teachers (results not reported).

[13]This is reported in the principal's questionnaire.

[14]The teachers who are promoted in less than one year are cases in which a teacher was hired in the middle of a school year and promoted in the next school year, making it less than one calendar year.

(at least ten years). The differences in the promotion rates and timing across different ranks provide a good opportunity to study differences in effort due to differing promotion rates, as well as to study how effort evolves over time.

5 Empirical Framework and Results

This section describes the empirical tests of the conditions and predictions of the model and presents the results of the empirical analysis. We present the estimating equation and results for each condition and prediction in turn. We begin with Condition 1 (that wages increase with promotion) and Condition 2 (that high evaluation scores lead to promotions). Before turning to the predictions on teacher effort, we present evidence to justify the validity of using evaluation scores as a measure of effort. Next, we present the tests and results for Predictions 1-7 on the determinants of performance. Depending on data availability and the relationships of interest, the empirical specifications examine cross-sectional variation controlling for county fixed effects (Predictions 1-4), or focus on panel variation using teacher fixed effects (Predictions 5-7).

In the regression analysis, we focus on three rank levels for both the promotion and for the performance regressions: primary 1, middle 2, and middle 1. Promotions routinely occur within a limited duration of time for lower ranks (primary 2 or middle 3), and there are no further promotions after primary high and middle high levels. Thus, the teachers with primary 1, middle 2 and middle 1 rank levels most clearly have promotion incentives.

5.1 Promotion and Wages

To test Condition 1, that wages increase with promotion, we estimate Mincer-type wage regressions in which the log of the monthly wage is regressed on rank level dummies, and controls are added for experience, experience squared, education level dummy variables, and the most recent evaluation score. The coefficients on the rank dummy variables reveal the returns to promotion after controlling for human capital factors that determine labour productivity. We control for county fixed effects since wages are set at the county level, and we cluster standard errors by county.

Tables 4 and 5 contain the results from wage regressions for primary school and middle school teachers, respectively. The first column presents results of a regression of the log of monthly wage on rank level dummy variables; the second column adds controls for teaching experience and teaching experience squared; and the final column includes dummy variables for the highest level of education completed, as well as the performance evaluation score for the most recent year (2006).[15] The omitted rank levels are primary 2 for the primary school teacher regressions, and middle rank 3 for the middle school teacher regressions.

Rank level premiums are significant and increasing at each higher rank. According to the specifications with full controls, at the primary level, compared to primary 2 teachers, wages for primary 1 teachers are 7.5% greater and wages for primary high teachers are 13.5% greater. For middle school teachers, compared to middle 2 teachers, wages for middle 1 teachers are 8% greater and wages for middle high teachers are 24% higher.[16] These results show that promotions are rewarded by significant salary increases independently of experience, education, and ability, confirming Condition 1 and providing *prima facia* evidence that promotions are used to incentivise effort. Similar results form the main findings of earlier empirical studies of promotion systems in the private sector. Additionally, the coefficients on the performance evaluation score are not statistically significant when the rank level dummies are included. This result suggests that recent performance does not directly affect wages; but rather influences wages via promotion to a higher rank. After controlling for education and rank, there is a positive return to experience of about 2% per year. However, these increases are not necessarily due to improving skill or effort with experience; they could reflect inflation adjustments and normal periodic increases in salaries for teachers and other public sector employees.

5.2 Promotion Determinants

To test Condition 2, we estimate a linear probability model[17] of whether teachers are promoted and test whether high evaluation scores influence the probability of being promoted. We estimate the following equation using retrospective panel data for primary 1, middle 2, and middle 1 rank teachers:

[15] We also try a version with experience dummy variables and find that the relationship is actually quite linear.

[16] When the results of Tables 4 and 5 are compared to regressions that do not include county fixed effects, there is little change in the R^2s.

[17] The marginal effects from probit regressions are almost identical.

$$pr_{irc,t+1} = \varphi_0 + \varphi_1 ev_{irct} + \varphi_2 \bar{p}_{rc} + \varphi_3 ln(n_{r\tau t}) + \varphi_4 a_{ir\tau t} + \varphi_5 X_{irct} + \varphi_6 X_{crt} + \gamma_c + \epsilon_{irct}, \quad (4)$$

where $pr_{irc,t+1}$ is a dummy variable for whether teacher i in rank r in county c is promoted to the next rank at time $t+1$, ev_{irct} is the teacher's evaluation score at time t, $\bar{p}_{rc}$ is the promotion rate for rank r in county c, $n_{r\tau t}$ is the number of teachers in rank r in district τ at time t, $a_{ir\tau t}$ is the ability/skill of teacher i in rank r in district τ at time t, X_{irct} are individual teacher level control variables, X_{crt} are county-rank level time-variant controls, γ_c is a county (or district, or school) level time invariant unobserved factor, and ϵ_{irct} is an individual teacher level time-variant unobserved factor. To reduce the possibility of omitted variable bias, we control for all teacher and other characteristics that may affect effort and be correlated with promotion decisions. The coefficient φ_1 tells us whether higher teacher evaluation scores lead to promotions, and thus, whether we can expect the promotion system to motivate teacher effort. We run three different models, one with county fixed effects, one with district fixed effects, and one with school fixed effects.[18]

Next, we briefly describe the right-hand side variables. ev_{irct} is an ordinal variable for the different evaluation score categories (fail, pass, good, excellent).[19] To measure $\bar{p}_{rc}$, the promotion rate, we use the proportion of teachers in a county that historically are promoted at that rank within a 'reasonable' number of years. Using the whole sample, we first calculate the number of years it took for each teacher to be promoted from one rank to another, accounting for education level (we have each teacher's entire promotion history). A 'reasonable' number of years is defined as the number of years within which, for each rank and education level overall, half of the teachers were promoted. Then for each rank, county, and education level, we calculate the proportion of teachers that were promoted within that number of years.[20] We only include promotions before 2003 to ensure that this is a predetermined variable for the sample being analysed, so this is a historical promotion rate in the county. We split the promotion rate

[18] Teacher fixed effects is not very well powered since there are too few teachers in any particular school with a promotion observed in the four years for which we have evaluation score data.

[19] Results are robust to using a dummy variable for an excellent or good evaluation score.

[20] There is no relationship between the promotion rate and county size, so measurement error whereby smaller schools may have promotion rates farther from 50% should not be a concern in this context.

into quintiles, and the average promotion rate in the quintiles is 11%, 37%, 52%, 68%, and 89% in the first to fifth quintiles, respectively. We leave the third quintile as the omitted category. We split the promotion rate into quintiles and omit the middle quintile because of Prediction 2; that average effort incentives are lower when the promotion rate is further from one half.

To measure n, we use the number of teachers in the same rank in the district in a given year.[21] In the regressions, we use $ln(n)$.[22] The ability variables are dummy variables for whether the ability index is in the top 10% or bottom 10% of the skill distribution of teachers in the same rank and district in the same year.

In X_{irct} we include a dummy variable for whether the teacher is female, dummy variables for the education level of the teacher (high school, vocational college, and college), the number of years of experience, the scores on each of the three teacher placement tests (education, psychology, and language), and dummy variables for whether the teacher is in the pre-eligibility period ($t = X - 5$ to $t = X - 1$) or post-eligibility period ($t \geq X$) for promotion.

Including county fixed effects sweeps away potential bias from time-invariant county-level unobservables. Since promotion rates are set at the county-rank level, we may be worried that teacher characteristics at the county-rank level affect the promotion rates that are chosen. To address this possibility, we also include county-rank level means of teacher characteristics (X_{crt}), including the proportion of female teachers, average experience, an interaction term between proportion female and average experience, and the proportion of teachers in the county-rank with different levels of education.

The results are presented in Table 6. The coefficient on the evaluation score is positive and significant at the 1% level. This confirms Condition 2: that higher evaluation scores do indeed lead to promotions.

[21] The number of teachers in each rank is adjusted by the under sampling rate at the school. This is calculated using the total number of teachers in the school as reported by the school principal. The principals did not report on the total number of teachers by rank, so the total number reported was compared to the number of teachers interviewed in the school to calculate the sampling rate and the under sampling rate of the school. The number of teachers in each rank was then adjusted as follows: number interviewed * (1 + under sampling rate). This assumes that that there were no systematic differences in the number of teachers that were not interviewed by rank, and the adjustment was applied in the same way for each rank. We check correlations between the sampling rate and the variables included in our regressions. The average correlation is 0.06, which is very low. Thus, we do not believe that bias due to correlations between the sampling rate and the number of teachers variable would occur due to measurement error.

[22] The regressions use the log number of teachers because we expect the effects to be proportional rather than linear; we do not expect the same effect going from two to three teachers compared to going from 22 to 23 teachers. Additionally, the distribution of the number of teachers in a rank in a district has a long upper tail. Finally, logs fit the data better; the R^2 is higher when log number of teachers is used.

One difficulty in estimating the above equation is simultaneity bias caused by the fact that promotion probability could influence teacher effort and thus their evaluation scores. The direction of this bias is unclear because changes in promotion probability are expected to have different effects on teachers at different parts of the skill distribution. Although this issue is of second order, as a robustness check, we use wage differences between a teacher's rank and the next higher rank in a given county (ΔEV)[23] as an instrument for evaluation scores.[24] We find that the coefficient on the instrumented evaluation score is larger, suggesting that our OLS coefficients if anything underestimate the impact of evaluation scores on promotion probability.[25]

5.3 Evaluation Scores and Effort

The remainder of the empirical work will focus on the determinants of teacher effort. As a proxy for effort, we will be using the teacher's evaluation scores. This is because four years of data are available on evaluation scores, compared to only one year of data for other effort measures. However, evaluation scores should be a strong proxy for effort after controlling for measures of skill. This section will test this claim.

To provide evidence that annual evaluation scores reflect actual effort exerted by teachers, we perform two tests. The first uses cross-sectional data on teacher time use which is available for all teachers for one year (2007). The second uses student test scores, which also are available

[23] To measure ΔEV, we use differences in log wages so that the coefficient magnitudes can be interpreted approximately as the effect of percentage changes in wages. However, all results presented here are robust to using absolute wages instead. The teacher herself is excluded from the average wage calculation for the rank and county in which she currently teaches.

[24] The exclusion restriction assumes that wage increases for different ranks within the same county are exogenous and are not influenced by the evaluation scores of teachers or by unobservables correlated with teacher evaluation scores. China's education system is highly decentralized. It is the job of the local school board to implement the national curriculum, and county governments have substantial authority in setting teachers' wage rates. Wage levels reflect the availability of fiscal revenue (Ding and Lehrer, 2007b), and the spending priorities of current leaders. Differences in the characteristics of teachers in different ranks within the same county are controlled for by including county-rank mean teacher characteristics described earlier. Furthermore, wages affect promotions only through evaluation scores, because a teacher can obtain a significantly higher wage only by receiving a high evaluation scores and being promoted. The wage regression results (Tables 4 and 5) show that after controlling for rank level, wages are not influenced by the annual performance evaluations, and so are plausibly exogenous. Additionally, there is no teacher union in China so concerns that may be present in other countries, such as high ranked teachers being able to negotiate higher wage gaps for themselves and being able to influence their likelihood of promotion, are not a concern in this context. As such, wage rates are taken to be exogenous to teachers.

[25] However, although wage differences statistically significantly increase evaluation scores in the first stage, the first-stage F-statistic is weak (7.3) and the second stage impact of evaluation scores on promotion probability is not statistically significant at conventional levels.

for one year (2007) for those teachers whose students were part of the GSCF child sample. We estimate Probit models of the probability of obtaining an 'excellent' or 'good' evaluation score including different fixed effects (county, district, school) and test whether time use and or test scores significantly predicts high evaluation scores after controlling for a full set of teacher control variables.[26]

The time use variable is the log of total number of hours spent on all teaching-related activities in a given week as described above.[27] Math and language test score data exist for some of the teachers' students. The GSCF primarily focuses on children, and for the new cohort of children included in the main survey in 2007, we have data on their test scores in math and language. Because the survey mainly tracked children and then subsequently interviewed the other teachers in that child's school, we do not have test score data for more than one year for a particular teacher (teachers cannot be linked to other rounds of the survey due to the lack of consistent ID codes) and for many teachers we have no test score data. Furthermore, the GSCF children are linked to their homeroom teachers, and we only use a math (or language) test score if the homeroom teacher teaches math (or language). Applying this criteria, there are too few math scores to support analysis, so we focus on language test scores. We first standardize the language test scores by the grade level of the student, and we then average over the teacher's students (ranging in number from 1-9 students per teacher). The fact that we do not have complete test score data for all of the teachers' students has two statistical consequences. The first is measurement error, which will bias our results toward zero, making our estimates conservative ones. Second, the degree of measurement error will depend on the number of student test scores available, leading to heteroscedasticity. We can correct for this problem using appropriate regression weights.[28]

[26]There is no significant difference in time use or test scores between the categories of 'excellent' and 'good' evaluation scores.

[27]Log of total hours is used because the data shows a long right tail. Additionally, logs fit the data better – the Log likelihood is higher.

[28]Define x to be student test scores and $\bar{x}$ to be the estimated mean test score for each teacher based on N observations, and μ_x to be the true population mean of x. The variance of $\bar{x}$ equals the variance of x divided by N, or $\sigma^2_{\bar{x}} = \sigma^2_x/N$. We want to estimate the equation $y = \beta\mu_x + \epsilon$, where y is the evaluation score and μ_x is the true mean of x, but we only have data on $\bar{x}$, which is a noisy measure of μ_x (in other words, $\bar{x} = \mu_x + \epsilon_x$). Then, the regression equation with estimated means of x and measurement error is: $y = \beta\bar{x} - \beta\epsilon_x + \epsilon$. Here, it is transparent that the measurement error becomes part of the measurement error of y. The variance of the error term of this regression can be written as: $S_y = \beta^2(S_x/N) + S_\epsilon$, where S_y, S_x, and S_ϵ are the sample variances for y, x, and ϵ, respectively. It is easy for us to estimate S_y and S_x for each observation, and then calculate S_ϵ. S_y is the variance of the regression error term, which can be estimated by squaring the predicted regression errors.

We include teacher controls including the age of the teacher, a dummy variable for a female teacher, dummy variables for the level of education of the teacher, and the ability, promotion quintile, and number of teachers.

The results for time use are presented in Table 7 Panel A. Column (1) includes county fixed effects, column (2) includes district fixed effects, and column (3) includes school fixed effects.[29] Standard errors are clustered at the level of the fixed effect. The log of total time use is positively and significantly related to the evaluation scores.[30]

We present the results for language test scores based on the same specification in Table 7 Panel B. Column (1) again includes county fixed effects, column (2) includes district fixed effects, and column (3) includes school fixed effects. Once again, standard errors are clustered at the level of the fixed effect. Student test scores are also positively and significantly related to the probability that a teacher receives an 'excellent' or 'good' evaluation score. These two relationships provide evidence that variation in evaluation scores reflects differences in actual effort exerted.

5.4 Determinants of Effort

To test the remaining predictions, we estimate two regression specifications for the determinants of teacher effort based on equation (3) – one that controls for county fixed effects and another that controls for individual teacher fixed effects. The first enables us to test the impact on performance of wage differences across ranks (Prediction 1), (historical) promotion rates (Prediction 2), relative skill (ability) (Prediction 3), and the number of teachers in the contest (Prediction 4). The second enables us to control for individual heterogeneity in ability and motivation to identify the impact on performance of time to promotion eligibility and changes in tournament conditions over time (Predictions 5 and 6), and also provides a second test of Predictions 3 and 4 using panel variation.

S_x is the variance of test scores, which is a constant that can be calculated directly from the test score data x. Using the above equation with the variance terms, we can then calculate S_ϵ for each observation and take its mean as a scalar estimate of S_ϵ for the sample. Now, we can estimate S_y as a function of N and three constants, using $S_y = \beta^2(S_x/N) + S_\epsilon$. Then, we can correct for heteroscedasticity by reweighting the sample by $1/\sqrt{S_y}$. This also is roughly equivalent to weighting the observations by $1/\sqrt{N}$, which produces nearly identical results.

[29] In the school fixed effects column, we lose observations because the Probit regression drops all cases in which all of the outcomes in a particular school are either positive or negative.

[30] We also run analogous regressions using each of the different categories of teacher time use separately, and find that the categories of the number of hours spent grading, tutoring, and disciplining students are most impactful.

In these regressions, we interpret the annual performance evaluation scores to be a measure of effort. The evaluation scores are the only measure for which multiple years of data are available, which is necessary for estimating the individual fixed effects specification.[31] As noted in the model description, evaluation scores also may reflect teachers' skill, so interpreting evaluation scores as a measure of effort requires that we control adequately for differences in teachers' skills in the regressions. Individual fixed effects control for all time-invariant skill differences across individuals, and in specifications without individual fixed effects, teacher skill can be controlled for separately using measures of education, experience, and teacher ability (as measured by placement tests). Once again, we focus on primary 1, middle 2, and middle 1 rank levels.

The first equation is specified as follows:

$$ev^*_{irct} = \zeta_0 + \zeta_1 \Delta w_{rc} + \zeta_2 \bar{p}_{rc} + \zeta_3 a_{10,ir\tau t} + \zeta_4 a_{90,ir\tau t} + \zeta_5 ln(n_{r\tau t}) + \zeta_6 ln(n_{r\tau t}) * a_{10,ir\tau t} + \zeta_7 ln(n_{r\tau t}) * a_{90,ir\tau t} + \zeta_8 X_{irct} + \zeta_9 X_{crt} + \gamma_c + \upsilon_{irct} \quad (5)$$

Here, ev^*_{irct} is again the evaluation score of teacher i at rank r in county c in year t. Δw_{rc} is the increase in wages if promoted to the next rank in the county, $\bar{p}_{rc}$ is the historical promotion rate for those with rank r in county c, $n_{r\tau t}$ is the number of teachers in the same rank in the same education district (τ), $a_{10,ir\tau t}$ and $a_{90,ir\tau t}$ are indicator variables for whether the ability of teacher i is below the 10th percentile or above the 90th percentile of teachers in the same rank in the same district, X_{irct} are teacher level controls, X_{crt} are district-rank controls as before, γ_C captures county-level unobserved, time-invariant factors, and υ_{irct} is an error term which includes omitted individual and district characteristics. Note that contests are between teachers in the same district, but the wages and promotion rates are set at the county level.

We expect that ζ_1 will be positive and significant (Prediction 1). This would tell us that promotion contests with higher wage gaps elicit higher levels of performance. This coefficient

[31] We treat the evaluation score as a continuous measure, rather than estimating an ordered probit model. This is because ordered probit (and ordered logit) cannot be estimated with teacher fixed effects, and because probit regressions of the same form where the outcomes are a dummy variable for receiving an excellent evaluation score, or for receiving an excellent or good evaluation score, yield consistent results.

is identified from different wage increases between rank levels within the same county (again, teacher *i's* wage is excluded from the calculation of the average for her current rank). Assuming that the ability distribution has a bell-shaped distribution, we expect that promotion rates nearer one half will elicit higher average performance than promotion rates further from one half (Prediction 2). To test this, we include dummy variables for four quintiles of promotion rates, omitting the middle quintile. We expect that these coefficients will be negative. We also expect that teachers with ability levels nearer the marginal skill percentile (equal to $1 - \bar{p}$) will exhibit better performance (Prediction 3). As such, we include dummy variables for teachers in the top and bottom 10% of the ability distribution, since these teachers are expected to be far from the marginal skill percentile. Since those at different extremes of the skill distribution may differ in the distance to the marginal skill percentile and have different intrinsic motivation, we include separate variables for the two extremes. The promotion model predicts that the coefficients on these dummies will be negative. Finally, we expect that the further a teacher is from the marginal skill percentile, the more performance will decrease with more teachers in the contest (Prediction 4). Thus, we interact skill (the dummy variables for teachers in the top and bottom 10% of skill) and the number of teachers (in logs). We expect these coefficients to be negative. We include the same teacher and county-rank level controls as in Table 6.

Table 8 presents the results from estimating equation (5). County fixed effects are controlled for, and standard errors are clustered at the county level. We find that average levels of performance are higher with higher wage gaps. The coefficient on the difference in log wages from one rank to another is positive and significant, confirming Prediction 1. On average, a 10% increase in the 'prize' increases the evaluation score by 0.04.

We also find evidence for Prediction 2; that average performance is higher the closer is p^* to 1/2. Since the omitted category is a dummy variable for promotion rates in the middle quintile, we expect performance incentives to be highest here, and to be lower for both lower and higher promotion rates. The coefficients on the dummies for the first, fourth and fifth quintiles are negative and that for the fourth quintile is marginally significant. The coefficient on the promotion rate in quintile 2 is of the opposite sign, but is not statistically different from zero. The coefficients on ability interacted with the number of teachers competing are imprecisely estimated. In addition, the education test score positively and significantly predicts higher evaluation scores, confirming that it captures a dimension of ability that predicts better performance.

Next, we estimate a second specification for the determinants of evaluation scores that controls for individual teacher fixed effects. This specification has the great advantage of controlling for time-invariant individual-level heterogeneity in skill or motivation. The estimating equation is the following:

$$ev^*_{irct} = \delta_1 \boldsymbol{D_{irct}} + \delta_2 a_{10,ir\tau t} + \delta_3 a_{90,ir\tau t} + \delta_4 ln(n_{irct}) + \delta_5 ln(n_{irct}) * a_{10,ir\tau t} + \delta_6 ln(n_{irct}) * a_{90,ir\tau t} + \phi_{irc} + v_{irct}. \tag{6}$$

Here, $\boldsymbol{D_{irct}}$ is a set of dummy variables for years before and after eligibility for promotion (accounting for the level of education), ϕ_{irc} is the individual teacher fixed effect, and v_{irct} is the error term. In $\boldsymbol{D_{irct}}$ we include dummy variables separately for each year from $t = X - 7$ and $t = X - 6$ to test whether effort incentives are zero when a teacher is more than five years before promotion eligibility (Prediction 5). We omit the dummy variable for $t = X - 5$. We include dummy variables from $t = X - 4$ to $t \geq X + 15$. With the reference period being $t = X - 5$, the period five years before promotion eligibility, we expect that the coefficients on the time dummy variables for $t = X - 4$ to $t = X$ should be increasing (Prediction 5). After many years of promotion eligibility, we expect that the coefficients will begin declining (Prediction 6). Thus, we predict an inverted-U shape in performance with respect to years to eligibility in a given rank.

The difference in wages and the promotion rate drop out with the inclusion of individual fixed effects because they do not vary for individual teachers over time.[32] Note that by controlling for teacher fixed effects, the coefficients on the ability variables and $ln(n)$ (plus their interaction) are identified from changes in the relative ability of teachers and the number of teachers over time as the pool of competitors changes (due to retirement and promotions). The inclusion of $ln(n)$

[32]In practice, wage rates and promotion rates could change over time. However, in our data we can only estimate wage gaps at the time of the survey and historical promotion rates, see Section 4.

interacted with the ability dummies provide a second test of Prediction 4 (that as the number of competitors increases, the effort levels of teachers far from the marginal skill percentile decrease). We expect the coefficients on these interaction terms to be negative ($\delta_5 < 0$ and $\delta_6 < 0$).

For this regression, which focuses on the behaviour of teachers over time, the inclusion of teacher fixed effects is essential to account for selection. The dummy variables for years after promotion eligibility would be biased without the teacher fixed effect, as it is likely that worse teachers would take longer to be promoted. However, even the coefficients on the dummy variables for the years preceding eligibility for promotion can be biased. The number of years that a teacher must wait before becoming eligible for promotion depends upon their education level. A teacher's education level is not exogenous and can also change if teachers earn degrees while teaching. As a result, we believe that any effort regressions including dummy variables for years after promotion eligibility should include a teacher fixed effect to account for selection.

The results of estimating equation 6 are reported in Table 9. Column (1) reports the coefficients, column (2) reports standard errors clustered at the teacher level (the level of the fixed effect), and column (3) reports two-way clustered standard errors at the teacher and district-rank levels.[33] Results are very consistent across both specifications. Confirming Prediction 4, Table 9 shows that the coefficients on the number of teachers interacted with ability in the top and bottom 10% of the ability distribution are negative, and the coefficient on the number of teachers interacted with ability in the top decile is marginally significant.

Figure 2 plots the coefficients on the time dummy variables from Table 9. The pattern of the dummy variables corresponds to Predictions 5 and 6: an inverted-U shape. The omitted category is $t = X - 5$, where we expect weak effort incentives. Consistent with Prediction 5, the coefficients on $t = X - 7$ and $t = X - 6$ are not statistically different from zero. The coefficients on the dummy variables from $t = X - 4$ to $t = X$ show that performance is increasing, apart from the first year ($t = X - 4$). As teachers get closer to eligibility for promotion, they work harder. The evaluation scores when $t = X - 2$, $t = X - 1$, and $t = X$ are significantly different than the time period $t = X - 5$, at the 5%, 1%, and 5% levels, respectively. The average evaluation score five years or more before eligibility ($t = X - 5$) is 2.5 (between a 'pass' and 'good' score). In $t = X - 2$ and $t = X - 1$, the evaluation score is more likely to be a 'good' or

[33]We also estimate a version that includes two-way clustered standard errors at the teacher and county-rank levels, and the results are almost identical, so we omit them here.

'excellent' score.

We may be concerned that the increasing trend in performance before promotion eligibility may be teachers learning over time as they are in the first few years of teaching at a specific rank level. However, in the pre-eligibility period, the average experience of primary 1, middle 2 and middle 1 teachers is 14, 9 and 14 years, respectively. These are not new teachers and they are performing the same job as in previous rank levels. Other research finds that teacher learning from experience comes in the initial few years of teaching (Hanushek and Rivkin, 2006). Further, the results are robust to dropping all teachers with fewer than five years of experience (7% of teachers in our sample).

Another concern is that principals may manipulate evaluation scores of teachers who are nearing promotions, awarding teachers who are becoming eligible with high evaluation scores so that they can be promoted, and for this reason the scores do not reflect true performance. Such manipulation could respond to promotion incentives in ways that are similar to effort incentives. For example, manipulation also might be expected to lead to an upward trend in scores in the years just before promotion eligibility. Other studies provide evidence against evaluation scores being manipulated in China. Park and Hannum (2001) showed that rank is a significant correlate in a regression of student test scores on various teacher characteristics using data from Gansu. Ding and Lehrer (2007a) find that teachers' ranks explain a much greater share of variation in student test scores than education or years of experience, also using Chinese data. The fact that in our data, some teachers in these rank levels are never promoted is also evidence that promotions are not inevitable, and that evaluations are not manipulated to ensure that everyone is promoted. Finally, because evaluation scores are announced at annual teacher meetings and because principals are also evaluated and promoted based on the performance of their school, this should reduce the incentive to inflate evaluation scores.[34] Finally, our tests in Section 5.3 show that evaluation scores are a good proxy for effort.

Figure 2 also shows that many years after promotion eligibility, performance declines. In fact, thirteen years following promotion eligibility, evaluation scores are on average almost two points lower (which is very large for a four point scale), and this difference is significant at

[34]Another possible type of manipulation is principals assigning teachers to easier or better classes or grades within the school so that their student's test scores are higher. However, in Gansu, in primary school there is generally only one class per grade (Park and Hannum, 2001), making such manipulation unlikely.

the 1% level. This provides positive evidence for Prediction 6, suggesting that teachers may be revising downwards their estimates of their relative skill, or responding to the reduction in fewer future periods they will be able to enjoy additional income from promotion.

We subject these results to a variety of robustness checks. First, we check that the results for the ability dummy variables are not sensitive to small changes in the cut-offs. We re-estimate the effort regressions using the top and bottom 15% of the ability distribution instead of the top and bottom 10%. The results are contained in Table 10, Column (1). The pattern of coefficients on the time dummy variables is preserved, and the interpretation of the main results is unaffected. We also check that the results are robust to the inclusion of the number of teachers variable, since this is a variable for which some values have been adjusted. Equation (6) is estimated again omitting schools with low sampling rates (below 25% and below 35%). Here as well, the magnitude of the time before and after promotion eligibility dummy coefficients changes slightly but not the signs (see Table 10 columns (2) and (3)). The coefficients on the log number of teachers interacted with the top and bottom 10% of the ability distribution are still negative. Most importantly however, the pattern of the time dummy variables remains unchanged, and the interpretation of the effect of the number of teachers also remains consistent with the previous results.

We test for heterogeneity in results for different groups. The regressions were conducted separately for male and female teachers (results not reported), and no differences were found. We also check whether Communist party membership makes a difference. Controlling for party membership does not change the results, and separating the results by party and non-party members shows an inverted-U pattern for both (results not reported).

Because the evaluation score data are recorded on a four-point scale, we also estimate the effort regression as an ordered probit model. However, it is not possible to include fixed effects in ordered probit models (Greene and Hensher, 2010), so we include only observations for the years before promotion eligibility in order to reduce concerns of selection bias associated with only non-promoted teachers remaining in the sample. Appendix Table 13 presents the results, which are consistent with the full model in Table 9.

In order to test Prediction 7, that performance incentives no longer exist after the highest promotion, we run an analogous regression of evaluation scores on time dummy variables for primary and middle high rank levels. Primary and middle high levels are the highest rank levels a teacher can achieve in primary and middle school; there are no more promotions after this. They are difficult ranks to attain, and becoming a '*gaoji*' (high level) teacher carries a great amount of prestige. Instead of pre- and post-promotion eligibility time dummy variables, we use dummy variables for the number of years (two to fifteen) since being promoted to the highest rank (primary high or middle high rank levels). Table 11 contains results of the regression for primary and middle high ranks, and the top panel of Figure 3 plots the time dummy variable coefficients for the number of years into the rank. The coefficients are steadily decreasing. The longer a teacher stays in these ranks, the lower is performance. This result could be consistent with lack of incentives in this rank, although the response is more gradual than theory might predict, perhaps reflecting gradual adjustment by teachers or the fact that these teachers are nearing retirement age and slowing down. By the time a teacher has been at primary or middle high level for 14 years, the evaluation scores are more than one point lower, and this effect is significant at the 1% level. At the primary and middle high levels, there are no predictions on the effect of being at the extremes of the skill distribution or the number of competitors, and as shown in Table 11, these coefficients are all insignificant.

One also might be worried about manipulation if principals decide that excellent and good evaluations should not be 'wasted' on high rank teachers. However, similar to the results for teachers with lower ranks, teachers with primary high and middle high rank levels exhibited a pattern of time use on teaching activities that mirrored the evaluation scores. In this case, total time spent on teaching activities declined with more years at the highest rank levels. To show this, we regress the log of time spent on teaching activities against the number of years into primary and middle high rank with the same specification as for evaluation scores, but only for the year 2007 (since that is the year for which we have time use data). We plot the coefficients on the number of years into the high rank levels in the bottom panel of Figure 3.[35] The log total number of hours spent on teaching activities declines over time.

[35]This result is not driven by high rank teachers spending a disproportionate amount of time on tutoring students outside of school. Less than 10% of total time is spent on tutoring, and high rank teachers spend approximately the same proportion of time tutoring as do teachers at lower rank levels.

6 Conclusion

In this study, we have examined the incentive properties of China's promotion system for public employees, introducing China's complex system of annual evaluations and promotions of primary and middle school teachers. As far as we are aware, this is the first paper to provide empirical evidence linking promotion incentives to work effort in the public sector, extending a limited body of previous research focused on the private sector. The Chinese system's reliance on promotion incentives also stands as an alternative model to pay for performance schemes in incentivising teacher performance, although to date it has received little attention.

Theoretical predictions from a tournament model of promotions are largely verified by the empirical findings. Promotions are associated with significant salary increases. Higher annual evaluation scores significantly increase the probability of being promoted. These two conditions suggest that the promotion system should motivate effort. Teachers exert more effort, measured by annual evaluation scores, when wage increases are greater. Average performance is greater when the promotion probability is closer to 50%, and increasing the number of competitors leads to relatively lower performance among those at the extremes of the skill distribution. Teacher performance increases over time as teachers get closer to the period of eligibility for promotion, and decreases if they are repeatedly passed over for promotion. Teachers who reach the highest rank exhibit a steady decline in performance. The evaluation scores are correlated with student test scores and teacher time spent on teaching, which is inconsistent with evaluation scores simply being manipulated to help teachers get promoted or for other political reasons. This is also evidence that evaluation scores are an appropriate proxy for effort.

The findings in this paper suggest that teachers respond to the incentives embedded in the design of the promotion system. One difference between China's promotion scheme and typical pay-for-performance schemes is that evaluations are based on a diversity of criteria and information sources (including test scores, peer and student assessments, classroom observation, evidence of innovation in educational practice or research, etc.) which are less likely than pay-for-performance schemes to lead teachers to focus only on test performance at the expense of other learning objectives. Unlike pay-for-performance schemes that reward recent performance, the promotion system provides dynamic incentives that may better enable teachers to set career goals, work hard to realize their career ambitions, and encourage them to make longer-term investments to becoming better teachers.

At the same time, the promotion system fails to provide strong effort incentives for some types of teachers. Those at the extreme ends of the skill distribution may perceive that their effort is unlikely to substantially influence their chances of being promoted, which is especially concerning for students of the lowest ability teachers. Teachers may exert less effort if they are far from being eligible for promotion or if they are repeatedly passed over for promotion. Teachers who reach the highest rank levels may lack incentives to continue to exert high levels of effort. For such teachers, there may be a role for more immediate incentives (e.g., bonus pay) or other reforms of the existing promotion system (e.g., creating additional rank levels, shortening the time to promotion eligibility, etc.). Given the diversity of teacher ability and effort incentives, it could be optimal to combine immediate performance incentives with promotion incentives as argued by Kwon (2006).

Acknowledgements

We thank Matthew Collin, Jishnu Das, Stefan Dercon, Eliana La Ferrara, Paul Glewwe, Clare Leaver, Lara Lebedinski, Derek Neal, Simon Quinn, Margaret Stevens, Francis Teal, Xuehui An, and seminar participants at Oxford, Bocconi, CSAE, CUHK, HKUST, IFPRI, ANU, Monash, UNSW, Fudan, and Peking University for helpful comments.

This paper was supported by the ESRC/DFID [RES167250250], the European Commission [214705 PITN-GA-2008-214705], and the Hong Kong Research Grants Council General Research Fund [692713].

References

Acemoglu, D., J. A. Robinson, and D. Woren (2012). *Why nations fail: the origins of power, prosperity, and poverty*, Volume 4. SciELO Chile.

Ashraf, N., O. Bandiera, and S. S. Lee (2014). Do-gooders and go-getters: career incentives, selection, and performance in public service delivery. Technical report, Suntory and Toyota International Centres for Economics and Related Disciplines, LSE.

Baker, G., M. J. Gibbs, and B. Holmstrom (1994). The internal economics of the firm: evidence from personnel data. *Quarterly Journal of Economics 109*(4), 881–919.

Barlevy, G. and D. Neal (2012). Pay for Percentile. *American Economic Review 102*(5), 1805–1831.

Campbell, D. (2008). Nonfinancial performance measures and promotion-based incentives. *Journal of Accounting Research 46*(2), 297–332.

Chan, K. W., V. Henderson, and K. Y. Tsui (2008). Spatial Dimensions of Chinese Economic Development. In T. Rawski and L. Brandt (Eds.), *China's Great Transformation: Origins, Mechanisms, and Consequences of the Post-Reform Economic Boom*, pp. 776–828. Cambridge University Press.

Cowgill, B. (2015). Competition and productivity in employee promotion contests.

Dal Bó, E., F. Finan, and M. A. Rossi (2013). Strengthening state capabilities: The role of financial incentives in the call to public service. *The Quarterly Journal of Economics 128*(3), 1169–1218.

Ding, W. and S. F. Lehrer (2007a). Do peers affect student achievement in China's secondary schools? *The Review of Economics and Statistics 89*(2), 300–312.

Ding, W. and S. F. Lehrer (2007b). The optimal policy to reward the value added by educators: Theory and evidence from China. In E. Hannum and A. Park (Eds.), *Education and Reform in China*. Routledge.

Dixit, A. (2002). Incentives and organizations in the public sector: An interpretative review. *Journal of Human Resources*, 696–727.

Ederer, F. (2010). Feedback and motivation in dynamic tournaments. *Journal of Economics & Management Strategy 19*(3), 733–769.

Enikolopov, R. (2011). Are bureaucrats really paid like bureaucrats? Technical report.

Fairburn, J. A. and J. M. Malcomson (2001). Performance, promotion, and the Peter Principle. *The Review of Economic Studies 68*(1), 45–66.

Ferraz, C. and F. Finan (2009). Motivating politicians: The impacts of monetary incentives on quality and performance. Technical report, National Bureau of Economic Research.

Gibbs, M. J. (1989). *Promotions, compensation, and firm organization.* Ph. D. thesis, University of Chicago, Department of Economics.

Gibbs, M. J. (1995). Incentive compensation in a corporate hierarchy. *Journal of Accounting and Economics 19*(2), 247–277.

Glewwe, P. (2002). Schools and skills in developing countries: Education policies and socioeconomic outcomes. *Journal of Economic Literature 40*(2), 436–482.

Glewwe, P., N. Ilias, and M. Kremer (2010). Teacher incentives. *American Economic Journal: Applied Economics 2*(3), 205–227.

Glewwe, P. and K. Muralidharan (2015). Improving school education outcomes in developing countries: evidence, knowledge gaps, and policy implications. *University of Oxford, Research on Improving Systems of Education (RISE).*

Greene, W. H. and D. A. Hensher (2010). *Modeling ordered choices: A primer.* Cambridge University Press.

Haeck, C. and F. Verboven (2012). The internal economics of a university: Evidence from personnel data. *Journal of Labor Economics 30*(3), 591–626.

Han, L. (2013). Is centralized teacher deployment more equitable? Evidence from rural China. *China Economic Review 24*, 65–76.

Han, L., M. Liu, and X. An (2014). Centralized deployment and teachers incentive: Evidence from reforms in rural China. *mimeo, HKUST*.

Hanushek, E. A. and S. G. Rivkin (2006). Teacher quality. *Handbook of the Economics of Education 2*, 1051–1078.

Holmstrom, B. and P. Milgrom (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization*, 24–52.

Jacob, B. A. and S. D. Levitt (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *Quarterly Journal of Economics 118*(3), 843–877.

Kwon, I. (2006). Incentives, wages, and promotions: theory and evidence. *RAND Journal of Economics 37*(1), 100–120.

Landry, P. F. (2008). *Decentralized Authoritarianism in China: the Communist Party's control of local elites in the post-Mao era*. Cambridge University Press New York.

Li, H. and L.-A. Zhou (2005). Political turnover and economic performance: the incentive role of personnel control in China. *Journal of Public Economics 89*(9), 1743–1762.

Macartney, H. (2014). The dynamic effects of educational accountability. Technical report, National Bureau of Economic Research.

McEwan, P. J. (2015). Improving learning in primary schools of developing countries a meta-analysis of randomized experiments. *Review of Educational Research 85*(3), 353–394.

Medoff, J. L. and K. G. Abraham (1980). Experience, performance, and earnings. *Quarterly Journal of Economics 95*(4), 703–736.

Muralidharan, K. and V. Sundararaman (2011). Teacher performance pay: Experimental evidence from india. *Journal of Political Economy 119*(1), pp. 39–77.

OECD (2014). *PISA 2012 Results in Focus: What 15-year-olds know and what they can do with what they know*. OECD.

Park, A. and E. Hannum (2001). Do teachers affect learning in developing countries? Evidence from matched student-teacher data from China. In *Conference Rethinking Social Science Research on the Developing World in the 21st Century, Park City, Utah, June*, pp. 7–11.

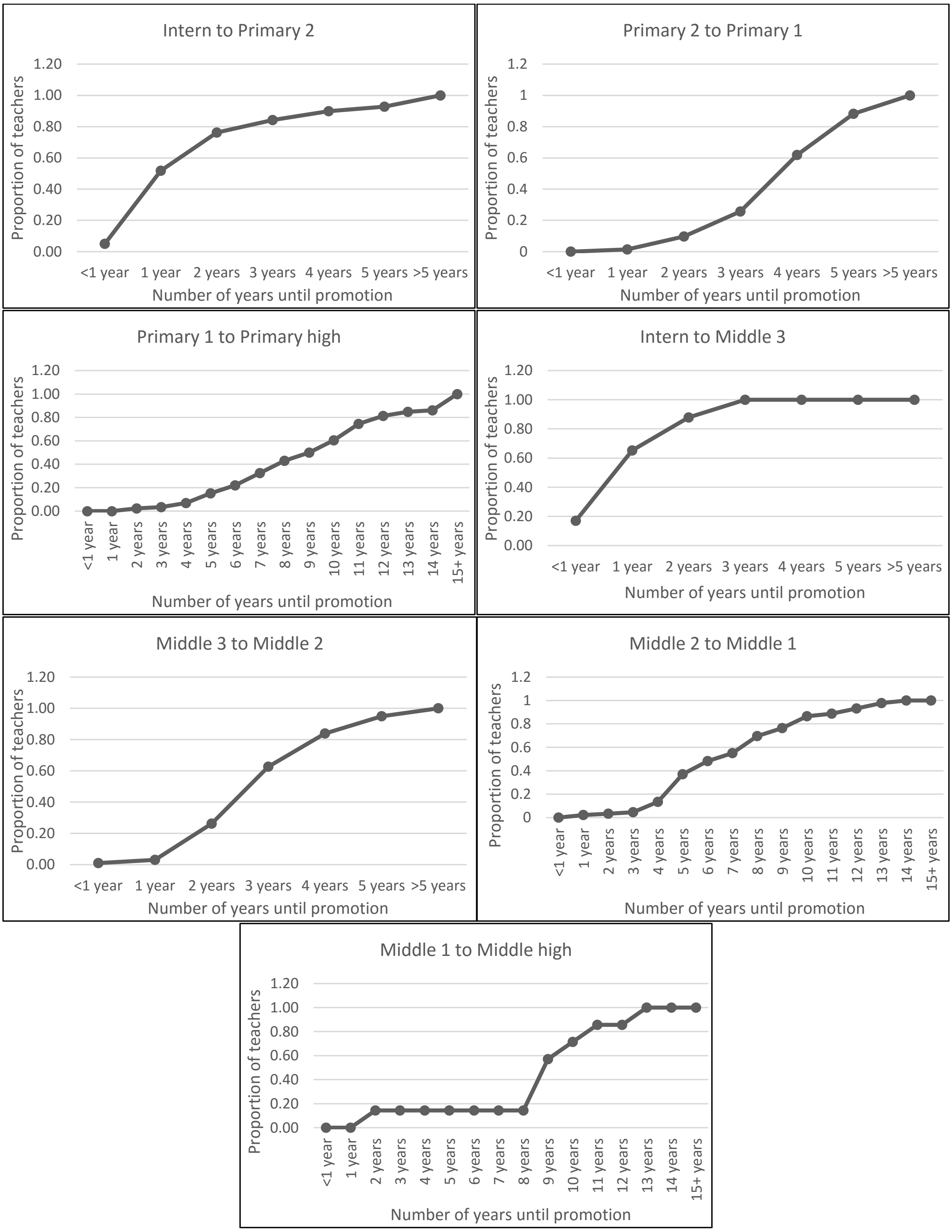
Podgursky, M. J. and M. G. Springer (2007). Teacher performance pay: A review. *Journal of Policy Analysis and Management 26*(4), 909–950.

Reback, R., J. Rockoff, and H. L. Schwartz (2014). Under pressure: Job security, resource allocation, and productivity in schools under no child left behind. *American Economic Journal: Economic Policy 6*(3), 207–241.

World Bank (2004). World Development Report: Making Services Work for Poor People. Technical report, World Bank.

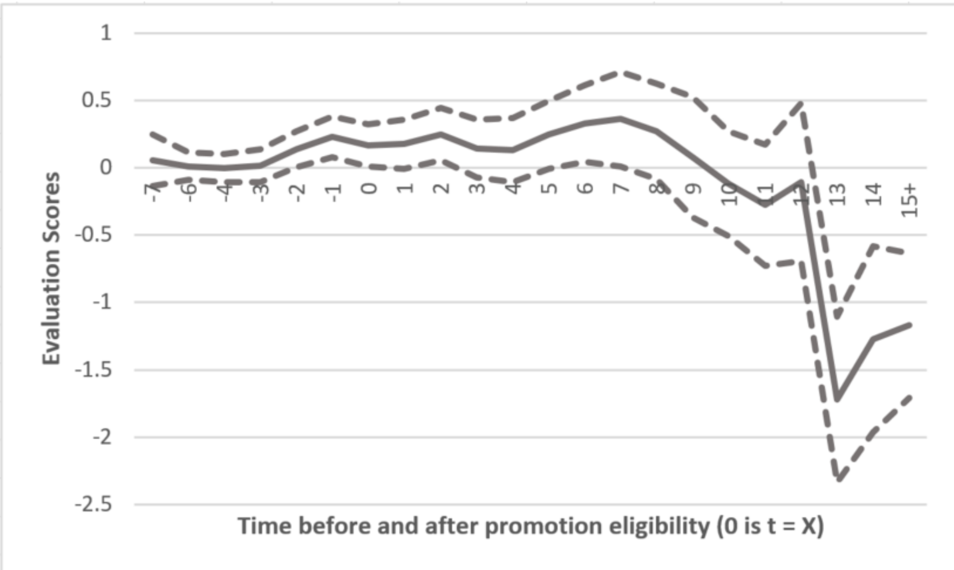
A Appendix: Figures

Figure 1: Number of Years to Promotion



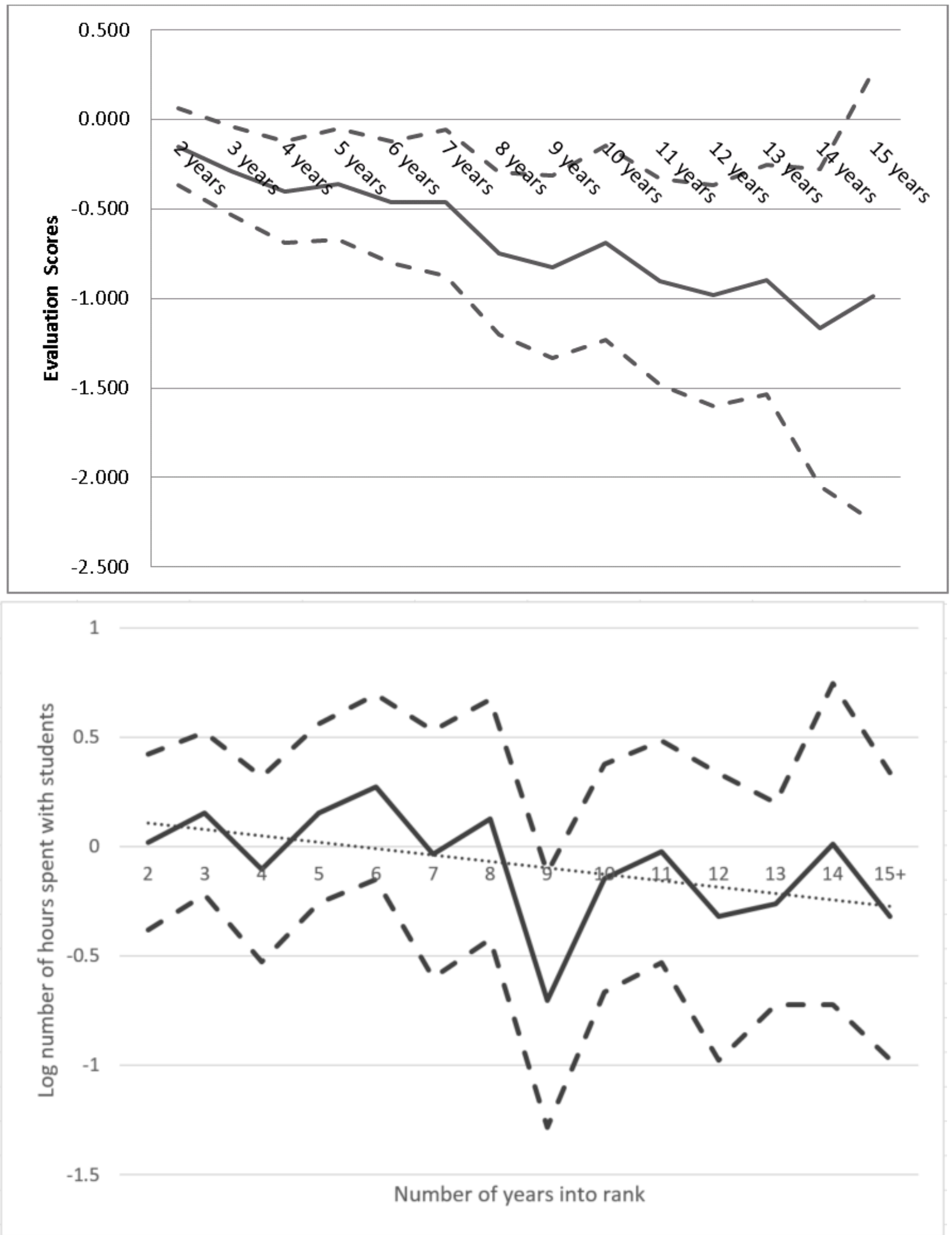
Source: Gansu 2007 Data.

Figure 2: Primary 1, middle 2 and middle 1, pre- and post-eligibility time dummies



Notes: Dashed lines are 95% confidence intervals.

Figure 3: Primary and middle high time dummies and log of total number of hours spent per week on teaching activities



Notes: Dashed lines are 95% confidence intervals. Total number of hours spent per week on teaching activities is the sum of the hours per week spent on teaching classes, grading, lesson plans, research, extra-curricular activities, home visits, and discipline.

B Appendix: Tables

Table 1: Characteristics of teachers in the sample, Round 3 (2007)

	Intern	Primary 2	Primary 1	Primary high	Middle 3	Middle 2	Middle 1	Middle high	All teachers
Number of teachers									
Total	135	141	544	348	127	516	277	12	2,100
Proportion Female	0.51	0.61	0.47	0.25	0.44	0.32	0.14	0.14	0.36
Basic Characteristics									
Average age	25.39	28.14	37.74	47.66	27.44	32.48	41.03	45.00	36.18
	(6.47)	(8.10)	(8.64)	(5.91)	(6.11)	(4.93)	(6.46)	(5.92)	(9.54)
Average years teaching	5.96	7.10	17.27	27.40	4.92	10.49	19.84	26.14	15.55
	(9.61)	(7.69)	(8.61)	(5.42)	(4.23)	(5.23)	(6.67)	(5.11)	(9.72)
Education - middle school	0.01	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.01
Education - vocational high school	0.09	0.29	0.25	0.49	0.05	0.03	0.10	0.00	0.19
Education - high school	0.03	0.02	0.09	0.23	0.00	0.00	0.02	0.00	0.06
Education - vocational college	0.73	0.61	0.52	0.23	0.81	0.52	0.51	0.29	0.52
Education - four year degree	0.13	0.08	0.12	0.04	0.14	0.44	0.37	0.71	0.22
Education placement test	-0.03	0.04	0.04	0.07	-0.28	-0.01	0.00	0.78	0.00
	(1.03)	(0.87)	(0.95)	(0.93)	(1.00)	(1.01)	(1.01)	(0.78)	(0.98)
Psychology placement test	-0.08	0.12	0.07	0.07	-0.32	-0.03	-0.05	0.77	0.00
	(0.99)	(0.94)	(0.91)	(0.93)	(1.03)	(1.02)	(1.05)	(0.69)	(0.98)
Language placement test	0.02	0.08	-0.01	0.08	0.02	-0.04	0.01	0.12	0.01
	(1.00)	(0.96)	(0.96)	(0.97)	(0.97)	(1.02)	(1.00)	(1.63)	(0.99)
Proportion published a paper	0.10	0.33	0.48	0.33	0.31	0.51	0.57	1.00	0.43
Proportion won a teaching award	0.39	0.74	0.84	0.87	0.67	0.81	0.89	1.00	0.79
Wages									
Monthly wage (CNY)	986.34	1,077.19	1,316.71	1,535.03	1,160.06	1,313.01	1,571.01	1,851.43	1,335.12
	(252.48)	(189.36)	(215.36)	(257.68)	(198.41)	(215.54)	(240.39)	(329.26)	(284.07)
Variance of monthly wage	63,748.12	35,856.60	46,381.78	66,401.35	39,367.09	46,458.80	57,785.92	108,414.30	51,144.41

Note: Interns include Primary and Middle school interns. Standard deviations in parentheses.

Table 2: Teacher Time Use, Round 3 (2007)

	Intern	Primary 2	Primary 1	Primary high	Middle 3	Middle 2	Middle 1	Middle high	All teachers
Hours per week teaching classes	15.63	19.12	18.03	17.73	14.34	13.66	13.11	10.17	15.91
	(5.89)	(5.67)	(5.87)	(6.65)	(5.11)	(4.03)	(4.27)	(4.69)	(5.80)
Hours per week spent grading	10.62	11.99	9.63	9.26	10.09	9.12	9.12	7.50	9.61
	(5.33)	(6.65)	(4.87)	(5.48)	(6.64)	(4.72)	(4.56)	(4.98)	(5.24)
Hours per week spent on lesson plans	9.83	9.91	9.31	9.15	9.83	9.18	9.69	8.25	9.40
	(6.43)	(5.06)	(5.07)	(5.70)	(5.84)	(4.92)	(4.93)	(5.63)	(5.27)
Hours per week spent on teaching research	3.12	3.38	2.89	2.74	2.73	2.90	2.81	2.00	2.89
	(2.87)	(2.68)	(2.13)	(2.22)	(2.31)	(2.95)	(1.93)	(1.13)	(2.45)
Hours per week spent on tutoring	5.33	5.59	4.75	4.40	5.87	4.97	4.70	2.83	4.89
	(3.43)	(3.79)	(3.28)	(3.86)	(3.98)	(3.74)	(3.60)	(2.08)	(3.64)
Hours per week spent on extra-curriculars	3.31	3.26	2.98	2.83	2.65	2.59	2.51	1.58	2.81
	(4.56)	(2.50)	(2.21)	(2.73)	(2.30)	(2.27)	(1.89)	(1.83)	(2.53)
Hours per week spent on home visits	2.35	2.35	1.89	1.76	2.53	1.96	1.80	0.63	1.97
	(3.65)	(2.23)	(2.48)	(1.63)	(2.83)	(2.34)	(2.15)	(0.64)	(2.39)
Hours per week spent on discipline	2.94	2.73	2.51	2.27	3.48	3.23	3.29	2.08	2.85
	(2.90)	(1.78)	(2.56)	(2.02)	(3.94)	(2.84)	(3.25)	(2.02)	(2.76)
Total hours per week on teaching activities	52.99	58.05	51.78	50.06	51.19	47.55	46.81	35.04	50.17
	(20.73)	(18.79)	(15.89)	(20.37)	(19.78)	(16.15)	(15.46)	(16.06)	(17.75)

Note: Interns include Primary and Middle school interns. Standard deviations in parentheses.

Table 3: Evaluation scores of teachers by rank (2003-2006)

	Interns		Primary 2		Primary 1		Primary high		Middle 3		Middle 2		Middle 1		Middle high	
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%
Fail	-	-	-	-	4	0.22	2	0.2	-	-	5	0.31	-	-	-	-
Pass	123	60.89	211	42.89	913	51.18	511	51.93	271	53.45	877	55.09	452	53.87	8	38.1
Good	53	26.24	194	39.43	608	34.08	339	34.45	163	32.15	469	29.46	248	29.56	10	47.62
Excellent	26	12.87	87	17.68	259	14.52	132	13.41	73	14.4	241	15.14	139	16.57	3	14.29
Total	202	100	492	100	1,784	100	984	100	507	100	1,592	100	839	100	21	100

Note: Interns include Primary and Middle school interns.

Table 4: Wage Premia - Primary School

	Outcome: Log of monthly wage, 2007		
	(1)	(2)	(3)
Primary 1	0.204***	0.084***	0.075***
	(0.030)	(0.023)	(0.025)
Primary high	0.361***	0.150***	0.132***
	(0.034)	(0.030)	(0.033)
Experience		0.021***	0.022***
		(0.004)	(0.004)
Experience squared		-3.32E-04***	-3.12E-04***
		(9.76E-05)	(9.35E-05)
Education - middle school			0.037
			(0.025)
Education - vocational high school			0.099***
			(0.015)
Education - high school			0.048***
			(0.011)
Education - vocational college			0.076
			(0.050)
Education - four year degree			0.090***
			(0.021)
Performance evaluation score, 2006			0.001
			(0.005)
Observations	1,029	1,029	1,026
R^2	0.400	0.543	0.553

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Regressions include county fixed effects and clustered standard errors at the county level. Performance evaluation score is measured on a scale of 1-4 (fail, pass, good, excellent).

Table 5: Wage Premia - Middle School

	Outcome: Log of monthly wage, 2007		
	(1)	(2)	(3)
Middle 2	0.134***	0.047*	0.040*
	(0.030)	(0.024)	(0.021)
Middle 1	0.318***	0.127***	0.119***
	(0.035)	(0.025)	(0.023)
Middle high	0.535***	0.301***	0.282***
	(0.080)	(0.063)	(0.058)
Experience		0.021***	0.022***
		(0.004)	(0.004)
Experience squared		-3.32E-04***	-3.47E-04***
		(9.67E-05)	(1.13E-04)
Education - vocational high school			-0.016
			(0.018)
Education - high school			-0.035
			(0.041)
Education - vocational college			0.090
			(0.112)
Education - four year degree			0.002
			(0.022)
Performance evaluation score, 2006			-0.003
			(0.005)
Observations	930	930	926
R^2	0.390	0.493	0.494

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Regressions include county fixed effects and clustered standard errors at the county level. Performance evaluation score is measured on a scale of 1-4 (fail, pass, good, excellent).

Table 6: Promotions and Evaluation Scores

Outcome: Promotion of teacher i in year t	(1) coef/se	(2) coef/se	(3) coef/se
Evaluation score	0.012***	0.010***	0.011***
	(0.004)	(0.004)	(0.004)
$t = X - 5$ to $t = X - 1$ dummy variable	-0.007	-0.009*	-0.011*
	(0.005)	(0.006)	(0.006)
$t \geq X$ dummy variable	0.022**	0.020*	0.021**
	(0.009)	(0.010)	(0.010)
Constant	-0.060*	-0.070**	-0.116**
	(0.033)	(0.034)	(0.052)
Number of observations	3,816	3,816	3,816
R^2	0.050	0.053	0.049
Fixed effect	County	District	School

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Regressions include standard errors clustered at the level of the fixed effect. Regressions inlucde controls for teacher education levels, experience, gender, the proportion of teachers promoted, teacher ability index (with dummies for the top and bottom 10%), the log number of teachers in the district at that rank, the log number of teachers interacted with ability in the top and bottom 10%, county-rank level variables (average education levels, proportion female, average experience, and average experience interacted with the proportion female). Regressions also include rank level and time dummy variables. Proportion promoted is the proportion of teachers in a county that are promoted at that rank within a 'reasonable' number of years. We first calculate the time it took for each teacher to be promoted from one rank to another. Then we calculate the number of years for which, for each rank and education level in a county, half of the teachers were promoted within that time. Then for each rank and county we calculate the proportion of teachers promoted within that 'reasonable' time. We only include promotions before 2003. This variable is split into quintiles, with the middle quintile omitted. The ability index was constructed using the teachers' scores on an education test, a language test and a psychology test that were taken before the teacher began teaching. Scores are first standardized by the year in which they were taken, and then for each year the following measure is calculated: $[S - ave(S)]/ave(S)$ where S is the standardized score on the tests for a teacher, and $ave(S)$ is the average of the scores of the teachers in the same rank in the same school, to capture relative ability. The index was constructed using Principal Components Analysis (PCA). Dummies for the top and bottom 10% of this index are included. Performance evaluation score is measured on a scale of 1-4 (fail, pass, good, excellent).

Table 7: Evaluation Scores, Time Use, and Test Scores

Outcome: excellent or good evaluation score			
Panel A: Time Use			
Fixed Effect	County	District	School
	Mfx/se	Mfx/se	Mfx/se
Log of total hours	0.10	0.080*	0.088*
	0.063	0.047	0.052
N	966	966	966
Panel B: Language Test Scores			
Fixed Effect	County	District	School
	Mfx/se	Mfx/se	Mfx/se
Average test scores (weighted)	0.010***	0.010**	0.010
	0.003	0.005	0.006
N	175	175	175

Notes: $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors clustered at the level of the fixed effect. Controls include the age of the teacher, a dummy variable for a female teacher, dummy variables for the level of education of the teacher, and the ability, promotion quintile, and number of teachers variables included in Table 6. Log of total time use first sums up the number of hours per week spent on grading homework, preparing lesson plans, participating in teaching and research activities, coaching of students outside class, organising extracurricular activities, home visits, and disciplining students, and then takes the log. Language test scores are standardized by the grade of the student, and then an average is calculated for each teacher. The regressions are weighted to account for the measurement error caused by teachers having different numbers of students' test scores in their average (see details in section 5.3).

Table 8: Parameters of the promotion system and effort incentives

Outcome: Evaluation score of teacher i in year t		
	(1)	
	Primary 2, Middle 2, Middle 1	
	coefficient	se
Difference in log wages (excluding teacher i)	0.936**	0.436
$t = X - 5$ to $t = X - 1$ dummy variable	0.013	0.031
$t \geq X$ dummy variable	0.096**	0.037
Education - high school	-0.150	0.126
Education - vocational college	-0.000	0.048
Education - college	0.050	0.058
Experience	0.005	0.004
Female	-0.040	0.036
Proportion promoted - quintile 1	-0.013	0.096
Proportion promoted - quintile 2	0.025	0.069
Proportion promoted - quintile 4	-0.117*	0.065
Proportion promoted - quintile 5	-0.115	0.068
Ability in bottom 10%	0.018	0.106
Ability in top 10%	0.020	0.150
Log Number of teachers in rank in township	-0.072***	0.023
Log teachers * ability in bottom 10%	-0.016	0.025
Log teachers * ability in top 10%	0.004	0.050
Average education in county rank - high school	0.845*	0.458
Average education in county rank - vocational college	0.266	0.261
Average education in county rank - college	0.279	0.264
Average proportion female in county rank	-0.140	0.378
Average experience in county rank	-0.011	0.009
Average proportion female * experience in county rank	0.016	0.018
Education test	0.066***	0.019
Psychology test	-0.009	0.024
Language test	-0.002	0.017
Constant	2.660***	0.259
Number of observations	3,423	
R^2	0.037	

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Regression includes county fixed effects and robust standard errors. Also includes rank level dummies, and year dummies. Proportion promoted is the proportion of teachers in a county that are promoted at that rank within a 'reasonable' number of years. We first calculate the time it took for each teacher to be promoted from one rank to another. Then we calculate the number of years for which, for each rank and education level in a county, half of the teachers were promoted within that time. Then for each rank and county we calculate the proportion of teachers promoted within that 'reasonable' time. We only include promotions before 2003. The ability index was constructed using the teachers' scores on an education test, a language test and a psychology test that were taken before the teacher began teaching. Scores are first standardized by the year in which they were taken, and then for each year the following measure is calculated: $[S - ave(S)]/ave(S)$ where S is the standardized score on the tests for a teacher, and $ave(S)$ is the average of the scores of the teachers in the same rank in the same school, to capture relative ability. The index was constructed using Principal Components Analysis (PCA). Dummies for the top and bottom 10% of this index are included. Evaluation score is measured on a scale of 1-4 (fail, pass, good, excellent).

Table 9: Primary 1, Middle 2, and Middle 1 rank effort incentives

Outcome: evaluation score of teacher i in year t	(1)		(2)
	coef	se	se
Relative age	0.300	0.377	0.385
Log number of teachers in rank in township (N)	0.015	0.082	0.079
Log N * ability top 10%	-0.091*	0.052	0.050
Log N * ability bottom 10%	-0.044	0.063	0.063
Ability in top 10%	0.322*	0.193	0.193
Ability in bottom 10%	0.038	0.201	0.201
t = X - 7	0.057	0.097	0.087
t = X - 6	0.011	0.051	0.053
t = X - 4	-0.002	0.054	0.052
t = X - 3	0.016	0.062	0.067
t = X - 2	0.136**	0.069	0.070
t = X - 1	0.228***	0.078	0.079
t = X	0.163**	0.080	0.083
t = X + 1	0.176*	0.093	0.090
t = X + 2	0.249**	0.100	0.085
t = X + 3	0.142	0.110	0.102
t = X + 4	0.130	0.121	0.110
t = X + 5	0.244*	0.128	0.115
t = X + 6	0.328**	0.144	0.146
t = X + 7	0.360**	0.179	0.192
t = X + 8	0.269	0.182	0.190
t = X + 9	0.078	0.228	0.220
t = X + 10	-0.117	0.199	0.204
t = X + 11	-0.278	0.230	0.204
t = X + 12	-0.108	0.299	0.259
t = X + 13	-1.722***	0.313	0.319
t = X + 14	-1.273***	0.351	0.356
t $\geq X + 15$	-1.171***	0.274	0.282
Number of observations	3,683		3,575
R^2	0.022		0.022
Level of clustering	Teacher		Teacher and district-rank

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Regression includes teacher fixed effects and clustered standard errors at the teacher level in column (1), and two-way clustering at the teacher and district-rank in column (2). The ability index was constructed using the teachers' scores on an education test, a language test and a psychology test that were taken before the teacher began teaching. Scores are first standardized by the year in which they were taken, and then for each year the following measure is calculated: $[S - ave(S)]/ave(S)$ where S is the standardized score on the tests for a teacher, and $ave(S)$ is the average of the scores of the teachers in the same rank in the same school, to capture relative ability. The index was constructed using Principal Components Analysis (PCA). Dummies for the top and bottom 10% of this index are included. Evaluation score is measured on a scale of 1-4 (fail, pass, good, excellent). t denotes time, X denotes the year in which a teacher becomes eligible to apply for promotion.

Table 10: Robustness Checks

	Outcome: evaluation score of teacher i in year t					
	(1)		(2)		(3)	
	Ability in top & bottom 15%		Teacher sampling rate $\geq$ 25%		Teacher sampling rate $\geq$ 35%	
Relative age	0.314	(0.319)	0.536	(0.339)	0.451	(0.344)
Log number of teachers in rank in township (N)	0.035	(0.070)	-0.020	(0.072)	-0.036	(0.075)
Ability in bottom 15%	-0.092*	(0.051)				
Ability in top 15%	-0.067	(0.051)				
Log N * ability bottom 15%	0.277*	(0.166)				
Log N * ability top 15%	0.221	(0.166)				
Ability in top 10%			0.291*	(0.173)	0.123	(0.183)
Ability in bottom 10%			0.060	(0.189)	0.036	(0.199)
Log N * ability top 10%			-0.078	(0.050)	-0.035	(0.052)
Log N * ability bottom 10%			-0.060	(0.060)	-0.046	(0.062)
t = X - 7	0.051	(0.103)	0.053	(0.107)	0.081	(0.107)
t = X - 6	0.011	(0.053)	0.019	(0.056)	0.029	(0.056)
t = X - 4	-0.002	(0.051)	0.004	(0.053)	-0.005	(0.054)
t = X - 3	0.015	(0.054)	0.039	(0.056)	0.031	(0.057)
t = X - 2	0.135**	(0.058)	0.112*	(0.061)	0.119*	(0.062)
t = X - 1	0.233***	(0.064)	0.198***	(0.067)	0.206***	(0.068)
t = X	0.169**	(0.068)	0.170**	(0.072)	0.203***	(0.073)
t = X + 1	0.181**	(0.077)	0.178**	(0.080)	0.206**	(0.081)
t = X + 2	0.251***	(0.083)	0.265***	(0.086)	0.307***	(0.088)
t = X + 3	0.146	(0.090)	0.162*	(0.095)	0.204**	(0.096)
t = X + 4	0.134	(0.101)	0.119	(0.105)	0.165	(0.107)
t = X + 5	0.244**	(0.110)	0.216*	(0.114)	0.285**	(0.117)
t = X + 6	0.320***	(0.122)	0.298**	(0.128)	0.357***	(0.130)
t = X + 7	0.355**	(0.144)	0.363**	(0.149)	0.466***	(0.152)
t = X + 8	0.259	(0.161)	0.185	(0.169)	0.296*	(0.173)
t = X + 9	0.067	(0.187)	0.040	(0.189)	0.159	(0.194)
t = X + 10	-0.143	(0.210)	-0.159	(0.212)	-0.052	(0.214)
t = X + 11	-0.277	(0.342)	-0.320	(0.339)	-0.205	(0.340)
t = X + 12	-0.110	(0.343)	-0.097	(0.362)	0.022	(0.363)
t = X + 13	-1.681**	(0.798)	-1.641*	(0.972)	-1.474	(0.970)
t = X + 14	-1.267*	(0.733)	-1.956**	(0.781)	-1.900**	(0.780)
t $\geq X + 15$	-1.191*	(0.708)	-1.121	(0.714)	-1.079	(0.712)
Observations	3,683		3,373		3,279	
R^2	0.022		0.023		0.024	

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Regression includes teacher fixed effects and clustered standard errors at the level of the teacher. The ability index was constructed using the teachers' scores on an education test, a language test and a psychology test that were taken before the teacher began teaching. Scores are first standardized by the year in which they were taken, and then for each year the following measure is calculated: $[S - ave(S)]/ave(S)$ where S is the standardized score on the tests for a teacher, and $ave(S)$ is the average of the scores of the teachers in the same rank in the same school, to caputure relative ability. The index was constructed using Principal Components Analysis (PCA). Dummies for the top and bottom 10% of this index are included. Evaluation score is measured on a scale of 1-4 (fail, pass, good, excellent). t denotes time, X denotes the year in which a teacher becomes eligible to apply for promotion.

Table 11: Primary and Middle high rank effort incentives

Outcome: evaluation score of teacher i in year t	coeff	se
Relative age	-1.218	(0.763)
Log number of teachers in rank in district (N)	-0.267	(0.309)
Log N * ability top 10%	0.139	(0.229)
Log N * ability bottom 10%	0.229	(0.342)
Ability in top 10%	-0.554	(0.798)
Ability in bottom 10%	-1.124	(1.234)
2 years in rank	-0.118	(0.107)
3 years in rank	-0.240*	(0.124)
4 years in rank	-0.350**	(0.142)
5 years in rank	-0.308**	(0.156)
6 years in rank	-0.407**	(0.171)
7 years in rank	-0.407**	(0.207)
8 years in rank	-0.687***	(0.230)
9 years in rank	-0.758***	(0.258)
10 years in rank	-0.619**	(0.275)
11 years in rank	-0.835***	(0.291)
12 years in rank	-0.912***	(0.312)
13 years in rank	-0.823**	(0.324)
14 years in rank	-1.089**	(0.450)
15 + years in rank	-0.912	(0.640)
Observations	648	
R^2	0.071	

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Regression includes teacher fixed effects and clustered standard errors at the teacher level. The ability index was constructed using the teachers' scores on an education test, a language test and a psychology test that were taken before the teacher began teaching. Scores are first standardized by the year in which they were taken, and then for each year the following measure is calculated: $[S - ave(S)]/ave(S)$ where S is the standardized score on the tests for a teacher, and $ave(S)$ is the average of the scores of the teachers in the same rank in the same school, to capture relative ability. The index was constructed using Principal Components Analysis (PCA). Dummies for the top and bottom 10% of this index are included. Evaluation score is measured on a scale of 1-4 (fail, pass, good, excellent).

C Appendix: Additional Tables

Table 12: Number of years until promotion eligibility - by education level

Rank change	**Education level**	**Number of years to wait**
Intern to Primary 2	Any	1 year after hired
Primary 2 to Primary 1	High school and below	4 years in the rank
	Vocational college	3 years in the rank
	College	1 year in the rank
Primary 1 to Primary high	High school and below	10 years in the rank
	Vocational college	7 years in the rank
	College	5 years in the rank
Intern to Middle 3	Any	1 year after hired
Middle 3 to Middle 2	High school and below	4 years in the rank
	Vocational College	3 years in the rank
	College	1 year in the rank
Middle 2 to Middle 1	High school and below	10 years in the rank
	Vocational college	7 years in the rank
	College	5 years in the rank
Middle 1 to Middle high	High school and below	25 years after hired (+ 5 years at Middle 1)
	Vocational college	15 years after hired (+ 5 years at Middle 1)
	College	5 years after Middle 1

Table 13: Primary 1, Middle 2, and Middle 1 rank effort incentives before promotion eligibility

Outcome: evaluation score of teacher i in year t		
	coef	se
Relative age	0.218	0.299
Log number of teachers in rank in township (N)	-0.034	0.041
Log N * ability top 10%	-0.092	0.065
Log N * ability bottom 10%	0.012	0.066
Ability in top 10%	0.378*	0.213
Ability in bottom 10%	0.091	0.214
t = X - 7	-0.055	0.154
t = X - 6	-0.074	0.081
t = X - 4	-0.075	0.087
t = X - 3	-0.088	0.094
t = X - 2	0.134	0.098
t = X - 1	0.242**	0.099
Cut 1	-2.961***	0.279
Cut 2	0.121	0.157
Cut 3	1.017***	0.160
Number of observations	1,887	

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Regression is estimated using ordered probit and includes clustered standard errors at the teacher level. Only observations in the time before promotion eligibility are included to reduce concerns of selection bias. The ability index was constructed using the teachers' scores on an education test, a language test and a psychology test that were taken before the teacher began teaching. Scores are first standardized by the year in which they were taken, and then for each year the following measure is calculated: $[S - ave(S)]/ave(S)$ where S is the standardized score on the tests for a teacher, and $ave(S)$ is the average of the scores of the teachers in the same rank in the same school, to capture relative ability. The index was constructed using Principal Components Analysis (PCA). Dummies for the top and bottom 10% of this index are included. Evaluation score is measured on a scale of 1-4 (fail, pass, good, excellent). t denotes time, X denotes the year in which a teacher becomes eligible to apply for promotion.

D Appendix: Proof of Prediction 5

Assume that teachers have careers lasting T periods. They are eligible for promotion after serving X years. Teachers can be promoted in any year t after they become eligible ($t > X$). As noted earlier, observed teacher performance in each year t is equal to the sum of skill, effort, and luck: $q_t = s + e_t + v_t$. The expected probability of promotion (p_t) in any given year t is based on teacher performance in the previous five years ($p_t(q_{t-1}, q_{t-2}, q_{t-3}, q_{t-4}, q_{t-5})$). Two simplifying assumptions are made: (1) that the impact of effort in one year on the probability of future promotion is independent of the impact of effort in other years, that is $d^2p_t/de_{t-i}de_{t-j} = 0$ for all i and j, and (2) that performance of effort in any of the five years before promotion has an equal impact on the probability of promotion ($dp_t/de_{t-1} = dp_t/de_{t-2} = ... = dp_t/de_{t-5}$).[36]

Now consider a teacher's optimal effort decision in year t. We can normalise the per-period utility from wages before promotion to equal zero, and define $U_h > 0$ to be the per-period utility from wages after promotion. In any given year j, the elements of lifetime expected discounted utility that are affected by effort in that year can be expressed as:

$$EV_j = -c(e_j) + Ep_{j+1}\sum_{t=j+1}^{T}\beta^{t-j}U_h + (1 - Ep_{j+1})[Ep_{j+2}\sum_{t=j+2}^{T}\beta^{t-j}U_h + (1 - Ep_{j+2})\{Ep_{j+3}\sum_{t=j+3}^{T}\beta^{t-j}U_h + (1 - Ep_{j+3})(Ep_{j+4}\sum_{t=j+4}^{T}\beta^{t-j}U_h + (1 - Ep_{j+4})Ep_{j+5}\sum_{t=j+5}^{T}\beta^{t-j}U_h)\}] \quad (7)$$

Here, $c(e_j)$ is the disutility of effort in year j, β is the discount rate, and Ep_t denotes the expected promotion probability given the distribution of draws of luck, where $p_t(s, e, \underline{E}, \underline{S})$ is defined as before. Current effort can alter the probability of being promoted in the subsequent five years, and once promoted the teacher enjoys higher utility until period T. Note that beyond year $j + 5$, the expected probability of promotion is not influenced by current year effort. As thus formulated, maximisation of (7) with respect to e_j defines optimal effort by the teacher.

In years in which there are more than five years before a teacher is eligible for promotion, effort incentives are zero. That is, effort incentives are zero whenever $t \leq X - 5$. Choice of

[36] These assumptions are made for tractability. As long as effort and performance in each of the previous five years increases the probability of promotion in years of promotion eligibility, the qualitative results are essentially the same.

effort in year $t = X - 5$ is such that equation (7) is maximised, where $j = X - 5$:

$$EV_{X-5} = -c(e_{X-5}) + Ep_{X-4} \sum_{t=X-4}^{T} \beta^{t-(X-5)} U_h + (1 - Ep_{X-4})[Ep_{X-3} \sum_{t=X-3}^{T} \beta^{t-(X-5)} U_h$$
$$+ (1 - Ep_{X-3})\{Ep_{X-2} \sum_{t=X-2}^{T} \beta^{t-(X-5)} U_h + (1 - Ep_{X-2})(Ep_{X-1} \sum_{t=X-1}^{T} \beta^{t-(X-5)} U_h$$
$$+ (1 - Ep_{X-1})Ep_X \sum_{t=X}^{T} \beta^{t-(X-5)} U_h)\}] \quad (8)$$

However, $Ep_{X-4} = Ep_{X-3} = Ep_{X-2} = Ep_{X-1} = Ep_X = 0$ since the teacher cannot be promoted until $t = X + 1$. As a result, all the terms in (8) drop out except for $-c(e_{X-5})$. Thus, effort is zero.

In the five years leading up to the year in which a teacher becomes eligible for promotion, effort incentives are increasing. That is, effort incentives are increasing from $t = X - 4$ to $t = X$. We can write out expressions for choice of effort in years $j = X - 4$, $j = X - 3$, $j = X - 2$, $j = X - 1$, and $j = X$ using equation (7). In $j = X - 4$:

$$EV_{X-4} = -c(e_{X-4}) + Ep_{X-3} \sum_{t=X-3}^{T} \beta^{t-(X-4)} U_h + (1 - Ep_{X-3})[Ep_{X-2} \sum_{t=X-2}^{T} \beta^{t-(X-4)} U_h$$
$$(1 - Ep_{X-2})\{Ep_{X-1} \sum_{t=X-1}^{T} \beta^{t-(X-4)} U_h + (1 - Ep_{X-1})(Ep_X \sum_{t=X}^{T} \beta^{t-(X-4)} U_h$$
$$+ (1 - Ep_X)Ep_{X+1} \sum_{t=X+1}^{T} \beta^{t-(X-4)} U_h)\}] \quad (9)$$

However, $Ep_{X-3} = Ep_{X-2} = Ep_{X-1} = Ep_X = 0$ since the teacher cannot be promoted until $t = X + 1$. As a result, (9) simplifies to:

$$EV_{X-4} = -c(e_{X-4}) + Ep_{X+1} \sum_{t=X+1}^{T} \beta^5 U_h \quad (10)$$

A teacher chooses effort in $j = X - 4$ in order to maximise:

$$\max_{e_{X-4}} \left\{ Ep_{X+1} \sum_{t=X+1}^{T} \beta^5 U_h - c(e_{X-4}) \right\} \tag{11}$$

The first order condition is:

$$c'(e_{X-4}) = \frac{dEp_{X+1}}{de_{X-4}} \sum_{t=X+1}^{T} \beta^5 U_h \tag{12}$$

This procedure can be repeated for $j = X - 3$, $j = X - 2$, $j = X - 1$, and $j = X$.

In $j = X-3$, after substiting $j = X-3$ and $Ep_{X-2} = Ep_{X-1} = Ep_X = 0$, the corresponding equation for expected lifetime utility is:

$$EV_{X-3} = -c(e_{X-3}) + Ep_{X+1} \sum_{t=X+1}^{T} \beta^4 U_h + (1 - Ep_{X+1}) Ep_{X+2} \sum_{t=X+2}^{T} \beta^5 U_h \tag{13}$$

The first order condition is now:

$$\begin{aligned} c'(e_{X-3}) = \frac{dEp_{X+1}}{de_{X-3}} \sum_{t=X+1}^{T} \beta^4 U_h + \frac{dEp_{X+2}}{de_{X-3}} \sum_{t=X+2}^{T} \beta^5 U_h \\ - Ep_{X+2} \frac{dEp_{X+1}}{de_{X-3}} \sum_{t=X+2}^{T} \beta^5 U_h - Ep_{X+1} \frac{dEp_{X+2}}{de_{X-3}} \sum_{t=X+2}^{T} \beta^5 U_h \end{aligned} \tag{14}$$

Because of the second simplifying assumption made in the model, that performance of effort in any of the past five years before promotion has an equal impact on the probability of promotion ($dp_t/de_{t-1} = dp_t/de_{t-2} = ... = dp_t/de_{t-5}$), $\frac{dEp_{X+1}}{de_{X-3}} = \frac{dEp_{X+2}}{de_{X-3}}$ and the first order condition simplifies to:

$$c'(e_{X-3}) = \frac{dEp_{X+1}}{de_{X-3}} \left(\sum_{t=X+1}^{T} \beta^4 U_h + \sum_{t=X+2}^{T} \beta^5 U_h (1 - Ep_{X+1} - Ep_{X+2}) \right) \tag{15}$$

Now we can compare (15) and (12). Again, because of simplifying assumption (2), we can substitute $\frac{dEp_{X+1}}{de_{X-3}} = \frac{dEp_{X+1}}{de_{X-4}}$ in (15) and compare:

$$\frac{dEp_{X+1}}{de_{X-4}} \sum_{t=X+1}^{T} \beta^5 U_h < \frac{dEp_{X+1}}{de_{X-4}} \left(\sum_{t=X+1}^{T} \beta^4 U_h + \sum_{t=X+2}^{T} \beta^5 U_h (1 - Ep_{X+1} - Ep_{X+2}) \right) \tag{16}$$

which holds as long as $Ep_{X+1}+Ep_{X+2} < 1$ since $\beta^5 < \beta^4$ and $\sum_{t=X+1}^{T}\beta^5 U_h < \sum_{t=X+2}^{T}\beta^5 U_h$. The sum of all expected probabilities is indeed less than one, since a teacher cannot be absolutely certain of promotion in the future. In $t = X - 2$ the FOC simplifies to:

$$c'(e_{X-2}) = \frac{dEp_{X+1}}{de_{X-2}}[\sum_{t=X+1}^{T}\beta^3 U_h + \sum_{t=X+2}^{T}\beta^4 U_h(1 - Ep_{X+1} - Ep_{X+2})+$$
$$(\sum_{t=X+3}^{T}\beta^5 U_h(1 - Ep_{X+1}(1 - Ep_{X+2} - Ep_{X+3}) - Ep_{X+2}(1 - Ep_{X+3}) - 2Ep_{X+3})] \tag{17}$$

which again is larger than the FOC for $j = X - 3$. The same procedure can be followed for $j = X - 1$ and $j = X$.

In the years after a teacher becomes eligible for promotion ($j > X$), effort incentives are no longer increasing. Assume that $j > X$. We can write an expression for $j + 1$:

$$EV_{j+1} = -c(e_{j+1}) + Ep_{j+2}\sum_{t=j+2}^{T}\beta U_h + (1 - Ep_{j+2})[Ep_{j+3}\sum_{t=j+3}^{T}\beta^2 U_h + (1 - Ep_{j+3})\{Ep_{j+4}$$
$$\sum_{t=j+4}^{T}\beta^3 U_h + (1 - Ep_{j+4})(Ep_{j+5}\sum_{t=j+5}^{T}\beta^4 U_h + (1 - Ep_{j+5})Ep_{j+6}\sum_{t=j+6}^{T}\beta^5 U_h)\}] \tag{18}$$

and compare it to the expression for j:

$$EV_j = -c(e_j) + Ep_{j+1}\sum_{t=j+1}^{T}\beta U_h + (1 - Ep_{j+1})[Ep_{j+2}\sum_{t=j+2}^{T}\beta^2 U_h + (1 - Ep_{j+2})\{Ep_{j+3}$$
$$\sum_{t=j+3}^{T}\beta^3 U_h + (1 - Ep_{j+3})(Ep_{j+4}\sum_{t=j+4}^{T}\beta^4 U_h + (1 - Ep_{j+4})Ep_{j+5}\sum_{t=j+5}^{T}\beta^5 U_h)\}] \tag{19}$$

The discounting terms are all equivalent, the expected probability of promotion is non-zero in all years in both cases (since $j > X$), and there are an equivalent number of years in which current effort will affect a teacher's expected probability of promotion (five). As a result, effort incentives are not increasing.