Managing Teachers in Low-Income Countries

Naureen Karachiwalla

University College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy in Economics

Trinity Term 2013

Managing Teachers in Low-Income Countries

Naureen Karachiwalla

University College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy in Economics

Trinity Term 2013

Abstract

Apart from the introduction (Chapter 1) and conclusion (Chapter 7), this thesis comprises five chapters organized into two parts:

Part I studies promotion incentives in the public sector, and focuses on the case of teachers in rural China. All teachers in China compete with their colleagues for rank promotions. I aim to answer two questions: first, whether the promotion system for teachers in China elicits effort from teachers, and second, how the design features of the promotion system affect effort incentives. Part I includes four chapters. Chapter 2 introduces the topic and provides a background on promotions for teachers in China. It also discusses related work in this area, and introduces the data that will be used in Part I.

Chapter 3 presents and tests a theoretical model of promotions as an incentive device. The model treats all teachers as identical in terms of their ability, and as such, focuses on average levels of teacher effort. It predicts that effort is exerted in response to potential promotions. In addition, the model also predicts that average effort incentives are higher in promotion contests in which the wage gap is higher, the promotion rate is closer to one half, the number of teachers competing for a promotion is higher (for promotion rates between 1/3 and 2/3), and the average age of teachers in the contest is lower, or the proportion of female teachers is lower. The model is used to derive an estimating equation by which to test predictions on average levels of teacher effort. An equation is estimated for the probability of promotion as a function of teacher effort, which is proxied by the teachers' annual performance evaluation scores. There is simultaneity present as effort increases the probability of promotion, but it is also the promise of promotion that motivates effort. As a result, effort is instrumented using wage changes, which are both informative (higher wage gaps are associated with higher effort) and valid (wages only affect promotions through effort). The second stage of the regression demonstrates that effort is indeed exerted by teachers in order to win promotions. The first stage confirms the predictions of the model with regards to wage gaps, the promotion rate, and the size and composition of the pool of competitors.

Chapter 4 extends the model of Chapter 3 in two ways: teachers are now treated as heterogeneous in ability, and a multi-period model of teacher effort over time is also added. This chapter focuses on individual levels of teacher effort, and on how the parameters of the promotion system interact with teacher characteristics to affect teacher effort. The predictions include that teachers in the extremes of the skill distribution will have lower incentives, and as the contest size increases these teachers will have effort incentives that

are lower still, that teachers who are five or more years from promotion eligibility will have zero effort, as will teachers in the highest rank, that teacher effort will increase in the five years leading up to promotion eligibility, and that teacher effort will decrease after a teacher is eligible for promotion but has been passed over several times. An effort equation is estimated that captures all of these components, and the predictions are largely affirmed by the data. Tests are conducted in order to alleviate concerns about selection, as well as measurement error in the performance evaluation scores. Chapter 5 concludes Part I.

Part II of this thesis looks at teacher labour markets, social distance, and learning outcomes in Punjab, Pakistan. Chapter 6 explores the link between the distribution of teachers in the labour market, caste differences between teachers and students, and child learning outcomes. Using rich longitudinal data from Pakistan that allows me to convincingly identify the causal effects of caste on learning outcomes, I show how the distribution of teachers across public schools induces particular matches of high and low caste teachers and students, and that these matches are highly predictive of test score outcomes. Specifically, low caste male children perform significantly better when taught by high caste teachers than when they are taught by low caste teachers. Several possible channels are explored, including discrimination in the classroom, role model effects, teacher quality, patronage, peer effects, and returns to education. Although the channel cannot be proven, the data points to high caste teachers being able to raise the already high returns to education for low caste children because they are able to assist these children in getting educational benefits and employment later on using their patronage networks. Low caste children therefore work harder to impress high caste teachers, and this results in higher learning outcomes.

Acknowledgements

This thesis has only been possible with the support and encouragement of a large number of people. I would especially like to thank both of my thesis supervisors. I have had the incredible fortune of working with two people I respect and admire both professionally and personally. It is rare for a supervisor to go above and beyond for his students the way that Professor Albert Park has. This includes not only reading every page and footnote (more than once), but also kindness and understanding at difficult times. I am also very lucky to have worked with Professor Stefan Dercon through two degrees. This is a supervisor who makes his students feel confident and encouraged, who always sticks up for his students, and whose door is always open for a friendly chat or for academic advice. I really wish to express my deepest gratitude to both of these extraordinary mentors.

Thanks also go to Professor Dercon for introducing me to Dr. Jishnu Das, who has been another wonderful mentor. I have learned a great deal working with him, and he has been enormously generous, particularly with his time. I would also like to thank Professor Eliana LaFerrara for the opportunity to participate in the Actors, Markets, and Institutions in Developing Countries (AMID) programme, as well as my fellow AMID Fellows and my friends at Bocconi University. My experience there really helped foster my confidence as a researcher.Professor Clare Leaver's course on public economics inspired both my thesis topic and my current research agenda. My thesis also benefitted from considerable advice and suggestions from Professor Frances Teal and Professor Margaret Stevens during my transfer interview, as well as from Professor Simon Quinn and Professor Paul Glewwe during my viva. I am very grateful as their advice has improved my papers a great deal. Thank you to Professor Quinn and Professor Glewwe for also being extremely kind examiners - it made a big difference. Many conference and seminar participants have also provided much appreciated suggestions. The staff at the Centre for the Study of African Economies, the Economics Department, the Graduate Studies Office and the Young Lives office, as well as at University College also made university and college life a lot easier and more pleasant.

Matt Collin deserves a very special thank you: for talking through exam questions, promotion models, regressions etc.; for sharing many a sunny (albeit slow) lunch outside of Stefan's office; and, for always being happy to help. Sonya Krutikova, Lara Lebedinski, Abhijeet Singh, and Kate Vyborny were all very kind in taking the time to read various chapters and provide detailed feedback. Bet Caeyers, Matt Collin, Ghada Fayad, Ricarda Gaentzsch, Torfinn Harding, Sebastian Konigs, Sonya Krutikova, Fah Jirasavetakul and Christoph Lakner, Alan Sanchez, Abhijeet Singh, and Aly Visram can always be counted on for a hike along the coast (a specific coast - others will not do), making and starring in movies, scampi, gag gifts for supervisors, giving talks to potential ODI fellows, a movie night, talking you through rough times, having five people think you are lost somewhere in Italy, enjoying Oxford in the snow, ironing your sheets, nicknames from adorable daughters, keeping you calm before your viva, educational sessions about movies from the 80's, presenting during the same session at every conference, Sunday morning coffee concerts, long talks over your favourite pizza, trips to Reading, or running around Milan in the rain, as well as much, much reassurance and advice.

Finally, I wish to thank my family and friends for their extensive encouragement and support, particularly my parents and my three sisters (Fareen Karachiwalla, Salima Jaffer, and Yasmin Mawani) as well as Neha Batura and Lindsey Craig, without whom this thesis would have been almost impossible, and far less enjoyable to complete.

Word Count

The approximate number of words in this thesis is 74,500, excluding the bibliography. This is calculated using the number of words on page 117 (423) multiplied by the number of pages (176).

Declaration

One chapter of this thesis (Chapter 4) is co-authored with Professor Albert Park at the Hong Kong University of Science and Technology (HKUST), and Part I will form the basis of a co-authored article for publication.

Contents

List of Figures

List of Tables

Chapter 1

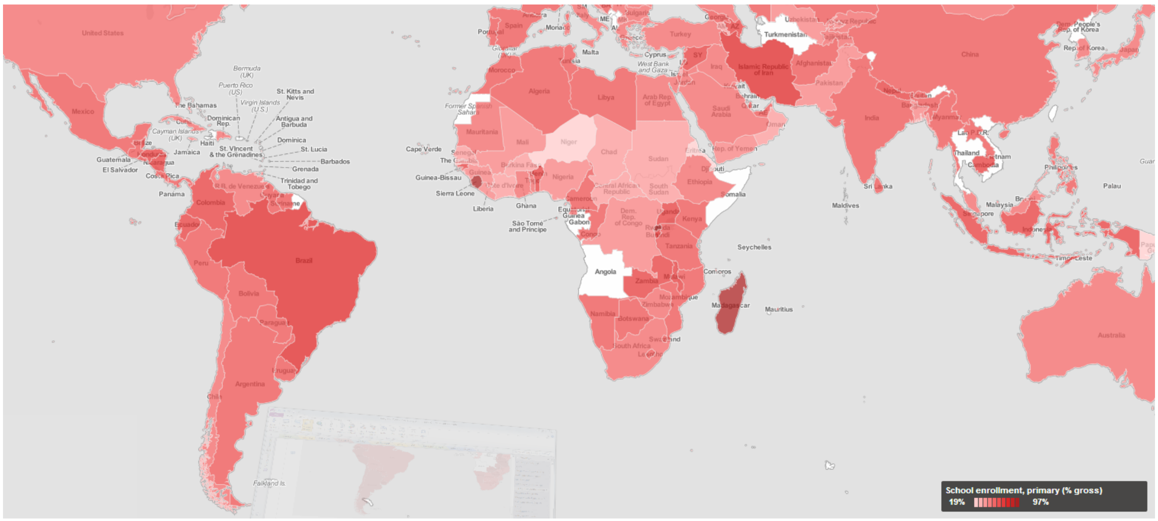
Introduction

Education is one of the foremost drivers of growth in an economy, fuelling the development of human capital (Mankiw et al., 1992; Barro, 2000, 2001). It is also a key determinant of individual economic and social well-being (Sen, 2001; Duflo, 2001; Glewwe, 2002). As a result, there has been a push in the international community towards funding and developing a strong education sector in low-income countries. The Millennium Development Goals (MDGs) include one very significant education-related goal; universal enrolment of primary school-aged children by 2015. Now that many low-income countries are coming close to achieving this goal (see Figure 1.1 below), the focus has shifted to retaining children in higher levels of education, and to the quality of education, and in particular, to improving learning outcomes (Glewwe et al., 2011). Learning is important, as this is what determines the skills that children will actually take to the labour market, and determines the way in which they will contribute to a growing economy (Hanushek and Woessmann, 2011). Studies have found that literacy and numeracy can have higher returns than primary school completion. In addition, improved learning in schools can be an important way to get children enrolled in school, as this would improve the returns to education (Das et al., 2006).

Teachers

Much of the focus on student learning has centered on teachers (Hanushek and Rivkin, 2006b). There are over 200 million primary and secondary school teachers worldwide

Figure 1.1: International Gross Enrolment Rates



Notes: Total enrolment in primary education as a share of the primary school-aged population. White indicates no data. Source: World Bank, World Development Indicators.

(UNESCO Institute for Statistics), and understanding this massive workforce has been a central preoccupation of education economists. We are interested in how teachers select into the profession and how to retain them, how to train them, how to motivate teacher effort, how teachers teach, and which teacher characteristics matter for student learning. This thesis focuses on two of these aspects of the teaching workforce: how to motivate teachers, and a particular characteristic of teachers, social distance between a teacher and student. Although there is agreement that both teacher incentives and social differences matter, it is not clear exactly how and how much they matter.

More motivated teachers should lead to more effort and to improved student learning. However, many countries have tried numerous methods of motivating teachers and have found mixed results. What are some places doing right, and what are ways some places can improve? Social distance between teachers and students, which includes factors such as gender, ethnicity, caste and religion, could either benefit students if there are complementarities of teacher and student characteristics in learning, or they could hinder learning if not. The examples that I use to look at these issues both come from Asia (China and Pakistan), but similar concerns are faced across low and middle-income countries and the lessons learned could be applied elsewhere. In this thesis, I ask two sets of questions. First, has the world's most populous and one of the fastest growing countries (China) found a successful way of motivating teachers? If so, how have they achieved this? Second, does social distance in the form of caste in one of Asia's largest economies (Pakistan) have impacts on learning that could be affecting its growth? If so, what are the potential mechanisms behind this?

Teacher Incentives and China

The interest in incentivising teachers has come out of the observation that, particularly in low-income countries, learning levels are low and teachers are often found either to be absent or not making full use of instructional time (Glewwe and Kremer, 2006; Bruns et al., 2011). 90% of children in lower-middle-income countries that participated in the Programme for International Student Assessment (PISA) test in 2009 scored below 400 points, the threshold for basic numeracy skills (Bruns et al., 2011). In a study of several countries, teacher absenteeism was found to be around 19% on average, but was as high as 40% in some Indian states (Kremer et al., 2005; Chaudhury et al., 2006). Teacher absence has a strong negative impact on learning (Glewwe et al., 2011).

In addition, teachers are part of the public service, and they often form a very large proportion of it. The effectiveness of incentives in the public sector in general, and for teachers in particular, has been found to be quite low. This is because the public sector faces special incentive problems. For example, public sector employees often have multiple tasks that they are responsible for, and these tasks may compete with one another. A teacher is responsible for teaching either one or often many subjects, instilling good values in students, administrative activities, disciplining students, and so on. Placing heavy incentives on one or a few tasks may cause effort to suffer in other dimensions (Holmstrom and Milgrom, 1991). Public sector employees are often also accountable to more than one individual. Teachers report to the school principal, but they are also accountable to parents, local and national education officials, and of course, students. If the objectives of these groups differ, it can be difficult to appropriately incentivise teacher effort (Dixit, 1997, 2002). Finally, the output of public sector employees is notoriously difficult to measure. Student learning can be measured by test scores, but even these are noisy, and they are affected by more than just the teacher. Other aspects of teacher output such as discipline and social values are even more difficult to measure. When output is less measurable, it is much more difficult to reward effort (Dixit, 1997, 2002).

China has become a model of economic growth, and its education system is viewed in high regard. The Chinese government has a unique method of incentivising public servants. It has instituted a system of evaluating and promoting public servants based on performance. Party-State officials of the same rank compete with one another based on economic growth rates in their locality (Li et al., 2012). This system of incentives extends to teachers as well, and teachers compete with others in their township for promotions. China's system differs from other countries in two respects: firstly, most countries have focused on the use of bonuses as a source of incentives for teachers, and secondly, although promotions are widely used amongst teachers, they are usually based on education and years of service rather than on merit and competition.

A teacher annual performance evaluation system developed alongside the promotion system. This evaluation system previously involved the school principal scoring teachers in several components of performance: acceptance of the principles of the Communist Party, loving the cause of education, being a good role model, lesson preparation, classroom teaching, homework, and most importantly, student test scores (Liu and Teddlie, 2005). In 2001, the system was reformed in two ways. First, it was no longer solely the principal who would undertake evaluations. Other teachers, students and parents, as well as local education officials would also be involved. Second, the performance indicators were streamlined to include four components (attitude - now not including acceptance of Communist Party values, attendance, preparation, and student test scores), and the weight of student test scores was reduced (Liu and Teddlie, 2005).

This is a highly sophisticated system based on merit, which incorporates both quantitative and qualitative indicators that are judged by several actors in the education system. It would not be an easy system to replicate elsewhere. It is possible that the reason this system might work so well is that it exists in the first place; adoption of such a sophisticated evaluation and promotion system is likely highly endogenous. The type of place that would adopt such a sophosticated system might also have people that respond to it quite well. However, there are lessons to be learned here regarding the extent to which such a complicated system works, and in what ways. The present study identifies some teachers who have low effort incentives, suggesting that efforts to reform the system could improve teacher performance.

Teacher and Student Differences and Pakistan

The second question I study in this thesis is how social distance between groups of individuals affects learning outcomes. I focus on caste in Pakistan. Pakistan provides a unique setting in which to study this topic, for two reasons. First, caste plays a large role in both social and economic exchanges across South Asia, including Pakistan, and it is closely tied to social and economic inequality (Mohmand and Gazdar, 2007). Barro (2000) and Deininger and Squire (1998) find that in poor countries, inequality slows growth. Inequality of opportunity may be particularly sinister: excluding a portion of the population from the opportunity to contribute to the economy can be especially harmful to overall growth (Marrero, 2009) and also for individual incomes (Bourguignon et al., 2007). Across South Asia, caste inequalities are obvious both for participation in the education system and for labour market outcomes.

In addition, Pakistan is the sixth most populous country in the world and an important political and economic player in South Asia. However, it is lagging behind in terms of indicators of education. Enrolment is lower than other Asian countries and learning levels are extremely low (Andrabi et al., 2007). In 2001-02, the primary net enrolment rate in Pakistan was only 51%; much lower than the corresponding 83% in India and 90% in Sri-Lanka, and Pakistan's enrolment is below what would be expected given its level of income (Das et al., 2006). Analysis of data from the Learning and Education Attainment in Punjab Schools (LEAPS) Survey revealed that by the time children leave Grade 5, only half are able to fill in a missing letter in a four-letter word accompanied by a picture, and only between 10-20% can construct very simple sentences in English. Only 54% of children can perform simple division, and less than a third are able to answer questions on simple algebra or fractions (Andrabi et al., 2012). In addition, there are gaps between children of differing socio-economic status, and parental education (Andrabi et al., 2012). If Pakistan is to capitalise on its rich resources, it will need to find a way to deliver inclusive and equitable growth.

Caste divisions in Pakistan are based on occupations, which have been persistent over time. Enrolment and school progression of low caste individuals is lower than that of high caste individuals (Jacoby and Mansuri, 2011). In the data used here, 47% of high caste children in the sample are enrolled in high school versus 36% of low caste children (LEAPS Data). Due to data constraints, there has been very limited study on caste differences in education in Pakistan. Jacoby and Mansuri (2011) study enrolment patterns based on caste in Punjab and Sindh provinces. Caste data have not previously been combined with learning data however, and this is the first study to do so.

Teachers are key input to learning. Many teacher characteristics that were originally thought to be critical for learning have turned out not to matter, including education, experience, and the teacher-pupil ratio (Glewwe and Kremer, 2006). Much of the variation in teacher effectiveness is unobserved (Rivkin et al., 2005). In a context in which caste filters into so many aspects of society, it is natural to wonder whether it may also be playing a role in learning, particularly due to the large differences in enrolment. Conditional on enrolment, does caste also play a role in the learning outcomes of children? One can envisage many ways in which caste could affect learning: through discrimination on the part of teachers, through role model effects, and through interactions between very different social groups, among others. The effects could be either positive or negative. Even a positive effect could illuminate mechanisms in the economy that were previously unknown if we can also provide some insight into why we see such an effect.

Plan of Thesis

Part I of this thesis studies empirically the performance evaluation and promotion system of rural teachers in Gansu province in North-West China. Chapter 2 provides an introduction to the topic, as well as the related literature. It also describes the evaluation and promotion system in Gansu, and describes the data that are used to test whether and how the promotion system incentivises teachers. Chapter 3 focuses on identifying whether promotions motivate effort. It looks at average levels of teacher effort and how this varies in promotion contests with different wage gaps, promotion rates, and the number of competitors. Chapter 4 focuses attention on individual teachers, and investigates how teacher effort varies over a teacher's career, and how it responds to changes in the competitive environment. Both Chapter 3 and Chapter 4 base the empirics on a theoretical model of promotions as incentives. Chapter 5 concludes Part I. Part II of this thesis focuses on social distance. Chapter 6 aims to identify the effect of child caste, teacher caste, and the interaction between child and teacher caste on learning outcomes in Pakistan, and offers some insights into the mechanisms that may be driving the results. The conclusion to this thesis (Chapter 7) will summarise the findings of the three substantive chapters and will outline lessons that can be learned from them.

This thesis attempts a substantial contribution to the literature on motivating teachers, and more generally, public servants. It studies a type of incentive system that has never been studied for teachers or for public servants in general, in order to draw lessons for other contexts. I also attempt to contribute to the literature on social distance in education, by identifying the effects of caste on learning outcomes in Pakistan, which has been impossible thus far due to data constraints. This prolific aspect of social dynamics in Pakistan is rarely studied, and its impact on education provides insights into inequality in one of Asia's largest countries.

Part I

Promotion Incentives in the Public Sector: Evidence from Chinese Schools

Chapter 2

Introduction to Part I

2.1 Promotion Incentives

Public servants comprise a large proportion of formal employment in low-income countries, but service delivery is often poor (Bruns et al., 2011). 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 for students (Glewwe, 2002; Bruns et al., 2011). 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; these are only weakly correlated with student achievement (Podgursky and Springer, 2007). As a result, there has been great interest in reforms that can incentivise teachers to exert greater effort, in particular through pay for performance schemes that link bonuses to student test scores (Muralidharan and Sundararaman, 2009).

China has taken a much different approach to incentivising teachers and other civil servants, establishing a sophisticated system of annual performance evaluations that feed into promotion decisions (Ding and Lehrer, 2001). 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 the international PISA exams (OECD, 2010). However, the design and performance of China's public service promotion system has received little attention or systematic study. Understanding how the system works could help unlock the puzzle of how China has grown so rapidly, despite having weak governance institutions that are viewed internationally as doing a poor job protecting property rights and guarding against corruption. Does China's public sector promotion system provide credible incentives for civil servants to exert effort to serve the public interest? If so, how does the design of the promotion system influence effort? Although I focus attention in this study on teachers, key features of the evaluation and promotion system are the same for all public employees in the country.

Part I of this thesis contributes to a small empirical literature that provides evidence on the performance of promotion systems. Nearly all previous studies examine promotion systems in the private sector. Of these, many study wage determination within companies to test whether pay increases substantially with rank (or promotion) rather than reflecting current marginal productivity (Medoff and Abraham, 1980, 1981; Baker et al., 1994a,b). Only a few link promotion incentives to direct measures of effort or performance (Gibbs, 1995; 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), public service jobs may involve multiple tasks or multiple principles to a greater degree than private sector jobs, making optimal incentives for specific tasks weaker (Holmstrom and Milgrom, 1991; Dixit, 1997), and civil servants may find ways to manipulate the system, as seen in evaluations of incentives provided to public agencies for job training programs in the US and UK (Heckman et al., 1997, 2002; Burgess et al., 2012). For these reasons, it is important to learn more about the actual performance of public sector promotion systems.

Unfortunately, there are almost no rigorous empirical studies of promotion systems in the public sector. Previous research 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). However, these studies focus only on high-level government officials and do not consider whether leaders respond to promotion incentives, making it difficult to distinguish whether higher growth rates result from selection or incentive effects. Perhaps the most closely related paper is a study by Haeck and Verboven (2012), which examines promotion incentives for professors within a European university; they find that faculty wages do not respond to the external labor market and that those with better research and teaching performance are more likely to be promoted.

Part I of this thesis will study this evaluation and promotion system in Gansu province in rural China. There are two questions that it will attempt to address: first, do promotions elicit effort? Second, how do the design features of the promotion system influence effort? Chapter 3 will look at whether promotions motivate effort. It assumes that teachers are identical and will focus on the effects of different types of promotion contests on average levels of teacher effort. A model of promotions as incentives is used to develop an estimating equation for whether promotions motivate effort, and that will also test predictions on how average levels of teacher effort may be affected as various components of the promotion system vary. The model predicts how the average level of effort of teachers will vary with the promotion rate, the difference in wages offered on promotion, and the interaction between the number of competitors competing for the promotion and the promotion rate. A four-year panel of teachers is then constructed from the Gansu Survey of Children and Families (GSCF) and a binary outcome model is estimated where promotions are regressed on the teachers' annual evaluation scores, which proxy for effort. The annual evaluation scores are endogenous, because although effort increases a teacher's probability of promotion, the probability of promotion is also what entices teachers to provide effort. Thus, there is simultaneity present. To correct for this, instrumental variables techniques are employed, using the average change in wages in each of 20 counties for each promotion level as an instrument for effort. The model predicts that the higher the increase in wage offered, the higher will be effort. Wages are a valid instrument as the wage structure for civil servants in China is well-defined and is exogenous to teachers' influence. Wages affect promotions only through effort because a teacher can only access an increased wage by working hard to get a promotion. I employ both a standard Probit model, as well as a Probit model with instrumental variables. Results show that average levels of effort are indeed higher with bigger wage gaps, and that promotions do provide incentives for effort. In addition, average levels of teacher effort are affected by the promotion rate and number of teachers in the manner in which the model predicts.

The aim of Chapter 4 is to test how individual teachers in rural China respond to the promotion system for public servants. It will focus on estimating an effort equation for individual teacher effort rather than average levels of effort, and will thus provide insight into the design of promotion systems. First, the one-period tournament model of promotions as incentives from Chapter 3 is extended to the case of teachers that differ by ability. It is used to derive predictions on how individual teacher effort may be affected by some additional features of the promotion system. The model predicts how the effort of teachers will vary with the perceived skill percentile of the teacher, and the interaction of this effect with the number of competitors competing for the promotion. Subsequently, a multi-period model that provides predictions on teacher effort over time is presented. The four-year panel of teachers is then used to estimate an extended effort equation. The predictions of the model are largely corroborated: effort increases as the teacher gets closer to the year in which she is eligible for a promotion; teachers that are repeatedly passed over for promotions tend to slack off; and the larger the number of teachers competing for a promotion, the lower are effort incentives for teachers at the extremes of the skill distribution. Although teachers are responding to the promotion system, the design of the incentive structure is important. This promotion system has design features that could be improved, and can provide lessons for the design of other promotion-based incentive systems.

The remainder of this chapter will outline some previous work on this topic, followed by an explanation of the data as well as the rationale behind the study. Chapter 3 will present the theoretical model, empirical specification, and results for the promotion equation and for average levels of teacher effort. Chapter 4 will present the model, empirical specification, and results for the individual teacher effort equation that incorporates heterogeneous teachers and a dynamic component. Chapter 5 will conclude Part I.

2.2 Related Literature

This section provides the context for this research. It outlines previous research and findings and describes how this thesis will build on them. There are four strands of literature to which this study contributes: the literature on motivating and incentivising public servants, and that for teachers, the literature concerning education and in particular teachers in China, and the labour economics literature that studies promotions as incentives through both theoretical models and empirical evidence.

Motivating Public Servants

The literature on motivating public servants is vast, both in terms of theory and empirics. Incentives in the public sector are generally deemed weaker than in the private sector. Three main problems are identified in the literature: measurement problems of outputs or outcomes, multiple tasks, and multiple principals. Dixit (2002) provides a useful review of the theoretical and empirical literature on incentives in the public sector. The theoretical literature incorporates the special problems of the public sector into incentive theory. Wilson (1989) describes different types of government agencies and what kinds of incentives are likely to be effective in that setting based on whether outputs, outcomes, both or neither can be observed. Dixit (2002) formally shows that optimal incentives are lower the less measurable is output. Holmstrom and Milgrom (1991) discuss the issue of public servants needing to perform multiple tasks. In this case, if tasks are effort substitutes, then the optimal incentive on each task is weaker. Public servants also tend to have multiple principals (politicians, head of organisation, the public) and Dixit (1997) shows that when principals behave non-cooperatively, incentives are also weaker. An important dimension of the work done on motivating public servants involves whether to provide incentives individually, or in teams. Corts (2007) shows that with a strong enough multi-task problem (when two or more tasks contribute to the same outcome in a way that cannot be separated) motivating agents in teams is generally better.

There are also many empirical studies on motivating public servants. In the United States, the best example is the Job Partnership Training Act (JPTA), in which job training centres were incentivised in order to try and maximise the employment and earnings of programme participants (outcomes). In the United Kingdom, Jobcentre Plus experimented with an incentive scheme in which offices were incentivised based on outputs. Both studies showed that incentives must be carefully designed because agents can find ways to game the system (Heckman et al., 1997, 2002; Burgess et al., 2012).

Teacher Incentives

Much of the empirical literature on motivating public servants has focused on the education sector and on motivating teachers. The studies on teacher incentives carried out thus far focus on performance pay (bonuses), [1] and have investigated the impacts of both individual and group incentive schemes on student outcomes and on teacher attendance. They attempt to glean whether basing pay on performance, or on closer monitoring, or both, leads to increased effort on the part of teachers, and then to better learning outcomes. They point to positive gains in both respects; however, they also uncover gaming behaviour in some cases.

Lavy (2002) and Lavy (2003) studies a group incentive program and an individual incentive pay program, respectively, in Israel. Lavy documents improved student achievement in both as a result of the schemes. Glewwe et al. (2010) undertake a randomised study in Kenya involving group incentives, and find that student test scores do indeed improve, but that teacher effort may suffer in other dimensions, as teachers were found to be 'teaching to the test'. Duflo et al. (2012) also undertake a randomised trial by installing tamper-proof cameras in Indian schools, coupled with high-powered financial incentives based on reported attendance of teachers. They find that attendance improved significantly, as did student achievement. Muralidharan and Sundararaman (2009) and Muralidharan (2011) study a randomised trial with both a group, and an individual incentive pay scheme based on student test scores in Andhra Pradesh, India. They find that the incentivised schools outperform the control schools, and that the individual incentive treatment group outperformed the group incentive treatment group.

Numerous studies have also been carried out in the United States. Figlio and Kenny (2007) consider individual incentives, and report positive effects on student achievement. However, their data are cross-sectional. Other studies in the US have pointed to evidence

[1] There are some key differences between bonuses and promotions. Firstly, wage increases from promotions are permanent, whereas bonuses are generally one-off payments. Another difference is that generally, promotions are not offered every year, but bonuses often are. They also have different risk properties. These differences have implications both for the management of the firm, and for effort incentives. This thesis will focus on effort incentives.

of gaming, particularly, teaching to the test, providing high calorie lunches on test days, and placing weaker students in special education classes to raise average scores (Jacob and Levitt, 2003; Figlio and Winicki, 2005; Figlio and Getzler, 2006). Understanding the various responses to different kinds of incentive schemes is important.

Teachers in China

A number of papers have used the GSCF data to study various aspects of the education system in China, including education management, financing and decentralisation, school attainment, and educational resources (see GSCF website). Zhao and Glewwe (2010) focus on the determinants of schooling attainment in Gansu, and find that out of the teacher characteristics, teacher experience matters in secondary school. Park and Hannum (2001) study the effects of educational inputs on student achievement in Gansu province, particularly, teachers as an input. They use teacher rank titles as a proxy for teacher quality, as they are meant to capture several dimensions of teacher ability, performance and effort. They find that teacher quality is extremely important, particularly for language scores of students. Part I of this thesis contributes to this literature in an attempt to improve our understanding of how China has achieved such a successful education system.

Other parts of China have also been studied, and these studies have included work on the performance evaluation system for teachers. Ding and Lehrer (2001) examine the evaluation and pay system in Jiangsu province in China. They present a model of a performance measure that combines both objective and subjective elements, and test the model empirically. They find that teacher salaries increase with rank, and that teachers who are male, older, and have more education, tend to be in higher ranks. However, this work does not study the effort incentives from the evaluations and promotions, and they do not have data on the annual performance evaluation scores of teachers.

Promotions as Incentives

The theoretical and empirical literature related to this study is concerned with the internal labour market of firms. This literature regards promotions as having two functions: to sort employees by ability, and to provide incentives for effort (Lazear, 1995).

The idea of promotions as an incentive device is closely linked to the theory of tournaments [2]. Tournament theory states that a firm can motivate employees by having them compete for a reward (a promotion) (Gibbs, 1994). Tournament models of promotion take the Lazear and Rosen (1981) rank-order tournament model as their starting point. Here, two players compete for a promotion (which implies an increase in wages) and the person with the higher output wins it. In their model, effort is unobservable and output depends on effort and a random component. They show that there exists a wage gap that will induce the first-best level of effort. Other papers that model promotion tournaments include Malcomson (1984), Rosen (1986), and Baker et al. (1988).

The literature on sorting of employees into ranks or jobs is based on Sattinger (1975), Rosen (1982) and Waldman (1984), as well as MacLeod and Malcomson (1988). In these models, agents are assumed to be heterogeneous and a job ladder is generated endogenously within the firm. MacLeod and Malcomson (1988) present a promotion model in which ability and effort are private information. The firm sets up an internal labour market with different ranks and higher wages at higher ranks. There is a distribution of ability types a and effort levels e. Disutility of effort is lower with higher ability ($c(e)/a$). The model predicts that employees will sort into ranks according to ability; the people who are promoted are those for whom their ability justifies their effort in working hard and being promoted. People with higher ability also work harder to get the next promotion.

Fairburn and Malcomson (2001) model a case where the firm would like to sort employees into *different jobs*. They note that it is generally a manager who reports on employees' performance and recommends promotions, and that these managers may be susceptible to 'influence activities' on the part of employees that may make the promotions inefficient in the sense that more employees could be promoted than would be optimal. However, if promotions are used in conjunction with bonuses to sort employees into different jobs, then even when influence is possible, effort will be increased [3]. Papers that study promo-

[2]There are two other strands of literature related to incentives. The first is that related to agency models and incentive contracting as developed by Mirrlees (1974), Mirrlees (1976), Hölmstrom (1979), Holmström (1999) and Shavell (1979), which focuses on the tradeoffs between risk and incentives. The efficient contract involves providing the employee with a bonus to elicit optimal effort. The second is the efficiency wage approach in which the firm elicits effort by paying a wage that is above the market-clearing level and threatening to fire the employee if he/she exhibits performance that is too low (Gibbons and Waldman, 2003).

[3]Many other papers study various incentives of managers, including Milgrom (1988), Milgrom and

tions as sorting devices highlight either learning or human capital acquisition as the way of adding dynamics to the promotion process (Gibbons and Waldman, 1999).

Gibbs (1989) extends the Lazear and Rosen (1981) basic tournament model to multi-person tournaments, and incorporates the use of an internal hierarchy in the firm. He presents both cases of identical and heterogeneous contestants. The paper shows that in the case of identical contestants, agents exert effort optimally in response to the promotion rate, the wage difference, and the number of competitors. In a multi-person tournament with heterogeneous agents competing for a series of promotions, employees with higher ability tend to be promoted, so employees sort into ranks according to ability, and the number of competitors and the beliefs on the distribution of skills amongst employees, all have an important influence on effort incentives. Gibbs' model is expounded in two further papers, Gibbs (1996) which compares tournaments and standards, and Gibbs (1991) which presents more fully the incentive effects for employees of varying ability.

Early empirical studies of tournaments used data from auto racing and chicken farming, rather than compensation and firm level data (see Knoeber and Thurman, 1994). Empirical studies of tournaments take the approach of testing tournament theory versus other wage theories such as marginal productivity, to see which one holds for the data (Medoff and Abraham, 1980, 1981; Baker et al., 1994a,b). The log of the wage is generally used as the outcome variable. For example, Baker et al. (1994b) use personnel data from a large firm in the US with multiple levels in the hierarchy and regress wages on rank titles. They find that the ranks explain 70% of variation in wages, and that individuals experience larger wage increases with promotions. They use this as evidence that the firm uses promotions as incentives. However, the effect of those incentives on effort is not studied. Much empirical evidence on tournament theory also focuses on CEO pay as a way to test tournament theory, as the wage gap at the top level is expected to be large (Bognanno, 2001).

However, these studies do not include a performance measure in their estimations and as a result, they do not unpack effort incentives. Furthermore, the studies focus on middle-level managers, generally in the United States, in the private sector. Gibbs

Roberts (1988), and Prendergast and Topel (1996).

(1995) uses the same personnel data as Baker et al. (1994b) along with performance ratings, and focuses on middle managers. He finds that both promotions and bonuses are used as incentives in this firm, and that employees that are repeatedly passed over for promotion have greatly reduced effort incentives. Campbell (2008) looks at non-financial performance measures. He studies managers of a fast food outlet, and he finds that these non-financial measures do indeed factor into promotion decisions. Furthermore, he provides evidence that promotion incentives affect behaviour; stronger incentives lead to greater improvements in qualitative performance indicators. Kwon (2006) studies effort in human capital accumulation but focuses on women at lower levels, and again in a private-sector setting. He finds that in his data, 50% of effort incentives are provided by each of bonuses and promotions, and that the optimal wage contract has both promotions and bonuses.

Part I contributes to all four of these strands of literature. It contributes to the study of incentives in the public sector, and to the literature on teacher incentives, by studying an aspect of public sector wage determinantion that has not been studied before: promotions. Most studies of incentives in the public sector and for teachers have focused on bonuses. In addition, promotions in the public sector (and for teachers) are generally not based on merit and competition. The unique case of teachers in rural China is a novel setting that can provide insights into the viability and design of this alternate incentive system. The chapters also contribute to the study of education in China. Although China's education system is viewed in very high regard, the determinants of its success have not been fully studied. The role played by the incentive system for teachers, who are arguably one of the most important inputs into children's education, is important to understand as it is a big component of the way in which teachers are managed in China. The chapters also contribute an empirical test of a general promotion-based incentive model. With the contribution of a multi-period model to extend the existing one-period models, the study contributes some additional predictions, as well as tests these additional predictions. Finally, the setting is the public sector in a low-income country. I am unaware of any other such studies.

The next section will outline the way in which the promotion system for teachers in rural China works.

2.3 Gansu Province and Teacher Promotions in China

Gansu province is located in a topographically diverse area of North West China, and is one of the country's poorest provinces. As of the 2010 national census, Gansu province had a population of 30.7 million people, 73% of whom live in rural areas. Educational attainment is relatively low compared to the rest of China. By 2000, less than 40% of the population had completed primary school, and less than 25% had completed middle school. School conditions and quality vary across the province; however, most principals report that the majority of the teachers in their school have the required qualifications (GSCF website).

In China, all civil servants compete with their colleagues for promotions. The specific number of ranks and the rules of the promotion contest vary by sector, but the basic design of performance evaluations feeding into promotion decisions is common. 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 one 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 status and regulations. Teachers are all hired as interns and then promoted [4]. Each rank corresponds to a progressively higher salary, and teachers compete for a limited number of spaces for promotions.

In Gansu province, the region being studied, there are specific rules on the years of service required before a teacher can apply for promotion to the next rank level. All teachers begin as interns in their first year, regardless of educational background, and can apply immediately for Primary level 2 or Middle level 3 in the second year of teaching. Teachers with a middle school education can apply for Primary 1 or Middle 2 rank after 4 years, with a vocational college degree after 3 years, and with a university degree after

[4]This is one of the key features of internal labour markets, as defined by Doeringer and Piore (1971).

1 year. Teachers that graduated from a middle school can apply for Primary high level or Middle level 1 after 10 years, with a vocational college degree after 7 years, and those graduating from a university after 5 years. To apply for Middle high level all teachers must have been at Middle 1 rank for at least five years, and teachers with middle school degrees also must have had at least 25 total years of service, those with vocational college degrees need at least 15 years of total service. Teachers with PhDs can apply after one year.

Apart from having to wait the appropriate number of years before applying to the next rank, promotions are also based on meeting some minimum requirements. The main requirement is to achieve either two 'good' evaluation scores or one 'excellent' evaluation score in the *past five years* [5]. These two requirements are consistent for all teachers. There are also other requirements that may be present, and these vary by township. For example, in some townships teachers must also publish an article in a teaching journal. 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. There are also a number of awards that can be won by teachers at the county and township level. Many schools also have their own school-wide teacher awards. In some townships, winning these awards can also help with promotions.

In addition, there are a specified number of positions available at each rank. Each year, the township education bureau announces the number of teachers that can be promoted to each rank level amongst all teachers in the township. In most cases, these quotas are established by the level above, in this case the county Education Bureau. Once a teacher has fulfilled the minimum conditions, she competes with other teachers in applying for rank promotions for the spaces available.

The annual teacher performance evaluations are based on a mixture of four different criteria: 1) student test scores, 2) work ethic/attitude, 3) preparation (lesson plans, homework, reporting etc.) and 4) teacher attendance. China operates within a decentralised education system, and a committee at the township education bureau comprised of the principals, vice-principals and some high ranking teachers in the township, chooses

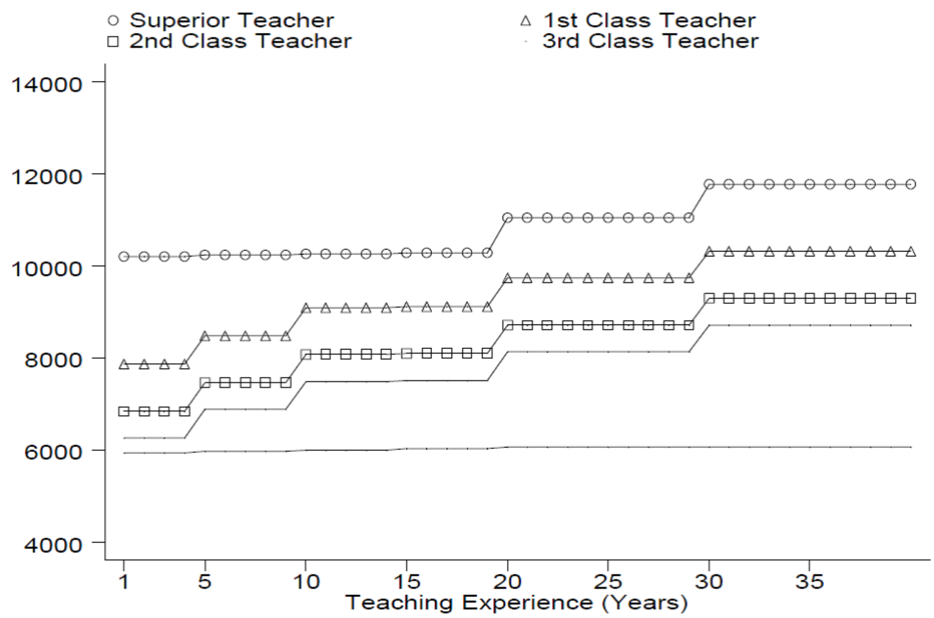
[5] Evaluations are conducted on a four point scale: excellent, good, pass, and fail.

the weight of each of the four criteria in the annual evaluation. This means that teachers are evaluated on several dimensions of their performance, some objective and some subjective. The committee conducts the annual evaluations on the basis of classroom observation, questionnaires from other teachers at the school, questionnaires from students at the school, data on student test scores, and principal reports of the teacher's attendance, preparation and attitude. Each component is worth a specified number of points (these vary at the township level), and the points for each category are added up for all teachers and the teachers are then ranked. The top 10% of teachers *in each school* can receive excellent evaluation scores, and the next 10% good evaluation scores. The names of the recipients of excellent and good evaluation scores are announced at an annual teacher meeting. Although some categories can be manipulated, categories such as test scores, are less easy to manipulate. Furthermore, since there is an explicit quantitative formula and the results are announced, it makes it less likely that committees will manipulate evaluation scores. Of course this will also vary by township. Promotions are also conducted by the committee at the township education bureau, and a similar system of points is used that incorporates the evaluation scores, publications, awards and other township specific requirements.

Principals are also evaluated annually and can receive promotions and salary increases based on the performance of their school. A committee from the township visits the school every year and collects test score data as well as administers a questionnaire to teachers about the principal's management of teachers, looks through the school accounts etc. As a result, we would expect that this system of promotions is incentive compatible, thereby avoiding to a large extent the problems of influence activities for which promotions are often criticised (see Fairburn and Malcomson, 2001).

Ding and Lehrer (2001) provide an excellent description of the way in which wages are determined in the Chinese education system. Figure 2.1 shows the salary schedule for secondary school teachers in a county in Jiangsu province in 1998. Wages for primary school teachers and for teachers in other provinces are determined in the same way, although the base rates and wage increases may be different. We can see that starting salaries are higher as the rank levels increase, and that the difference is greater for higher rank levels. As such, there is a great incentive for teachers to increase their rank levels through promotions as early as possible. The authors also note that there is a sense of prestige derived from high rank levels, and that teachers are also able to obtain higher wages from activities other than teaching, such as tutoring outside of school. It is important to note that these wage gaps vary at the county level. In addition, it is difficult for teachers to move to other counties. Teachers are moved only for very specific reasons, such as a spouse in another county, or because they have subject-specific skills that are needed in another county. This is important because it means that teachers cannot choose to move to schools that have higher wage increases.

Figure 2.1: Sample Salary Scale for Teachers



Source: Ding and Lehrer (2001).

The next section will discuss the data that is used to look at promotion incentives in rural China.

2.4 The Gansu Survey of Children and Families (GSCF)

The Gansu Survey of Children and Families (GSCF) is a longitudinal study focussing on children and education in Gansu province in rural China. It is the most comprehensive survey on education in rural China, and as such, much information is available that would not be in most surveys. One hundred villages in 42 townships in 20 counties were sampled. Because this is a sample of rural schools, there is generally only one main primary school in each village, and one middle school in each township, and these were the schools that were sampled. In conjunction with the main child questionnaire, a detailed school survey was carried out in each of three waves (2000, 2004 and 2007) that included interviews of the principal at each school that the children attended, the child's homeroom teacher, as well as interviews of (in most cases) all the teachers at the school.

A panel was constructed for each teacher for the years 2003-2006 from the third round of data that was collected in 2007. The data set was constructed in this way because it was in the third round that data was collected on evaluation scores from 2003-2006, as well as full promotion histories of all the teachers. The 200 schools that were surveyed included interviews of almost 2400 teachers. The average sampling rate of teachers in the schools is 85% and only 23% of the schools have a sampling rate lower than 25% [6]. In 2007, individual and household questionnaires, as well as achievement tests, were also administered to 1500 children aged ten to fifteen.

Table 2.1 shows the mean and standard deviation of the weights placed on the four performance evaluation criteria for schools in our sample. Student test scores generally have the highest weight (34.2% on average), followed by preparation, attitude, and finally attendance.

Table 2.1: Weights placed on performance evaluation criteria

Criteria	Mean percentage weight	Standard deviation
Attitude	23.22%	10.57
Preparation	29.45%	11.39
Attendance	13.16%	5.82
Test scores	34.17%	15.59

Tables 2.2 and 2.3 contain the results from wage regressions for Primary School and for Middle school, respectively, that estimate the wage premiums for different ranks. The first column is a regression of the log of monthly wages on rank level dummies. The

[6]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.

second column also includes controls for experience in addition to the rank levels. The final column includes dummy variables for the highest level of education completed, as well as the performance evaluation score from the most recent year. All regressions include county fixed effects, as well as standard errors clustered at the county level. The omitted category is that of Primary 2 at the primary level, and that of Middle rank 3 at the Middle school level. Rank level premiums are significant and increasing as a teacher gets promoted. At the Primary level, the increase from Primary 2 to Primary 1 is 27% and to Primary high a further 28%. At middle school, the increase between levels is 20%. Controlling for individual characteristics reduces these premiums, but they remain high, especially at the top ranks.

Table 2.4 provides some basic data on the teachers in the sample. There are 2,353 teachers, 39% of whom are female. The data is broken down by current rank of the teacher. We see that in accordance with the findings of Ding and Lehrer (2001) for Jiangsu province, teachers in higher ranks tend to be male (the proportion increases sharply over ranks), tend to be older and have more teaching experience, and tend to have higher levels of education. In terms of measures of 'ability', the teachers in higher ranks tend to have performed slightly better on their teaching tests (education, psychology and language tests, which are taken at the beginning of their teaching career), and tend to have received more teaching awards over the course of their career. A greater proportion of higher rank teachers have also published papers on teaching methods in teaching journals. Some of these differences may be related to the greater number of years of experience. Looking at time use, it seems teachers at lower levels are spending more hours per week on grading, lesson plans, and with students. Again this could be a learning effect; teachers that are in higher ranks have been teaching for longer and thus may need to spend less time on such tasks. They also teach slightly fewer classes. Interestingly, there is quite a sharp decline in Middle high rank, with these teachers putting in the least amount of time on various tasks. So it seems that there is a correlation between rank and teacher ability; better teachers tend to get promoted, but the effect on effort remains to be seen, thus providing scope to study teacher effort in greater detail.

Table 2.5 provides a breakdown of evaluation scores by rank of the teacher for the years

Table 2.2: Wage Premia - Primary School

Outcome: log of monthly wage	(1)	(2)	(3)
Primary 1	0.267***	0.166***	0.118***
	(0.031)	(0.047)	(0.042)
Primary high	0.433***	0.243***	0.180***
	(0.030)	(0.057)	(0.053)
Experience		0.015**	0.017***
		(0.006)	(0.005)
Experience squared		-1.778e-4	-1.492e-4
		(1.394e-4)	(1.145e-4)
Education - middle school			0.029
			(0.044)
Education - high school			0.010
			(0.043)
Education - vocational middle school			0.105**
			(0.050)
Education - vocational college			0.102
			(0.069)
Education - four year degree			0.084
			(0.056)
Education - graduate school			0.218*
			(0.116)
Year end assessment result 2006			0.002
			(0.010)
Observations	1066	1066	1060
R^2	0.403	0.447	0.478

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.

Table 2.3: Wage Premia - Middle School

Outcome: log of monthly wage	(1)	(2)	(3)
Middle 2	0.176***	0.083***	0.066***
	(0.027)	(0.025)	(0.025)
Middle 1	0.372***	0.176***	0.155***
	(0.028)	(0.033)	(0.032)
Middle high	0.577***	0.343***	0.302***
	(0.052)	(0.056)	(0.051)
Experience		0.023***	0.022***
		(0.004)	(0.005)
Experience squared		-3.848e-4***	-3.378e-4***
		(1.018e-4)	(1.248e-4)
Education - high school			-0.089
			(0.057)
Education - vocational middle school			-0.018
			(0.025)
Education - vocational college			0.079
			(0.132)
Education - four year degree			-0.004
			(0.024)
Education - graduate school			0.008
			(0.029)
Year end assessment result 2006			-0.005
			(0.007)
Observations	950	950	945
R^2	0.562	0.610	0.616

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.

Table 2.4: 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	331	163	553	354	133	525	281	13	2,353
Male	43%	42%	55%	75%	51%	63%	83%	85%	61%
Female	57%	58%	45%	25%	49%	37%	17%	15%	39%
Basic characteristics									
Average Age	29.09	28.3	36.7	48	26.8	32.3	40.6	47.2	36.5
Average Years teaching	7	7	16.3	27.6	4	10.1	19.7	27.6	14.4
Education - primary	1%	1%	0%	0%	0%	0%	0%	0%	0.2%
Education - middle school	5%	2%	3%	3%	1%	2%	1%	0%	2.5%
Education - high school	23%	7%	22%	48%	2%	2%	10%	0%	17.6%
Education - vocational middle	20%	55%	63%	45%	13%	32%	47%	54%	42.3%
Education - vocational tech	1%	1%	0%	0%	0%	0%	1%	0%	0.3%
Education - vocational college	45%	34%	12%	4%	83%	57%	38%	46%	34.2%
Education - four year degree	6%	1%	0%	0%	2%	7%	3%	0%	3%
Time use									
Hours per week spent grading	10.4	11.62	9.62	9.24	9.92	9.13	9.16	7.38	9.63
Hours per week on lesson plans	9.98	9.66	9.18	9.11	9.61	9.23	9.7	7.92	9.38
Published a paper (proportion)	12.1%	25.2%	46.5%	35.9%	17.3%	45.9%	57.3%	100.0%	38.3%
Number of classes taught	17.57	19.34	18.05	17.79	14.34	13.7	13.13	9.84	16.02
Hours/week spent with students	14.18	13.67	12.09	11.24	14.58	12.82	12.44	7.03	12.64
Wages									
Monthly wage (CNY)	606	974	1,287	1,511	1,015	1,271	1,535	1,866	1,218

2003-2006. There is four years of data on many teachers, and the total number of teacher-year observations is 7,894. Not many teachers fail, and the majority of the evaluation scores is a score of ‘pass’. There is also evidence of the ‘curve’ described earlier; the proportions of teachers receiving each type of evaluation score remain quite similar across ranks. Thus, where a teacher falls in the distribution is of particular importance. Table 2.6 reports the average monthly salary (and its standard deviation) for teachers at each rank level. As expected, salaries increase with rank. The biggest jumps are from Primary 2 to Primary 1, followed by Middle 3 to Middle 2, and then Middle 1 to Middle high. As Table 2.5 shows, the salary increases are 32.4%, 25.1%, and 21.5% respectively on average.

Figure 2.2 and Table 2.7 (in the appendix) show how long promotions tend to take for each of the rank levels. Figure 2.2 is a set of Kaplan-Meier survival functions for how long a teacher ‘survives’ at each rank level. Each panel is for the number of years from one rank level to the rank level above. Most people that get promoted from Primary level 2 to Primary level 1 do so around 4-5 years after having entered Primary 2 rank; promotion is nearly automatic. We see that practically everyone is promoted at some point, and that the maximum number of years at this level until promotion is 24 years in the sample. For Primary level 1 to Primary high level, the distribution in terms of number of years until promotion is far more spread out. There seems to be a steady stream of promotions, but most people get promoted to Primary high level between 5 and 11 years after promotion to Primary 1. Just under 1/3 of Primary 1 teachers never get promoted to Primary high. At Middle level 3 rank, every teacher is promoted within 9 years, but the majority of promotions appear to be between two and five years. At this level, promotion is also almost automatic. The promotions from Middle 2 to Middle 1 look similar to those of Primary 1 to Primary high. They are more spread out, but the majority occur between 5 and 10 years. Approximately 25% of those at Middle level 2 are not (or have not yet been) promoted to Middle level 1. Very few people are promoted from Middle 1 to Middle high level. In this sample, less than 25% have been promoted to Middle high level. The promotions take considerably longer; anywhere above nine years. Once the teachers reach this level though, as Table 2.4 demonstrated, there may either be

Table 2.5: 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	-	-	3	0.4	4	0.2	1	0.1	-	-	5	0.3	1	0.1	-	-
Pass	236	43.62	396	48.4	973	49.9	526	50.7	473	56.1	906	54.9	802	51.1	16	55.2
Good	215	39.74	291	35.5	675	34.6	378	36.5	253	30.0	487	29.5	502	32.0	10	34.5
Excellent	90	16.64	129	15.8	297	15.2	132	12.7	117	13.9	251	15.2	263	16.8	3	10.3
Total	819	100.0	1,949	100.0	1,037	100.0	843	100.0	1,649	100.0	1,568	100.0	29	100.0		

Note: Interns include Primary and Middle school interns.

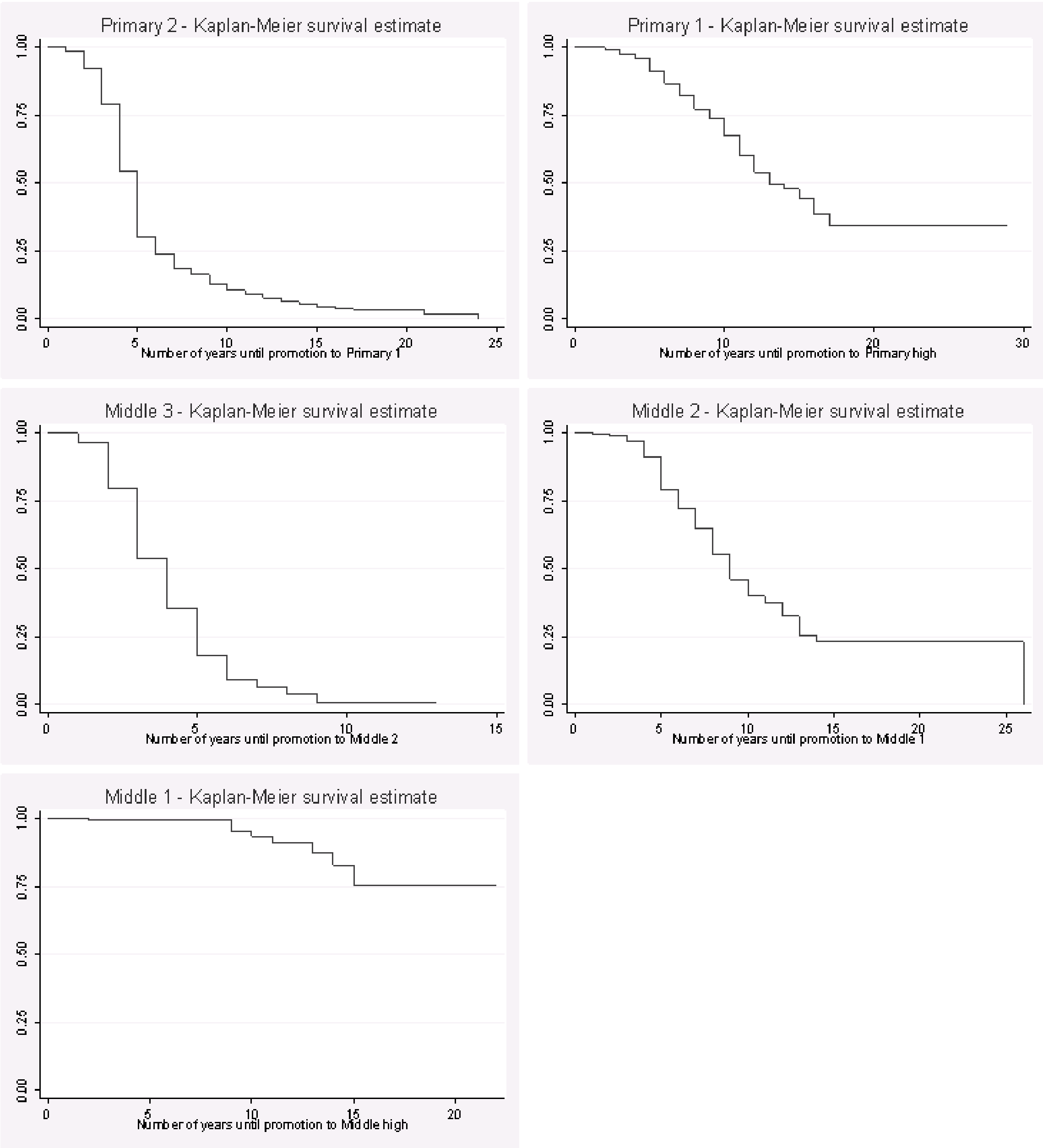
Table 2.6: Salaries across ranks (CNY / month)

	Average salary	Standard deviation	Increase
Primary 2	974.15	261.23	-
Primary 1	1289.63	230.19	32.4%
Primary high	1511.27	271.48	17.2%
Middle 3	1015.87	235.4	-
Middle 2	1270.69	229.79	25.1%
Middle 1	1534.88	233.46	20.8%
Middle high	1865.62	263.54	21.5%

a decline in effort, or a learning effect. Because of the differences in the promotion rates and timing across different ranks, this provides a good opportunity to study differences in effort due to different promotion rates, as well as to study how effort evolves over time.

The next chapter will present a model of promotions that allows me to develop an estimating equation to test whether promotions motivate effort. It will also estimate the effects of promotion contests with different characteristics on average levels of teacher effort.

Figure 2.2: Kaplan-Meier Survival Functions



Source: Gansu 2007 Data.

2.5 Appendix Chapter 2 - Additional Tables

Table 2.7: Proportion of teachers in the sample promoted after x years

	Primary 2 to Primary 1	Primary 1 to Primary high	Middle 3 to Middle 2	Middle 2 to Middle 1	Middle 1 to Middle high
1 year	2.08	-	5.05	1.15	-
2 years	7.62	3.85	21.3	1.72	11.11
3 years	15.47	4.33	28.88	4.6	-
4 years	27.02	4.81	19.13	10.92	-
5 years	24.71	11.54	14.08	20.69	-
6 years	5.54	11.54	6.5	10.92	-
7 years	4.62	9.62	1.81	9.2	-
8 years	2.31	10.58	1.44	12.07	-
9 years	3	6.25	1.81	10.34	33.33
10 years	1.85	10.1	-	6.32	11.11
11 years	0.92	9.62	-	2.3	11.11
12 years	1.15	6.25	-	4.02	-
13 years	1.15	3.37	-	4.02	11.11
14 years	0.69	1.44	-	1.15	11.11
15 years	0.69	1.92	-	-	11.11
16 years	0.23	2.88	-	-	-
17 years	0.23	1.92	-	-	-
18 years	0.23	-	-	-	-
19 years	-	-	-	-	-
20 years	-	-	-	-	-
21 years	0.23	-	-	-	-
22 years	-	-	-	-	-
23 years	-	-	-	-	-
24 years	0.23	-	-	-	-
25 years	-	-	-	-	-
26 years	-	-	-	0.57	-
Longest without promotion to next level	20 years	29 years	20 years	20 years	22 years

Chapter 3

Do Promotions Motivate Effort?

3.1 A Model of Promotions as Incentives

This section will outline a model of promotions as incentives. The following exposition is based primarily on Gibbs (1995) and Gibbs (1989). These are the only models in the literature that incorporate multiple contestants into a tournament model. They are well-matched to the context of Chinese teachers, with the same promotion structure and a performance evaluation feeding into the promotion decision.

A school wishes to motivate teachers' effort by offering rewards from promotion. A school hires teachers into the lowest rank and promotions are then offered sequentially, based on performance. The hierarchy is assumed to be in a steady state. Teachers contribute effort, e, to their job (this includes diligence, care etc., not just hours), but effort is costly and gives them disutility $C(e)$ which is increasing in e at an increasing rate so that C' and $C'' > 0$. Performance of teacher i is measured by $q_i = e_i + \pi_i$, where q_i is output, e_i is effort and $\pi_i = \epsilon_i + \mu$ is a stochastic error term, which comprises individual error (or 'luck') and a common error. Assume that ϵ_i is i.i.d. and has a symmetric distribution around a mode of 0 and $E(\pi_i) = E(\epsilon_i) = E(\mu) = 0$. Let $F(\epsilon)$ be the CDF and $f(\epsilon)$ be the PDF of the luck term. We can further assume that $d^2q/de^2 = 0$ so there are no scale economies in effort. Teachers do not differ in terms of their ability. Effort is chosen before the 'luck' draw is known for that time period.

When n teachers are competing for k promotion slots, we have a tournament structure.

Each teacher's probability of promotion depends on his/her effort and the efforts of others. A teacher's probability of promotion can be written as $p = p(e, \underline{e})$ where $\underline{e}$ is a vector of other teachers' efforts. The school offers a benefit/reward from promotion, ΔEV, which is the change in expected lifetime utility gained from winning a promotion relative to not being promoted (this could be in the form $(W_2 - W_1)$, different wages at different ranks where $W_2 > W_1$ if rank 2 is higher than rank 1, but could also include non-pecuniary benefits). ΔEV also includes the option value of possible future promotions.

The teacher chooses effort to maximise the one-period expected reward minus the disutility of effort:

$$\max_{e} \{p(e, \underline{e})\Delta EV - C(e)\} \tag{3.1}$$

The first order condition sets the marginal cost of effort equal to the gain in utility, times the marginal effect of effort on the probability of promotion, $\frac{dp}{de}$:

$$C'(e) = \frac{dp(e, \underline{e})}{de}\Delta EV \tag{3.2}$$

Gibbs terms $\frac{dp}{de}$ the marginal probability of effort (MPE). As (3.2) shows, incentives depend on two things: ΔEV and $\frac{dp}{de}$, the sensitivity of the probability of promotion to changes in effort. The larger the prize (ΔEV), the greater the incentives. The second order condition is satisfied as long as the slope of the probability function does not increase too quickly in effort:

$$\frac{d^2p}{de^2}\Delta EV - C''(e^*) \leq 0 \tag{3.3}$$

It is clear that the younger individuals are at the time of the tournament, the greater the effort incentive, because being promoted produces a longer stream of income and utility differences. In addition, since all teachers are identical, e^* will be the same for everyone, and it represents average effort. The prediction above then suggests that if the average age of teachers in one tournament is lower than that in another, or if the proportion of female teachers is lower, then we should see average effort being higher, given that women retire five years earlier than men in China.

In order to study $\frac{dp}{de}$, the structure of the promotion contest must be specified. The promotion rate (which is fixed) is $p^* = k/n$. To win a promotion, the teacher must place k^{th} or higher, so he/she must beat at least $n - k$ co-workers in pair-wise comparisons of performance.

The probability that teacher i beats teacher g is $pr(q_i > q_g) = pr(e_i + \epsilon_i + \mu > e_g + \epsilon_g + \mu) = pr(\epsilon_g < e_i - e_g + \epsilon_i) = F(e_i - e_g + \epsilon_i)$ [1]. A symmetric Bayesian Nash Equilibrium (BNE) is appropriate for identical teachers and optimal effort for all teachers is e^*. The probability that teacher i beats any other teacher is $F(e_i - e^* + \epsilon_i)$ and the probability that teacher i loses to any opponent is $1-F(e_i-e^*+\epsilon_i)$. To finish j^{th} from the top out of n teachers, one must beat $n-j$ teachers and lose to $j-1$ teachers. The number of ways to choose $n-j$ elements from $n-1$ is $\binom{n-1}{n-j} = \frac{(n-1)!}{(n-j)!(j-1)!}$ and the probability of promotion conditional on ϵ_i is the sum of the conditional probabilities of placing first through k^{th}. Integrating out ϵ_i gives the unconditional probability of promotion:

$$p(e, \underline{e}^*) = \sum_{j=1}^{k} \binom{n-1}{n-j} \int F(e_i - e^* + \epsilon_i)^{n-j}[1 - F(e_i - e^* + \epsilon_i)]^{j-1} f(\epsilon) d\epsilon \tag{3.4}$$

If $p^* = 0$ or 1 ($k = 0$ or n) so that either everyone or no-one is promoted, then $p(e, \underline{e})$ is constant and the MPE and incentives are zero. Nobody has an incentive to work.

To solve for the MPE we need to differentiate with respect to e, and we can rearrange and substitute in the symmetric NE condition $e_i = e^*$. We can also use change of variables $\Psi \equiv F(\epsilon)$. Since $f d\epsilon = dF = d\Psi$, and $\epsilon = F^{-1}(\Psi)$ we can write [2]:

$$\frac{dp}{de} = \int (n-k) \binom{n-1}{n-k} \Psi^{n-k-1}(1-\Psi)^{k-1} f(F^{-1}(\Psi)) d\Psi \tag{3.5}$$

Equation (3.5) shows that the MPE is independent of equilibrium effort; it depends only on k and n. Incentives come from the increase in the likelihood of winning the promotion when effort is increased at the margin. Holding efforts of competitors fixed,

[1] The common error term cancels out.
[2] See Section 3.6 in the Appendix for full derivation.

changing effort by a small amount has the effect of decreasing the luck required to beat the luck of the k^{th} highest scorer. The integrand in equation (3.5), except for $f(F^{-1}(\Psi))$, has the form of a Beta density function $\beta(k, n-k)$. Thus, Ψ has the effect of a Beta-distributed random variable and,

$$\frac{dp}{de} = E_{\Psi} f(F^{-1}(\Psi)) \tag{3.6}$$

When $f(\epsilon)$ is symmetric and unimodal at 0 then $f(F^{-1}(\Psi))$ is symmetric over [0, 1], unimodal at 1/2 and $f(F^{-1}(1/2)) = f(0)$.

Gibbs (1989) then provides a number of propositions and predictions using properties of Beta densities, some of which are summarised here, providing the intuition [3]. The most relevant properties of Beta densities are as follows:

1. For $k, n-k > 1$, $\beta(k, n-k)$ is uni-modal at $(n-k-1)/(n-2)$ with mean $(n-k)/n = 1 - p^*$;

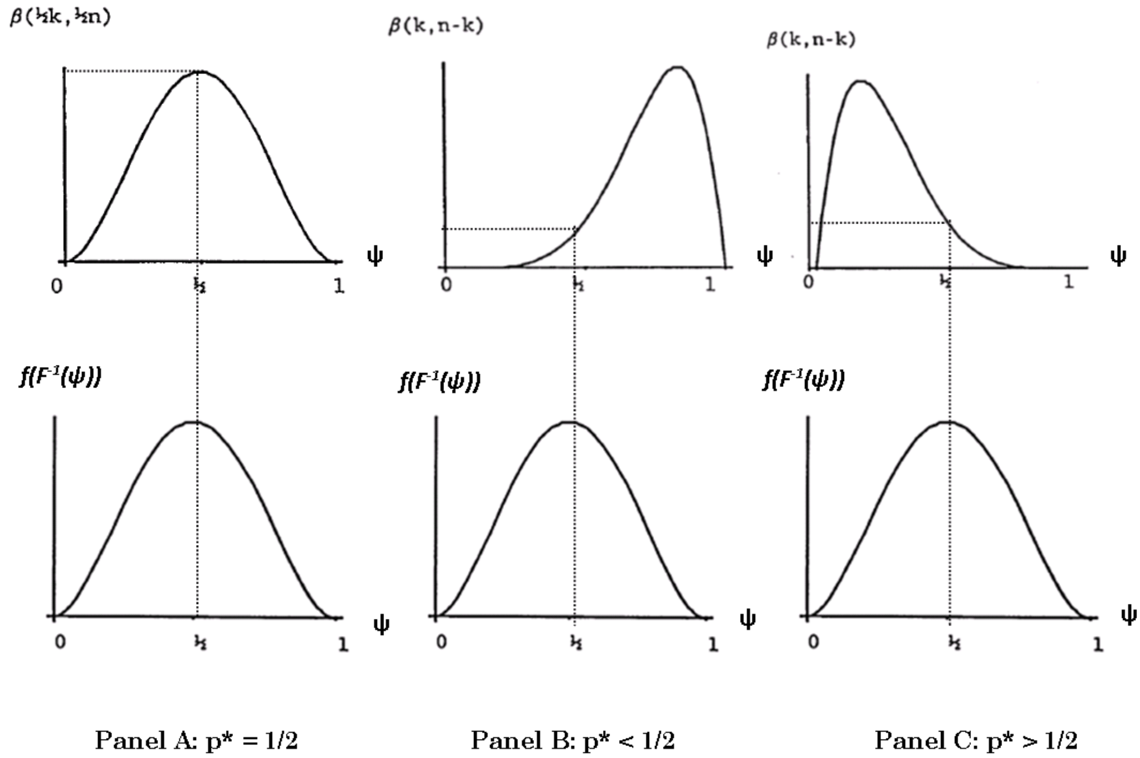
2. $\beta(k, n-k)$ is symmetric about one half for $k \geq 1$ if and only if $k = n/2$ (when half of the teachers are promoted). $\beta(k, n-k)$ is skewed to the right if $k > n/2$, is skewed to the left if $k < n/2$, and becomes increasingly skewed as k/n deviates from 1/2 in either direction. Figure 3.1 below illustrates. Panel A is for $p^* = k/n = 1/2$, Panel B is $p^* < 1/2$ and Panel C is $p^* > 1/2$; and,

3. Holding k/n fixed, as n (and k) approaches infinity, $\beta(k, n-k)$ approaches a distribution with all of its mass at $1 - p^*$.

The bottom half of Panels A, B and C show the (re-scaled) distribution of luck draws, which is the same regardless of the promotion rate. Middle values of luck are most likely; very high or very low draws of luck are less likely. The MPE is a weighted sum of the Beta density values (from the top figures), where the weights are the f (luck) distributions plotted in the bottom figures. The top half of Panels A, B, and C shows the MPE for every draw of luck when $p^* = 1/2$, $p^* < 1/2$, and $p^* < 1/2$, respectively. The weighted average is highest when $p^* = 1/2$ since in this case high Beta density values are given

[3] The reader is referred to Gibbs (1989) for proofs of the predictions.

Figure 3.1: Effort Incentives: $p^* = 1/2$, $p^* < 1/2$, $p^* > 1/2$



high weight and low Beta density values are given low weight. This means promotion incentives are highest when the promotion rate is 1/2. Intuitively, all teachers have an equal chance (50%) of promotion ex-ante, so many teachers are on the margin of being promoted versus not being promoted and small amounts of luck can decide who wins the promotions. It is in this case that effort makes the biggest difference to one's probability of promotion. If the promotion rate were greater than 1/2, a teacher would need a low draw of luck in order not to be promoted. Incentives are lower here, since more than half of teachers will get promoted and low values of luck are unlikely, so the teacher does not have to work as hard to be promoted. Similarly, if the promotion rate was less than 1/2, a teacher would need a high draw of luck in order to be promoted. Incentives are lower here as well, since few teachers will be promoted and high draws of luck are unlikely so the teacher's effort would not be very productive in getting the teacher promoted. When the promotion rate diverges from 1/2, high Beta density values are given low weight. Effort incentives are highest when the promotion contest is close.

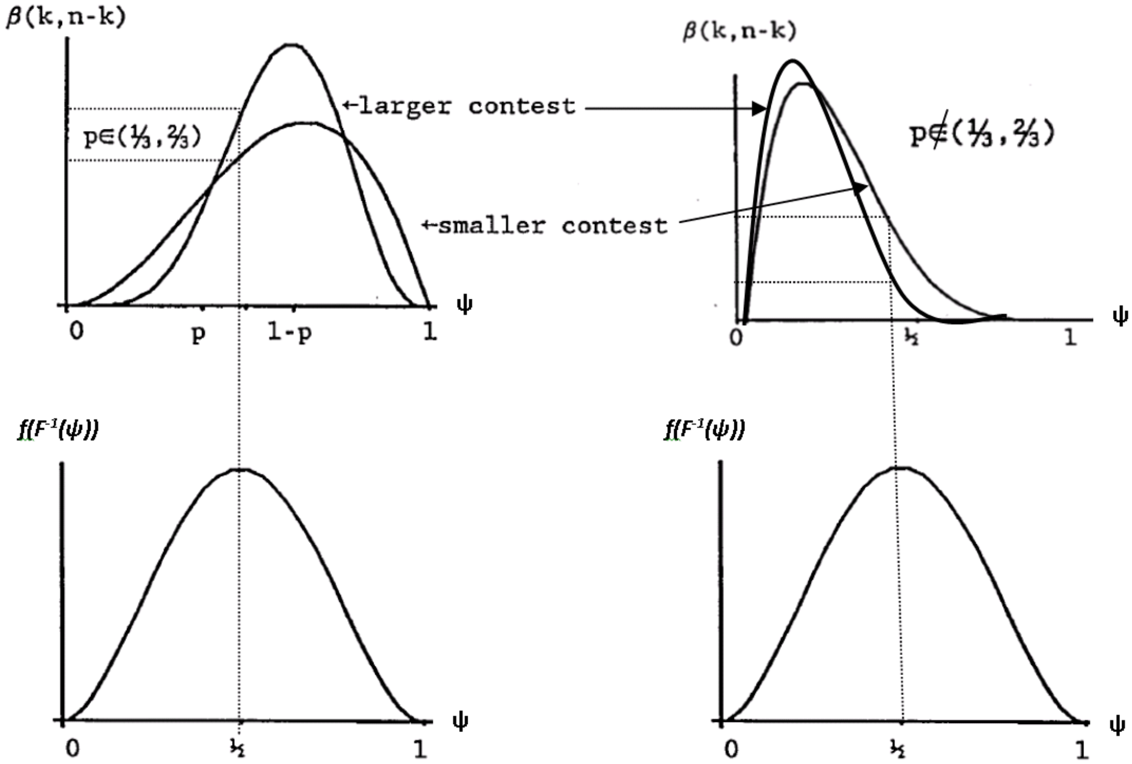
I will now describe predictions of the model regarding the number of competitors (n) and the number of people promoted (k). Changes in k or n change the skew of the Beta densities (property 2 above) because they change p^*, and the mode of the Beta density is $1 - p^*$ (property 1 above). In a contest where fewer than half are promoted ($p^* < 1/2$), if another person is added to the contest ($n + 1$ contestants instead of n) leaving k fixed, then effort is less likely to affect your chances of promotion (dp/de will be lower) since p^* will be even further from 1/2. The Beta density gets skewed further left. If more than half get promoted ($p^* > 1/2$) then adding another person means effort is more likely to affect your chances of promotion, as adding another contestant will reduce p^* closer to 1/2 and the Beta density will be less skewed. Intuitively, as promotion becomes more or less likely (p^* approaches 0 or 1), more extreme good or bad luck is needed to win or lose the promotion. Extreme luck is less likely than luck draws closer to 1/2, so marginal effort is less likely to make a difference.

Another result is that when f is symmetric and unimodal, adding one more winner ($k+1$ instead of k winners) while keeping the number of contestants, n, the same, increases the MPE for $k < 1/2(n - 1)$, and decreases it for $k > 1/2(n - 1)$, a special case of which is $k = 1/2n$ or $p^* = 1/2$. If less than half of people get promoted, then adding another winner will increase the effect of effort on your probability of promotion as p^* will be closer to 1/2, and the reverse is true for more than half being promoted (p^* will be further from 1/2).

Finally, increasing the number of contestants, n, keeping the probability of promotion k/n fixed, means that the Beta density gets taller (property 3 above) at $1 - p^*$. This raises the MPE if $p^* = 1/2$, and generally does so for middle values of p^* as well (between 1/3 and 2/3) [4]. For values of p^* closer to 0 or 1, increasing n lowers the MPE. As Figure 3.2 below shows, the weighted average is higher when n increases for middle values of p^*, and is lower for values of p^* closer to zero or one. The greater the number of contestants, the smaller the variance in the order statistic that must be beaten to win a promotion. For middle values of p^*, smaller amounts of luck are more likely to affect which teacher wins the promotion.

[4] Gibbs (1989) conducts numerical analysis and finds that increasing the number of contestants keeping

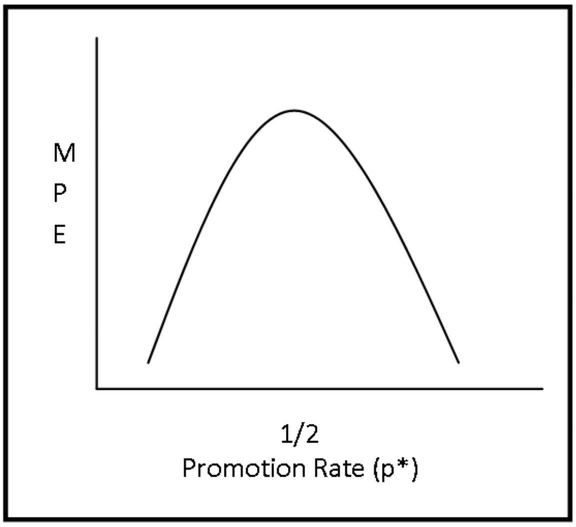
Figure 3.2: Effort incentives: p^* and the size of the contest



Gibbs (1989) also presents predictions on effort for promotions in which teachers must beat a standard instead of compete in a tournament. In this case, the number of teachers is not predicted to matter, as other teachers' actions do not matter for an individual teacher's promotion. This is the only difference in predictions between the tournament and the standards models.

In sum then, holding the prize fixed, **incentives rise as the probability of promotion, p^*, rises from 0 to 1/2, incentives peak at $p^* = 1/2$, and incentives decrease as p^* rises from 1/2 to 1**. The model predicts an inverted U-shaped relationship between incentives and the probability of promotion. Intuitively, if the probability of promotion is exactly half, each teacher is on the margin of being promoted versus not being promoted. It is in this case that effort is most likely to affect the probability of promotion, and incentives are sharpest. Thus, any change in k or n that brings p^* closer to 1/2 causes incentives to increase. In addition, the effect on incentives is symmetric with respect to $|p^* - 1/2|$, the absolute difference of the promotion rate to 1/2. Thus if p^* is very high or very low we can expect very low effort.

Figure 3.3: Effort and the Promotion Rate



In this thesis I do not focus on what the optimal promotion structure should be. Here, the promotion system is fixed (the combinations of p^* and ΔEV are given exogenously). However, it is interesting to note that when the promotion system is being chosen, the

the promotion rate fixed never decreases incentives for promotion rates between 1/3 and 2/3.

further the promotion rate is from 1/2, the higher the ΔEV should be. This is because for very high and very low promotion rates, effort incentives are low. Thus, higher wage differences tend to be offered in order to keep competitors motivated. For very high promotion rates, with high wage increments, even the slightest chance that one could lose the promotion is an incentive to work hard. For very low promotion rates, a high wage increase can motivate effort even though promotion is unlikely; the expected value of the future stream of income can still be high.

To summarise, the Gibbs model predicts the following:

1. Teachers will exert effort in order to win promotions;
2. The larger the expected prize (wage differential), the higher will be effort;
3. Incentives are sharpest when $p^* = 1/2$ and they decrease as p^* moves away from 1/2 in either direction;
4. For values of p^* between 1/3 and 2/3, increasing the number of contestants while leaving $p^* = k/n$ fixed increases incentives, and for values of p^* closer to zero or one, incentives decrease;
5. Contests in which the average age of teachers is higher or in which the proportion of female teachers is higher have lower effort incentives; and,
6. The further p^* is from 1/2, the higher the wage differential tends to be.

The next section will present the empirical specification that will be used to test the model and Predictions 1 - 5. Prediction 6 will not be tested in this data as the promotion system is set by the government and is taken here to be exogenous.

3.2 Testing Promotions

The model specified above leads to a number of testable implications. This section will outline how the model's predictions will be tested, and will discuss some potential shortfalls of the estimation strategy, as well as how these will be addressed.

The model has two main behavioural equations. The first regards effort. From the FOC (equation (3.2)) we see that effort depends on the gain in expected utility (in this case, the difference in wages between two rank levels that a teacher would get if promoted), and the marginal effect of effort on promotion probability (MPE). The MPE depends on the number of competitors (n) and the number promoted (k), or, the promotion rate ($p^* = k/n$). Thus,

$$\begin{aligned} e &= f(\Delta EV, MPE), MPE = h(n, p^*) \\ e &= e(\Delta EV, n, p^*) \end{aligned} \tag{3.7}$$

The second behavioural equation regards promotion. This is the main idea behind the model and is the one I will focus on in this chapter. The equation states that promotion is driven by a teacher's own effort and the effort of other teachers. The model maintains that higher effort increases the probability of promotion. That is,

$$p_i = p(e_i, \underline{e}) = p(e_i^*) \tag{3.8}$$

where p_i is an individual teacher's probability of promotion, e_i^* is the teacher's optimal effort (which is the same for everyone in the same rank and township and so represents average effort). We can estimate (3.8) as follows:

$$pr_{irc,t+1} = \varphi_0 + \varphi_1 e_{irct} + \varphi_2 p_{rc}^* + \varphi_3 n_{r\tau t} + \omega_c + \epsilon_{irct} \tag{3.9}$$

where $pr_{irc,t+1}$ is a dummy variable for whether teacher i is promoted at time $t+1$ at rank r in county c, e_{irct} is the effort of the teacher at time t, p_{rc}^* is the promotion rate at rank r in county c (I use the average historical rate at that rank in a county), $n_{r\tau t}$ [5] is the number of teachers in the same rank in the same township (τ) , ω_c is the unobservable county factor, and ϵ_{irct} is the error term. The historical promotion rate and number of teachers are included as controls for heterogeneity in contests (differences in the expected promotion probability and for heterogeneity in the size of contests between rank levels).

[5] The contests are between teachers in the same township, but the wages and promotion rates are set at the county level.

The two behavioural equations present a problem of simultaneity. Increasing effort increases the probability of promotion, and it is the promise of promotion that increases effort. To solve this problem, I use instrumental variables, and propose using wages as an instrument. This would capture ΔEV. Specifically, I maintain that the increase in wages between levels is both an informative and a valid instrument. It is informative because it is the promise of a higher wage that motivates teachers to work harder for a promotion. I will be able to test whether this is the case from the first stage regressions. The change in wages between levels is also a valid instrument. The wage structure is fixed and is set by the county government; it is exogenous to teachers' influence. China's education system is highly decentralised. It is the job of the local school board to implement the national curriculum, and county governments have significant authority in setting teachers' wage rates. The wage rates that are set reflect availability of fiscal revenue (Ding and Lehrer, 2001), and can also reflect personal preferences of current leaders. Furthermore, wages affect promotions only through effort, because a teacher can access an increased wage only by working hard to get a promotion. A regression of various controls, and the measure of teacher effort (the performance evaluation scores) reveals that neither of these systematically influences wages (see Tables 2.2 and 2.3). In addition, there is no teacher's 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 at the county level are taken to be exogenous to teachers.

Equation (3.9) will be estimated with evaluation scores as a proxy for effort, and the evaluation scores will be instrumented using change in log wages at the county level for each rank:

$$\begin{aligned} &First\ stage: \ ev_{irct} = \zeta_0 + \zeta_1 \Delta w_{rc} + \zeta_2 p^*_{rc} + \zeta_3 n_{r\tau t} + \omega_c + \upsilon_{irct} \\ &Second\ stage: \ pr_{irc,t+1} = \varphi_0 + \varphi_1 \widehat{ev_{irct}} + \varphi_2 p^*_{rc} + \varphi_3 n_{r\tau t} + \omega_c + \epsilon_{irct} \end{aligned} \tag{3.10}$$

where ev_{irct} is the evaluation score of teacher i at time t at rank r in county c, Δw_{rc} is the change in the log of the average wage at the county level for that rank, and $\widehat{ev_{irct}}$

are the predicted values from the first stage regression.

Wages and promotion rates are set at the county level. We may then be worried that there are unobserved factors at the county level (ω_c) that affect the way in which the wage rates and promotion rates are chosen. Consequently, I also include county dummies and estimate a county fixed effects model, and as a result, time invariant county level unobservables are swept away:

$$\begin{aligned} &First\ stage: \ ev_{irct} = \zeta_0 + \zeta_1 \Delta w_{rc} + \zeta_2 p^*_{rc} + \zeta_3 n_{r\tau t} + \gamma_C + v_{irct} \\ &Second\ stage: \ pr_{irc,t+1} = \varphi_0 + \varphi_1 \widehat{ev_{irct}} + \varphi_2 p^*_{rc} + \varphi_3 n_{r\tau t} + \gamma_C + \epsilon_{irct} \end{aligned} \tag{3.11}$$

where γ_C is a set of county dummies. Identification of ζ_1 is off differences in wage increases associated with promotion to different ranks within the same county.

From the first stage regressions, a number of predictions can be tested. We expect that ζ_1 will be a positive and significant coefficient (Prediction 2). This would tell us that promotion contests with higher wage gaps elicit higher levels of effort. It would also tell us that wages are an informative instrument. In order to test Prediction 3, I include the promotion rates (calculated at the county level for each rank) as the proportion of teachers in a county that are promoted at that rank within a 'reasonable' number of years [6], and split this into quintiles, omitting the third quintile. Table A2 in the appendix provides the mean promotion rates of the five quintiles. The average promotion rate in the quintiles is 11%, 37%, 52%, 68%, and 89% in the first to fifth quintile, respectively. I leave the third quintile as the omitted category, so I expect all the coefficients on these promotion rate quintile dummies to be negative, and I expect the coefficient for the first quintile to be smaller than that of the second quintile, and the coefficient for the fifth quintile to be smaller than that of the fourth quintile, as these are the quintiles with promotion rates furthest from 1/2. The number of teachers in the competition [7] is also interacted with

[6] I first calculate the number of years it took for each teacher to be promoted from one rank to another (I have each teacher's entire promotion history). Then I calculate the year for which, for each rank and education level in a county, half of the teachers were promoted within that time. This is the 'reasonable' number of years. Then for each rank and county I calculate the proportion of teachers that were promoted within that number of years. I only include promotions before 2003, and so this is a historical promotion rate in the county.

[7] The number of teachers in each rank is adjusted by the undersampling rate at the school. This is

the proportion promoted quintile dummies. I expect the coefficients on the interactions for the first, second, fourth and fifth quintiles to be negative, as an increase in the number of teachers competing is predicted to lower effort incentives when the promotion rate is further from 1/2 (Prediction 4). I also include the average age and the proportion of female teachers in a rank in a township, as well as the interaction between these two variables to test Prediction 5. I expect these coefficients to be negative as well.

From the second stage, I can test Prediction 1; the main prediction of the model. The coefficient φ_1 will tell us whether teachers are working hard for promotions, and thus, whether the promotion structure in Gansu does motivate teacher effort. I expect it also to be positive and significant. Note that this estimation strategy takes advantage of the fact that we observe teachers at multiple ranks within each county, and identification of the effort variable is driven by wage increases at different ranks within each county.

Measurement error could be a potential source of bias. We would be most concerned with measurement error in the evaluation scores, as measurement error here could result in attenuation bias. Evaluations are likely an imperfect measure of teacher effort, but they do incorporate several dimensions of teacher effort, as they are based on student test scores, teacher attitude, attendance, and preparation. Chapter 4 provides further evidence on the validity of evaluation scores as a measure for effort.

We would also be concerned with measurement error on the number of teachers. As mentioned previously, not all teachers were interviewed in all schools. As such, I have also tried restricting the sample to schools with a sampling rate of 25% and above and 35% and above to check that the results are robust to sampling rates and the adjustment described above.

Since wage increases and promotion rates are set at the county level, but the competition is between teachers at the township level (n), we may worry about differences in unobserved average skills between townships. Since wage gaps and promotion rates are

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 to calculate the sampling rate and the undersampling rate. The number of teachers in each rank was then adjusted as follows: number interviewed * (1 + undersampling 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.

common across townships in a county, the change in wages and promotion rate quintile variables will not be biased. However, if higher skilled (on average) townships tend to have more teachers, this would bias these coefficients. This could occur if, for example, there are some townships in a county that are more urbanised. Migration to urban townships within the county (or simply urban townships having larger populations) may also mean that there are more teachers competing. As a result, I also identify the township level variables (number of teachers competing) using a township fixed effects regression that will control for unobserved township factors.

Finally, we may also be concerned if teachers are moving to counties that have higher promotion rates or wage rates. If, for example, the hardest working teachers moved to counties with the highest wage rates, this would bias coefficient estimates as well and would invalidate the instrument. However, it is not possible for teachers to move counties based on wage rates. Teachers are rarely transferred across counties and would need a legitimate reason to do so. Only 20 teachers (0.85%) in the sample have switched counties between 2003 and 2007. However, I will also test the robustness of the results to restricting the regression to the sample of teachers who have not switched counties during this time.

The next section will present the results from estimating the empirical models above.

3.3 Results

This section presents the findings from estimation of the empirical model. First, I will discuss the evaluation scores in greater detail, and then move on to the main regressions, which are the promotion regressions estimating equation (3.11).

3.3.1 Effort

I begin with a regression of the evaluation scores on the change in log wages, the first stage of the instrumental variables estimation of the promotion equation. Table 3.1 contains the results from the first stage regressions of a Probit regression with an endogenous variable [8]. The results in the first column include only the change in log wages and the average age

[8] These results are identical for 2SLS and Probit with an endogenous variable, since the first stage is just a linear regression in both cases.

and proportion female variables, and the second column adds the proportion of teachers promoted variable split into quintiles. The third column also adds to this the number of teachers competing, interacted with the proportion promoted quintiles. The fourth column is the same specification as column (3) but with township fixed effects instead, to control for unobserved heterogeneity between townships in the same county. We can see that a larger wage gap elicits higher evaluation scores. A 1% higher wage gap increases the evaluation score by 0.00359 on average (taking the third column). This effect is significant at the 5% level. It is also an important effect. For example, from Primary 2 to Primary 1, the wage increase is 32.4% and this would induce an increase in evaluation scores of 0.12. The coefficients on the proportion promoted quintiles are negative as expected in column (2), but the magnitude of the coefficients is not what is expected, as the coefficient on the first quintile is not smaller than that of the second quintile and the coefficient on the fifth quintile is not smaller than that of the fourth quintile. All the coefficients, although they have the expected sign in column (2), are quite imprecisely estimated. There is evidence that Prediction 3 could hold, but it is weak. The coefficients on the number of teachers in a contest are also imprecisely estimated, and those for the number of teachers interacted with the proportion promoted quintiles 1, and 5 have the predicted sign (column (4)). The evidence for Prediction 4 is also weak. Older teachers do not seem to have lower incentives, but the higher the proportion of female teachers in a rank in a township, the lower are incentives (Prediction 5). However, again these coefficients are imprecisely estimated.

The instrument, the change in the log of average wages from one rank to another in a county, performs quite well. In all county fixed effects specifications the coefficient on change in log wages is positive as expected, and is statistically significant at the 5% level. The F-statistic for the significance of the instrument (a test that the coefficient on the instrument is zero) reveals that we do not have a weak instrument. A widely used rule of thumb suggested by Staiger and Stock (1997) suggests that an F-statistic of less than 10 indicates weak instruments. In all four first stage regressions, the F statistic is over 200, so there is no reason to be concerned about weak instruments.

I also conduct three falsification tests to further investigate the validity of the instru-

Table 3.1: IV and IV Probit first stage results

	Outcome: evaluation score of teacher i, in year t			
	(1)	(2)	(3)	(4)
Change in log wages	0.435***	0.422***	0.359**	0.445*
	(0.150)	(0.134)	(0.176)	(0.262)
Average age in rank in township	0.002	0.002	0.002	0.003
	(0.003)	(0.003)	(0.003)	(0.003)
Proportion female in rank in township	-0.084	-0.080	-0.091	-0.069
	(0.259)	(0.266)	(0.267)	(0.237)
Average age * proportion female	0.001	0.001	0.002	0.003
	(0.008)	(0.008)	(0.008)	(0.008)
Proportion promoted - quintile 1		-0.014	0.132	0.108
		(0.077)	(0.182)	(0.186)
Proportion promoted - quintile 2		-0.021	-0.107	-0.052
		(0.066)	(0.096)	(0.103)
Proportion promoted - quintile 4		-0.088*	-0.059	-0.076
		(0.052)	(0.146)	(0.127)
Proportion promoted - quintile 5		-0.019	0.017	0.041
		(0.042)	(0.149)	(0.129)
Number of teachers in rank in township (N)			0.002	0.021
			(0.033)	(0.028)
P* quintile 1 * N			-0.050	-0.055
			(0.058)	(0.062)
P* quintile 2 * N			0.022	0.008
			(0.022)	(0.022)
P* quintile 4 * N			-0.008	0.002
			(0.034)	(0.032)
P* quintile 5 * N			-0.011	-0.015
			(0.046)	(0.038)
Constant	2.394***	2.437***	2.437***	2.299***
	(0.117)	(0.138)	(0.158)	(0.164)
Number of observations	5,111	5,111	5,111	5,114
F Statistic	317.93	209.15	565.43	439.85
Fixed effect	County	County	County	Township

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Regressions include time dummies and standard errors clustered at the level of the fixed effect.

ment. First, we may be worried that wages have a direct effect on the probability of promotion. This could be the case if, for example, higher wage gaps are associated with very high or very low promotion rates and so by construction a teacher may have a higher or lower probability of promotion with a high wage gap. We can test this by looking at teacher interns at primary and middle schools. Interns are automatically promoted to the lowest rank level after one year. As such, wages should not affect their effort or their promotions. When I include the wage gap for interns directly in the promotion regression, the wage gap has no effect. The results are contained in Table 3.2, Column (1). As is evident here, evaluation scores do not affect the probability of promotion either.

The second falsification test assigns each teacher the wage gap at the rank below, instead of the wage gap he or she currently faces. A valid instrument would mean that a wage gap that is no longer relevant should have no effect on effort. Column (2) of Table 3.2 reports the result. The coefficient is small and insignificant; the 'false' wage gap does not predict effort [9].

Thirdly, we may be worried that there are unobserved factors at the rank level within a county that determine the differences in both wage gaps and evaluation scores at different levels. This could happen if, for example, the wage gap from the lowest level to the next is low and teachers in that rank in a school tend to receive low evaluation scores. If this conjecture is true and these patterns are similar across counties, then randomly assigning wage gaps at the same level from a different county, should also have an effect on evaluation scores. Column (3) of Table 3.2 shows that there is no such effect; the coefficient is small and not significantly different from zero. These three tests indicate a valid instrument.

3.3.2 Promotions

I now discuss the promotion regressions. Estimation is conducted as follows: I first estimate equation (3.11) using Probit, which accounts for the binary nature of the outcome variable. I then estimate equation (3.11) using Probit with an instrumental variable,

[9] I do not include the wage gap for the rank above as another falsification test because future wage increases could also motivate current teacher effort.

Table 3.2: IV Falsification Tests

Outcome	(1) Promotion	(2) Evaluation	(3) Evaluation
Evaluation score	-0.002		
	(0.002)		
Change in log wages (interns)	-0.014		
	(0.024)		
Change in log wages (of rank below)		0.080	
		(0.070)	
Change in log wages (randomly assigned counties)			-0.141
			(0.121)
Average age in rank in township	-0.001	0.001	0.001
	(0.001)	(0.003)	(0.003)
Proportion female in rank in township	-0.041	-0.249	-0.075
	(0.051)	(0.255)	(0.274)
Average age * proportion female	0.000	0.008	0.004
	(0.001)	(0.008)	(0.009)
Proportion promoted - quintile 1		0.184*	0.257*
		(0.101)	(0.155)
Proportion promoted - quintile 2		-0.116	-0.009
		(0.072)	(0.080)
Proportion promoted - quintile 4		0.030	0.062
		(0.122)	(0.144)
Proportion promoted - quintile 5		0.030	0.103
		(0.139)	(0.157)
Number of teachers (N)		-0.014	0.013
		(0.033)	(0.033)
P* quintile 1 * N		-0.049	-0.073
		(0.040)	(0.050)
P* quintile 2 * N		0.030	0.006
		(0.024)	(0.025)
P* quintile 4 * N		-0.022	-0.033
		(0.030)	(0.036)
P* quintile 5 * N		-0.006	-0.026
		(0.048)	(0.050)
Constant	0.053	2.502***	2.438***
	(0.049)	(0.164)	(0.171)
Number of observations	504	5,394	5,196

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.

which corrects both for the binary outcome and for the endogeneity of the evaluation score [10]. This improves on the linear treatment of the 2SLS method as it incorporates the nonlinearity of the second stage equation while generating the correct standard errors for the coefficients, and is the preferred model. Furthermore, the models are estimated as before, beginning with a basic specification that includes only the evaluation scores (or instrumented evaluation scores) and the average age of teachers in a county in a rank, proportion of female teachers in the county in that rank, and the interaction of these two variables. Next, I include the proportion of teachers promoted quintiles, and then finally I add the variables for the number of teachers competing in a township.

Table 3.3 contains the marginal effects of the evaluation scores from the Probit with IV models. This is the full model, and Column (3) of Table 3.3 is the preferred specification. We can see that the marginal effect for the evaluation score (instrumented with change in log wages) is positive and significant. An increase in the evaluation score of one unit (for example, from 'pass' to 'good' or from 'good' to 'excellent') increases the probability of promotion by almost 30% and this is significant at the 1% level of significance. This result indicates that the effect of effort on promotions is positive and significant, and so effort is being exerted by teachers in order to earn promotions, and we can affirm Prediction 1. An increase in the probability of promotion of 30% seems high, but because there are only four possible values of the evaluation score (and more realistically three, since very few people fail), increasing promotion probability by about one third makes sense.

We can compare this to the Probit marginal effect without the IV, which is contained in Table 3.4. The effect is also positive and significant, but a one unit increase in the evaluation score leads to an increase in the probability of promotion of only 1% here. This specification fails to account for the endogeneity of the evaluation score. Hypothesis tests on the correlation between the error terms in the first and second stage regressions (equivalent to testing whether the endogenous variable is indeed endogenous) reveal correlations that are statistically significant, and so the evaluation scores are indeed endogenous. These are reported in Table 3.3 in the row labelled 'Rho'. The hypothesis is

[10] OLS and 2SLS specifications were also run. These results are contained in Appendix Tables 3.5 and 3.6, respectively. Coefficient estimates from the Probit and Probit with IV specifications are also included in the Appendix, in Tables 3.7 and 3.8, respectively.

Table 3.3: IV Probit marginal effects

Outcome: promotion of teacher i in year t			
	(1)	(2)	(3)
Marginal effect - evaluation score			
(instrumented with change in log wages)	0.152	0.193**	0.273***
	(0.118)	(0.091)	(0.072)
Rho	-0.573	-0.681**	-0.851**
	(0.316)	(0.217)	(0.145)
Sigma	0.725***	0.725***	0.725***
	(0.008)	(0.008)	(0.008)
Promotion rate quintiles included	No	Yes	Yes
Number of teachers included	No	No	Yes
Number of observations	5,111	5,111	5,111

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.

that the correlation is equal to zero. We reject this in columns (2) and (3). The sign of Rho is negative, which is consistent with the difference in the evaluation score coefficients between the naive Probit models and in the Probit with IV models (that the coefficient is larger for Probit with IV) [11]. This makes sense, as high ability people are likely to have lower effort, but a higher probability of promotion.

Table 3.4: Probit marginal effects

Outcome: promotion of teacher i in time t			
	(1)	(2)	(3)
Marginal effect - evaluation score	0.011***	0.012***	0.012***
	(0.004)	(0.004)	(0.004)
Promotion rate quintiles included	No	Yes	Yes
Number of teachers included	No	No	Yes
Number of observations	5,111	5,111	5,111

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.

[11] This same pattern also holds in comparing the OLS and IV results, see Appendix.

3.3.3 Robustness Checks

Because the number of teachers in a rank in a township is a variable that has been adjusted, as a check, I restrict the sample first to schools that have a sampling rate of 25% and over and then to schools that have a sampling rate of 35% and over. Table 3.9 in the Appendix contains the first stage results from the IV and Probit with IV regressions for both the 25% and above and 35% and above samples. Tables 3.10 and 3.11 contain the coefficient estimates from the OLS and Probit regressions, as well as the second stage coefficients from the IV and Probit with IV regressions for the sample of 25% and above and 35% and above, respectively. The coefficient estimates change slightly, but remain quite consistent. In particular, the interpretation of the results on Predictions 1 and 2 do not change. The coefficients on the average age of teachers in the rank is also consistent with the previous results. The proportion promoted quintile dummies and number of teachers competing variables also change slightly, but the signs are generally consistent, and they were very imprecisely estimated to begin with. As a result, the results appear to be robust to the adjustment of this variable.

We may be worried that the 20 teachers that have switched counties over the sample period, may be driving the results. As a check, I also restrict the sample to drop these teachers. The results are contained in Appendix Tables 3.12 and 3.13. Here again, the results are almost identical to the main results. The teachers that have moved counties are not driving the results.

3.4 Summary

This chapter has presented a theoretical model of promotions as incentives for teacher effort. The aim was to assess whether the promotion system for teachers in Gansu province, China, motivates teacher effort. The first order condition for effort was derived and then differentiated in order to understand the effects of the promotion system on the average effort levels of teachers. This led to a function that resembled a Beta density, and the properties of Beta densities were used to understand the incentive effects. The main behavioural equation of the model was then used as a basis for the empirical specification to test the model's predictions. This behavioural equation states that effort is exerted to earn promotions; promotions motivate effort.

Panel data from the Gansu Survey of Children and Families (GSCF) were used to test five predictions. Two main types of models are estimated: a Probit model that accounted for the binary nature of the outcome, and a Probit with IV model that accounted for the endogeneity of the evaluation scores (which proxied for effort) as well as corrected for the binary outcome. I find that the predictions of the model are largely corroborated by the data. Effort is indeed exerted in order to earn promotions (Prediction 1). The first stage of the IV regressions showed that promotion contests with higher wage gaps between rank levels lead to higher average effort (Prediction 2). In addition, contests in which the promotion rate is closer to 1/2 have higher effort incentives (Prediction 3). However, the coefficients are imprecisely estimated. The number of teachers in a contest does matter (Prediction 4), but again the evidence is weak. This is perhaps due to the fact that the promotion system in Gansu is a combination of a tournament and a standard. The evidence for the prediction that contests with older teachers and a higher proportion of female teachers would have lower incentives is weak (Prediction 5).

The next chapter will extend the theoretical and empirical models in this chapter and will focus on the second major question of the model: how does individual teacher effort vary with the interaction between teacher characteristics and the parameters of the promotion contest? What can this tell us about how best to design promotion-based incentive systems in the public sector?

3.5 Appendix 1 Chapter 3 - Additional Tables

Table 3.5: OLS coefficient estimates

Outcome: promotion of teacher i in year t			
	(1)	(2)	(3)
	Basic	+ p*	+ n
Evaluation score	0.011**	0.012**	0.012**
	(0.005)	(0.004)	(0.004)
Average age in rank in township	-0.005***	-0.004**	-0.004**
	(0.002)	(0.002)	(0.002)
Proportion female in rank in township	0.146	0.126	0.134
	(0.092)	(0.100)	(0.096)
Average age * proportion female	-0.004	-0.004	-0.003
	(0.003)	(0.003)	(0.003)
Proportion promoted - quintile 1		-0.008	-0.002
		(0.026)	(0.054)
Proportion promoted - quintile 2		0.017	0.098
		(0.018)	(0.077)
Proportion promoted - quintile 4		0.044	0.117
		(0.031)	(0.079)
Proportion promoted - quintile 5		0.012	0.070
		(0.016)	(0.052)
Log Number of teachers in rank in township			-0.003
			(0.005)
P* quintile 1 * N			-0.003
			(0.013)
P* quintile 2 * N			-0.022
			(0.018)
P* quintile 4 * N			-0.021
			(0.016)
P* quintile 5 * N			-0.019
			(0.015)
Observations	5703	5703	5703
R^2	0.023	0.028	0.034

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

All regressions include county fixed effects and standard errors clustered at the county level.

Table 3.6: IV coefficient estimates

Outcome: promotion of teacher i in year t	(1) Basic	(2) + p*	(3) + n
Evaluation score (instrumented	0.231**	0.286**	0.424**
with change in log wages)	(0.109)	(0.124)	(0.186)
Average age in rank in township	-0.004***	-0.004***	-0.004***
	(0.001)	(0.001)	(0.001)
Proportion female in rank in township	0.173**	0.153*	0.171*
	(0.073)	(0.080)	(0.099)
Average age * proportion female	-0.005**	-0.004*	-0.004
	(0.002)	(0.002)	(0.003)
Proportion promoted - quintile 1		-0.022	-0.149**
		(0.015)	(0.066)
Proportion promoted - quintile 2		0.009	0.092*
		(0.015)	(0.053)
Proportion promoted - quintile 4		0.039**	0.032
		(0.019)	(0.051)
Proportion promoted - quintile 5		-0.004	-0.008
		(0.015)	(0.049)
Log Number of teachers in rank in township			-0.019**
			(0.009)
P* quintile 1 * N			0.041*
			(0.021)
P* quintile 2 * N			-0.020
			(0.014)
P* quintile 4 * N			0.006
			(0.014)
P* quintile 5 * N			-0.001
			(0.015)
Observations	5114	5114	5114

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

All regressions include county fixed effects and standard errors clustered at the county level.

Table 3.7: Probit coefficient estimates

Outcome: promotion of teacher i in year t			
	(1)	(2)	(3)
	Basic	+ p*	+ n
Evaluation score	0.100***	0.103***	0.112***
	(0.039S)	(0.037)	(0.035)
Average age in rank in township	-0.043***	-0.042***	-0.034***
	(0.015)	(0.015)	(0.013)
Proportion female in rank in township	0.689	0.688	0.750
	(0.804)	(0.843)	(0.775)
Average age * proportion female	-0.020	-0.022	-0.018
	(0.026)	(0.027)	(0.026)
Proportion promoted - quintile 1		-0.059	0.022
		(0.217)	(0.446)
Proportion promoted - quintile 2		0.206	0.610
		(0.162)	(0.390)
Proportion promoted - quintile 4		0.330	0.639
		(0.207)	(0.430)
Proportion promoted - quintile 5		0.096	0.514
		(0.145)	(0.354)
Log Number of teachers in rank in township			-0.037
			(0.059)
P* quintile 1 * N			-0.031
			(0.125)
P* quintile 2 * N			-0.117
			(0.096)
P* quintile 4 * N			-0.098
			(0.113)
P* quintile 5 * N			-0.149
			(0.105)
Observations	5703	5703	5703

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

All regressions include county fixed effects and standard errors clustered at the county level.

Table 3.8: IV Probit coefficient estimates

Outcome: promotion of teacher i in year t			
	(1)	(2)	(3)
	Basic	+ p*	+ n
Evaluation score (instrumented	0.873**	1.014***	1.240***
with change in log wages)	(0.428)	(0.291)	(0.184)
Average age in rank in township	-0.028	-0.026	-0.015
	(0.017)	(0.016)	(0.010)
Proportion female in rank in township	0.872	0.728	0.612
	(0.799)	(0.775)	(0.670)
Average age * proportion female	-0.027	-0.023	-0.014
	(0.026)	(0.025)	(0.021)
Proportion promoted - quintile 1		-0.151	-0.491
		(0.211)	(0.348)
Proportion promoted - quintile 2		0.073	0.297
		(0.146)	(0.235)
Proportion promoted - quintile 4		0.156	0.143
		(0.160)	(0.311)
Proportion promoted - quintile 5		-0.039	-0.015
		(0.126)	(0.367)
Log Number of teachers in rank in township			-0.080
			(0.059)
P* quintile 1 * N			0.150
			(0.105)
P* quintile 2 * N			-0.049
			(0.061)
P* quintile 4 * N			0.025
			(0.076)
P* quintile 5 * N			0.001
			(0.112)
Observations	5114	5114	5114

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

All regressions include county fixed effects and standard errors clustered at the county level.

Table 3.9: First stage regressions, schools with sampling rates of teachers 25% and 35% and above

	Outcome: evaluation score of teacher i in year t							
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
		25%+ sampling rate				35%+ sampling rate		
	Basic	+ p*	+ n	Town	Basic	+ p*	+ n	Town
Change in log wages	0.383**	0.365***	0.240	0.253	0.394**	0.369***	0.257	0.279
	(0.157)	(0.137)	(0.182)	(0.276)	(0.162)	(0.139)	(0.189)	(0.282)
Average age in rank in township	0.004	0.003	0.004	0.006*	0.003	0.003	0.003	0.006
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
Proportion female in rank in township	-0.084	-0.057	-0.067	0.003	-0.060	-0.034	-0.049	0.061
	(0.255)	(0.271)	(0.276)	(0.257)	(0.264)	(0.285)	(0.287)	(0.284)
Average age * proportion female	0.002	0.001	0.001	0.001	0.001	0.000	0.000	-0.001
	(0.007)	(0.008)	(0.009)	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)
Proportion promoted - quintile 1		0.014	0.307	0.322		0.013	0.266	0.254
		(0.086)	(0.189)	(0.202)		(0.090)	(0.211)	(0.234)
Proportion promoted - quintile 2		-0.003	-0.002	-0.020		-0.007	-0.011	-0.039
		(0.066)	(0.115)	(0.103)		(0.067)	(0.114)	(0.107)
Proportion promoted - quintile 4		-0.067	-0.002	-0.087		-0.078	-0.085	-0.158
		(0.058)	(0.165)	(0.143)		(0.058)	(0.178)	(0.160)
Proportion promoted - quintile 5		-0.011	-0.019	-0.027		-0.010	-0.037	-0.042
		(0.054)	(0.166)	(0.139)		(0.056)	(0.163)	(0.144)
Log Number of teachers in rank in township			0.017	0.026			0.015	0.020
			(0.038)	(0.030)			(0.038)	(0.032)
P* quintile 1 * N			-0.094	-0.094			-0.080	-0.072
			(0.058)	(0.067)			(0.063)	(0.076)
P* quintile 2 * N			-0.002	0.004			-0.000	0.008
			(0.026)	(0.024)			(0.025)	(0.024)
P* quintile 4 * N			-0.017	0.005			0.002	0.022
			(0.040)	(0.040)			(0.045)	(0.044)
P* quintile 5 * N			0.008	0.008			0.013	0.012
			(0.048)	(0.041)			(0.048)	(0.042)
Observations	4657	4657	4657	4551	4466	4466	4466	4327
Fixed effect	County	County	County	Township	County	County	County	Township

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Standard errors clustered at the level of the fixed effect.

Table 3.10: OLS, Probit, IV and IV Probit (second stage) - schools with sampling rate of teachers 25% and above

	Outcome: promotion of teacher i in year t											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
		OLS			IV			Probit			IV Probit	
	Basic	+ p*	+ n	Basic	+ p*	+ n	Basic	+ p*	+ n	Basic	+ p*	+ n
Evaluation score	0.008	0.009*	0.009*	0.269**	0.341**	0.634	0.080*	0.081**	0.091**	0.880*	1.064***	1.316***
	(0.005)	(0.005)	(0.004)	(0.137)	(0.162)	(0.392)	(0.042)	(0.041)	(0.039)	(0.530)	(0.317)	(0.164)
Average age	-0.005***	-0.005**	-0.004**	-0.005***	-0.005***	-0.005**	-0.048***	-0.046***	-0.036***	-0.032*	-0.029*	-0.015
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.015)	(0.015)	(0.014)	(0.018)	(0.017)	(0.012)
Prop. female	0.146	0.124	0.133	0.168**	0.136	0.163	0.771	0.791	0.811	0.920	0.699	0.415
	(0.098)	(0.108)	(0.103)	(0.082)	(0.092)	(0.139)	(0.865)	(0.911)	(0.804)	(0.862)	(0.808)	(0.627)
Avg age*female	-0.004	-0.004	-0.003	-0.005**	-0.004	-0.004	-0.025	-0.027	-0.021	-0.032	-0.025	-0.011
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.004)	(0.029)	(0.030)	(0.027)	(0.028)	(0.026)	(0.020)
P* - quintile 1		-0.009	0.038		-0.033*	-0.270		-0.058	0.426		-0.178	-0.540
		(0.028)	(0.069)		(0.018)	(0.170)		(0.230)	(0.470)		(0.223)	(0.350)
P* - quintile 2		0.018	0.120		0.006	0.068		0.231	0.758**		0.071	0.130
		(0.019)	(0.077)		(0.017)	(0.072)		(0.166)	(0.383)		(0.147)	(0.215)
P* - quintile 4		0.046	0.161*		0.039*	0.041		0.353*	0.865*		0.153	0.048
		(0.032)	(0.092)		(0.021)	(0.078)		(0.206)	(0.473)		(0.146)	(0.329)
P* - quintile 5		0.014	0.093		-0.002	0.035		0.107	0.678*		-0.027	0.089
		(0.017)	(0.058)		(0.017)	(0.069)		(0.152)	(0.382)		(0.134)	(0.347)
Log (N)			-0.001			-0.027*			-0.010			-0.073
			(0.005)			(0.016)			(0.060)			(0.064)
P* quintile 1 * N			-0.017			0.072			-0.174			0.132
			(0.016)			(0.051)			(0.134)			(0.125)
P* quintile 2 * N			-0.028			-0.015			-0.159*			-0.023
			(0.017)			(0.019)			(0.093)			(0.053)
P* quintile 4 * N			-0.032			0.005			-0.161			0.022
			(0.019)			(0.022)			(0.123)			(0.080)
P* quintile 5 * N			-0.026			-0.016			-0.197*			-0.047
			(0.016)			(0.020)			(0.111)			(0.100)
Observations	5131	5131	5131	4657	4657	4657	5131	5131	5131	4657	4657	4657

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

All regressions include county fixed effects and standard errors clustered at the county level.

Table 3.11: OLS, Probit, IV and IV Probit (second stage) - schools with sampling rate of teachers 35% and above

	Outcome: promotion of teacher i in year t											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
		OLS			IV			Probit			IV Probit	
	Basic	+ p*	+ n	Basic	+ p*	+ n	Basic	+ p*	+ n	Basic	+ p*	+ n
Evaluation score	0.009*	0.010*	0.010**	0.221*	0.290*	0.528	0.086**	0.090**	0.101***	0.724	0.969**	1.289***
	(0.005)	(0.005)	(0.004)	(0.125)	(0.149)	(0.323)	(0.042)	(0.041)	(0.039)	(0.654)	(0.412)	(0.216)
Average age	-0.005***	-0.005**	-0.004**	-0.005***	-0.006***	-0.005***	-0.049***	-0.047***	-0.036***	-0.036*	-0.033*	-0.017
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.002)	(0.015)	(0.015)	(0.014)	(0.019)	(0.019)	(0.014)
Prop. female	0.142	0.123	0.131	0.158**	0.128	0.148	0.784	0.842	0.836	0.996	0.791	0.448
	(0.105)	(0.113)	(0.109)	(0.079)	(0.088)	(0.125)	(0.900)	(0.932)	(0.815)	(0.916)	(0.874)	(0.684)
Avg age*female	-0.004	-0.004	-0.003	-0.005**	-0.004	-0.003	-0.026	-0.030	-0.021	-0.035	-0.029	-0.011
	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.004)	(0.029)	(0.031)	(0.027)	(0.030)	(0.029)	(0.022)
P* - quintile 1		-0.006	0.041		-0.026	-0.213		-0.041	0.419		-0.163	-0.513
		(0.029)	(0.066)		(0.017)	(0.135)		(0.243)	(0.420)		(0.240)	(0.374)
P* - quintile 2		0.025	0.137*		0.017	0.090		0.290*	0.809**		0.141	0.178
		(0.019)	(0.076)		(0.016)	(0.064)		(0.168)	(0.344)		(0.156)	(0.218)
P* - quintile 4		0.047	0.160		0.043**	0.080		0.386*	0.846*		0.207	0.139
		(0.032)	(0.095)		(0.020)	(0.073)		(0.215)	(0.477)		(0.163)	(0.361)
P* - quintile 5		0.015	0.077		0.004	0.029		0.132	0.530		0.018	0.056
		(0.016)	(0.057)		(0.016)	(0.062)		(0.146)	(0.377)		(0.135)	(0.370)
Log (N)			-0.004			-0.027**			-0.046			-0.097
			(0.005)			(0.014)			(0.061)			(0.072)
P* quintile 1 * N			-0.017			0.055			-0.167			0.121
			(0.014)			(0.040)			(0.118)			(0.129)
P* quintile 2 * N			-0.030*			-0.017			-0.151*			-0.021
			(0.017)			(0.016)			(0.081)			(0.049)
P* quintile 4 * N			-0.031			-0.005			-0.142			0.011
			(0.020)			(0.018)			(0.129)			(0.088)
P* quintile 5 * N			-0.020			-0.012			-0.135			-0.027
			(0.016)			(0.018)			(0.111)			(0.105)
Observations	4885	4885	4885	4466	4466	4466	4885	4885	4885	4466	4466	4466

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

All regressions include county fixed effects and standard errors clustered at the county level.

Table 3.12: First stage regressions, teachers who have not switched counties

Outcome: evaluation score of teacher i in year t				
	(1)	(2)	(3)	(4)
	Basic	+ p*	+ n	Township
Change in log wages	0.440***	0.426***	0.358**	0.381
	(0.145)	(0.131)	(0.172)	(0.263)
Average age in rank in township	0.002	0.002	0.002	0.003
	(0.003)	(0.003)	(0.003)	(0.004)
Proportion female in rank in township	-0.089	-0.088	-0.096	-0.080
	(0.261)	(0.270)	(0.268)	(0.244)
Average age * proportion female	0.001	0.001	0.002	0.004
	(0.008)	(0.008)	(0.008)	(0.008)
Proportion promoted - quintile 1		-0.015	0.143	0.166
		(0.075)	(0.178)	(0.191)
Proportion promoted - quintile 2		-0.025	-0.116	-0.100
		(0.066)	(0.099)	(0.106)
Proportion promoted - quintile 4		-0.089*	-0.051	-0.139
		(0.052)	(0.146)	(0.133)
Proportion promoted - quintile 5		-0.019	0.035	-0.000
		(0.042)	(0.148)	(0.129)
Log Number of teachers in rank in township			0.004	0.015
			(0.033)	(0.028)
P* quintile 1 * N			-0.053	-0.058
			(0.057)	(0.063)
P* quintile 2 * N			0.024	0.020
			(0.022)	(0.024)
P* quintile 4 * N			-0.011	0.015
			(0.035)	(0.035)
P* quintile 5 * N			-0.017	-0.004
			(0.046)	(0.038)
Observations	5091	5091	5091	4891
Fixed effect	County	County	County	Township

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Standard errors clustered at level of fixed effect.

Table 3.13: OLS, Probit, IV and IV Probit (second stage) - teachers who have not switched counties

	Outcome: promotion of teacher i in year t											
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
		OLS			IV			Probit			IV Probit	
	Basic	+ p*	+ n	Basic	+ p*	+ n	Basic	+ p*	+ n	Basic	+ p*	+ n
Evaluation score	0.011**	0.012**	0.012**	0.233**	0.290**	0.437**	0.102***	0.105***	0.114***	0.876**	1.019***	1.253***
	(0.005)	(0.005)	(0.004)	(0.109)	(0.124)	(0.191)	(0.039)	(0.038)	(0.036)	(0.419)	(0.282)	(0.169)
Average age	-0.004**	-0.004**	-0.003**	-0.004***	-0.004***	-0.003**	-0.043***	-0.041***	-0.033**	-0.028	-0.026	-0.014
	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.015)	(0.015)	(0.013)	(0.017)	(0.016)	(0.010)
Prop. female	0.153	0.134	0.143	0.178**	0.159**	0.179*	0.723	0.733	0.805	0.890	0.747	0.556
	(0.093)	(0.101)	(0.096)	(0.074)	(0.080)	(0.101)	(0.805)	(0.848)	(0.778)	(0.797)	(0.774)	(0.650)
Avg age*female	-0.004	-0.004	-0.003	-0.005**	-0.004*	-0.004	-0.021	-0.023	-0.019	-0.028	-0.024	-0.013
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.003)	(0.026)	(0.028)	(0.026)	(0.026)	(0.025)	(0.021)
P* - quintile 1		-0.007	0.002		-0.022	-0.157**		-0.047	0.059		-0.149	-0.493
		(0.026)	(0.054)		(0.015)	(0.069)		(0.219)	(0.456)		(0.209)	(0.324)
P* - quintile 2		0.018	0.103		0.010	0.097*		0.217	0.647*		0.078	0.241
		(0.018)	(0.077)		(0.015)	(0.055)		(0.163)	(0.392)		(0.146)	(0.205)
P* - quintile 4		0.046	0.122		0.040**	0.032		0.344	0.678		0.161	0.034
		(0.031)	(0.079)		(0.019)	(0.052)		(0.210)	(0.436)		(0.160)	(0.314)
P* - quintile 5		0.013	0.074		-0.004	-0.015		0.106	0.551		-0.040	-0.044
		(0.016)	(0.052)		(0.015)	(0.051)		(0.147)	(0.359)		(0.124)	(0.345)
Log (N)			-0.003			-0.020**			-0.030			-0.083
			(0.005)			(0.010)			(0.061)			(0.058)
P* quintile 1 * N			-0.004			0.043**			-0.039			0.130
			(0.013)			(0.022)			(0.127)			(0.104)
P* quintile 2 * N			-0.023			-0.021			-0.125			-0.041
			(0.018)			(0.014)			(0.096)			(0.053)
P* quintile 4 * N			-0.022			0.007			-0.106			0.039
			(0.016)			(0.014)			(0.114)			(0.077)
P* quintile 5 * N			-0.020			0.002			-0.158			-0.003
			(0.015)			(0.015)			(0.106)			(0.103)
Observations	5679	5679	5679	5091	5091	5091	5679	5679	5679	5091	5091	5091

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

All regressions include county fixed effects and standard errors clustered at the county level.

3.6 Appendix 2 Chapter 3 - Full Derivation of Model

Full derivation of one-period model of promotion incentives.

The probability that teacher i beats teacher g is $pr(q_i > q_g) = pr(e_i + \epsilon_i + \mu > e_g + \epsilon_g + \mu) = pr(\epsilon_g < e_i - e_g + \epsilon_i) = F(e_i - e_g + \epsilon_i)$ [12]. A symmetric Bayesian Nash Equilibrium (BNE) is appropriate for identical teachers and optimal effort for all teachers is e^*. The probability that teacher i beats any other teacher is $F(e_i - e^* + \epsilon_i)$ and the probability that teacher i loses to any opponent is $1 - F(e_i - e^* + \epsilon_i)$. To finish j^{th} from the top out of n teachers, one must beat $n - j$ teachers and lose to $j - 1$ teachers. The probability of doing so for any given partition of competitors, conditional on i's luck ϵ_i is $F(e_i - e^* + \epsilon_i)^{n-j}[1 - F(e_i - e^* + \epsilon_i)]^{j-1}$. The number of ways to choose $n - j$ elements from $n - 1$ is $\binom{n-1}{n-j} = \frac{(n-1)!}{(n-j)!(j-1)!}$ so the probability of placing exactly j^{th} from the top conditional on ϵ_i is:

$$pr(place\ j^{th}\ out\ of\ n \mid \epsilon_i) = \binom{n-1}{n-j} F(e_i - e^* + \epsilon_i)^{n-j}[1 - F(e_i - e^* + \epsilon_i)]^{j-1} \tag{3.12}$$

The probability of promotion conditional on ϵ_i is the sum of the conditional probabilities of placing first through k^{th}. Integrating out ϵ_i gives the unconditional probability of promotion:

$$p(e, \underline{e}^*) = \sum_{j=1}^{k} \binom{n-1}{n-j} \int F(e_i - e^* + \epsilon_i)^{n-j}[1 - F(e_i - e^* + \epsilon_i)]^{j-1} f(\epsilon) d\epsilon \tag{3.13}$$

To solve for the MPE we need to differentiate with respect to e and substitute in the symmetric NE condition $e_i = e^*$:

$$\frac{dp}{de} = \sum_{j=1}^{k} \binom{n-1}{n-j} \int \left((n-j)F(\epsilon)^{n-j-1}(1 - F(\epsilon))^{j-1} - (j-1)F(\epsilon)^{n-j}(1 - F(\epsilon))^{j-2}\right) f(\epsilon)^2 d\epsilon \tag{3.14}$$

[12] The common error term cancels out.

where the subscript i has been dropped for notational convenience. Since $(m-1)\binom{n-1}{n-m} = (n-(m-1))\binom{n-1}{n-(m-1)}$ and since the second expression is zero if $j = 1$, this becomes:

$$\begin{aligned}\frac{dp}{de} &= \sum_{j=1}^{k}(n-j)\binom{n-1}{n-j}\int F(\epsilon)^{n-j-1}(1-F(\epsilon))^{j-1}f(\epsilon)^2 d\epsilon \\ &- \sum_{m=2}^{k}(n-(m-1))\binom{n-1}{n-(m-1)}\int F(\epsilon)^{n-m}(1-F(\epsilon))^{m-2}f(\epsilon)^2 d\epsilon\end{aligned} \tag{3.15}$$

Letting $j = m-1$, all the terms in the second sum cancel terms in the first sum, leaving only the term $j = k$:

$$\frac{dp}{de} = (n-k)\binom{n-1}{n-k}\int F(\epsilon)^{n-k-1}(1-F(\epsilon))^{k-1}f(\epsilon)^2 d\epsilon \tag{3.16}$$

A teacher is trying to beat the $n-k-1^{st}$ order statistic (the random variable that is the $n-k-1^{st}$ score (from the bottom) among competitors). Denote this order statistic ς. Subtracting off optimal effort of competitors gives the luck ϵ required to win, $\lambda = \varsigma - e^*$. With a symmetric NE, effort terms cancel out and λ is also an order statistic (the $n-k-1^{st}$ (from the bottom) among the $n-1$ competitors). Let this random variable have CDF $G(\lambda)$ and PDF $g(\lambda)$. For any given luck ϵ, the probability of promotion equals:

$$pr(promoted \mid \epsilon) = p(e+\epsilon > \lambda \mid \epsilon) = p(e+\epsilon > e^* + \lambda \mid \epsilon) = G(e - e^* + \epsilon) \tag{3.17}$$

Integrating over the luck distribution provides the unconditional probability of promotion, $p = \int G(e-e^*+\epsilon)f(\epsilon)d\epsilon$ where the integrand except for $f(\epsilon)$ is an order statistic density.

We can simplify (3.16) by making use of change of variables $\Psi \equiv F(\epsilon)$. Since $f d\epsilon = dF = d\Psi$, and $\epsilon = F^{-1}(\Psi)$ we can write:

$$\frac{dp}{de} = \int (n-k)\binom{n-1}{n-k}\Psi^{n-k-1}(1-\Psi)^{k-1}f(F^{-1}(\Psi))d\Psi \tag{3.18}$$

This integrand, except for $f(F^{-1}(\Psi))$, has the form of a Beta density function $\beta(k, n-k)$. Thus, Ψ has the effect of a Beta-distributed random variable and,

$$\frac{dp}{de} = E_{\Psi} f(F^{-1}(\Psi)) \tag{3.19}$$

Chapter 4

What Influences Effort?

4.1 An Extended Model of Promotions as Incentives

This chapter will focus on the second question of this thesis: what influences effort? It will examine how individual teacher effort varies with teacher characteristics, and with the parameters of the promotion system. As in Chapter 3, the theoretical model is based primarily on Gibbs (1995) and Gibbs (1989). First, the one-period model will be extended to include teachers that differ in their ability. Subsequently, an addition to the model that develops predictions on teaching effort over time will be presented. The models will be used to generate testable implications to take to the data. An effort equation focusing on individual teachers will be estimated in order to see how various design features of the promotion system interact with teacher characteristics to influence teacher effort.

4.1.1 One-Period Model of Promotions

The basics of the one-period model are as before; a school wishes to motivate teachers' effort by offering rewards from promotion. A school hires teachers into the lowest rank and promotions are then offered sequentially, based on performance. The wage increases and number of levels in the heirarchy are again assumed to be exogenously given. In this chapter, teachers differ in skill, s. Principals do not know individual teachers' s. Teachers know their skill but not that of others. All teachers have skill drawn from the same distribution $B(s)$ with density $b(s)$. Teachers and principals know that skill is drawn from this distribution. Skill is assumed to have a symmetric distribution with $E(s) = 0$ since it is only differences from mean skill that matter. Teachers contribute effort, e, to their job (this includes diligence, care etc., not just hours), and effort is costly and gives them disutility $C(e)$, which is increasing in e at an increasing rate so that C' and $C'' > 0$. Performance is now measured by $q_i = s_i + e_i + v_i$, where q_i is output, e_i is effort and $v_i = \epsilon_i + \mu$ is an error term, which comprises individual error (or 'luck') and a common error. I assume that v_i has a symmetric distribution around a mode of 0 and $E(v_i) = E(\epsilon_i) = E(\mu) = 0$. Let $F(\epsilon)$ be the CDF and $f(\epsilon)$ be the PDF of the error term. I further assume that $d^2q/de^2 = 0$ and $d^2q/dsde = 0$ so there are no scale economies in effort and no productive interactions between effort and skill.

There are n teachers competing for k promotion slots. The k teachers with the highest evaluation scores win the promotions. Each teacher's probability of promotion depends on his/her effort, skill and the efforts of others. A teacher's probability of promotion can now be written as $p = p(s, e, \underline{e})$ where $\underline{e}$ is a vector of other teachers' efforts. The school offers a benefit/reward from promotion, ΔEV as before, which is the change in expected lifetime utility gained from winning a promotion relative to not being promoted. Again, ΔEV also includes the option value of possible future promotions.

The maximisation problem and First Order Condition (FOC) are as before in Chapter 3. If n teachers are competing for a fixed number, k, promotion slots, then this is a tournament structure. To win a promotion, a teacher must place k^{th} or higher, or must outperform at least $n - k$ teachers in pair-wise comparisons of performance. A teacher treats the effort of a competitor as a function of the random variable that is the competitor's skill, and treats all co-workers as ex-ante identical. However, a symmetric Bayesian Nash Equilibrium is no longer appropriate. Given CDF $R(q_j)$ and PDF $r(q_j)$ over a competitor's performance, the probability that teacher i outperforms teacher g is $pr(q_i > q_g) = pr(e_i + s_i + \epsilon_i + \mu > e_g + s_g + \epsilon_g + \mu) = pr(e_g + s_g + \epsilon_g + < e_i + s_i + \epsilon_i) = R(e_i + s_i + \epsilon_i)$. The probability that teacher i outperforms any opponent is $R(e_i + s_i + \epsilon_i)$ and the probability that teacher i loses to any opponent is $1 - R(e_i + s_i + \epsilon_i)$. The logic of the derivation of the marginal probability of effort (MPE) in Chapter 3 is still applicable. Since all competitors are treated as ex-ante identical, they are treated as identical in deriving the formula for the marginal probability of promotion:

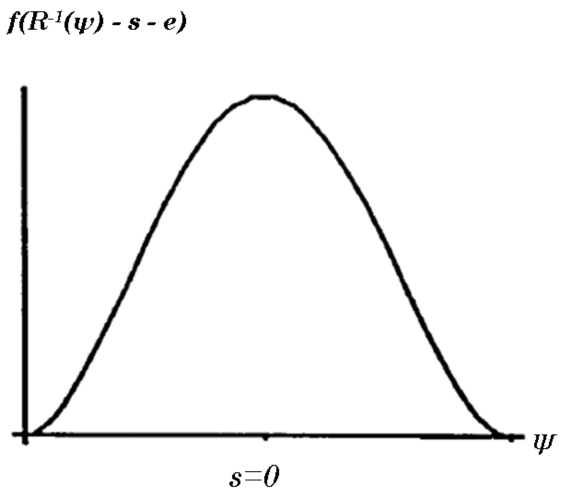
$$\frac{dp}{de} = (n-k)\binom{n-1}{n-k}\int R(e_i+s_i+\epsilon_i)^{n-k-1}(1-R(e_i+s_i+\epsilon_i))^{k-1}r(e_i+s_i+\epsilon_i)f(\epsilon)^2 d\epsilon \quad (4.1)$$

The order statistic part of this equation depends on the CDF $R(.)$ and PDF $r(.)$ of the random output of teacher i's competitors. However, the only stochastic part of his output to integrate over (from the teacher's point of view) is ϵ so $f(\epsilon)$ appears instead of an additional $r(.)$. Since e and s are constants, $d(e+s+\epsilon) = d\epsilon$. We can simplify (4.1) by using change of variables $\Psi \equiv R(e+s+\epsilon)$:

$$\frac{dp}{de} = \int (n-k)\binom{n-1}{n-k}\Psi^{n-k-1}(1-\Psi)^{k-1}f(R^{-1}(\Psi)-s-e)d\Psi \quad (4.2)$$

where the subscript i has been dropped for notational convenience. When s is distributed symmetrically, and when variations in optimal effort across skill types are small relative to variations in s and ϵ, then q_j will also be distributed approximately symmetrically and $R^{-1}(\Psi)$ will be approximately symmetric about the effort of a teacher with average skill, $s = 0$. Figure 4.1 shows this graphically.

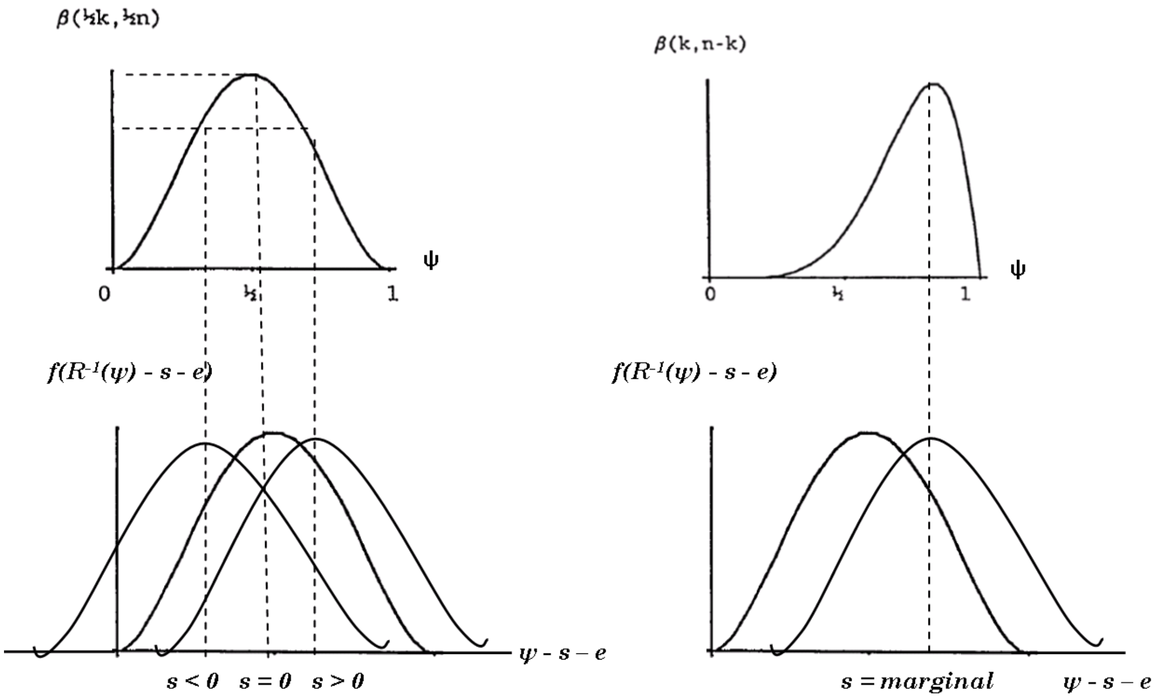
Figure 4.1: Skill, luck and effort



Apart from $f(R^{-1}(\Psi)-s-e)$, (4.2) has the form of a Beta density function $\beta(k, n-k)$ and just as in Chapter 3, properties 1 - 3 of Beta densities also apply here.

For an employee with average skill, incentives are highest when the promotion rate is 1/2, as they will be on the margin of being promoted versus not being promoted. High skill teachers do not need to work as hard to be promoted, they would require a very low draw of luck in order not to be promoted. However, very low draws of luck are unlikely. Figure 4.2 shows that $f(R^{-1}(\Psi) - s - e)$ will shift right over [0,1] for higher than average skilled teachers, given that variations in effort are not too large. Skill has the effect of decreasing the luck required to win. So incentives are low for high skilled teachers. Low skill teachers would need a very high draw of luck in order to win a promotion. Very high draws of luck are also unlikely, so low skilled teachers also have low incentives.

Figure 4.2: The marginal-skilled teacher



Define a marginal-skilled teacher as one whose skill shifts $f(.)$ so that it peaks in the same region as the beta density in (4.2). The marginal skill level depends on the promotion rate, p^*, since the mode of $\beta(k, n-k) = 1 - p^*$. This is the teacher for whom effort incentives are highest; it is exactly the point at which the highest Beta density values are given the highest weight. Intuitively, if for example 1/3 of teachers are promoted, the teacher with the highest incentives is the one who believes she is of the

2/3 skill percentile of teachers (from the bottom). She is on the margin between being promoted and not being promoted, and small amounts of luck can determine whether she wins the promotion or not, so effort will make the most difference to her probability of promotion. Incentives decrease monotonically from the marginal skill percentile for those both below and above this skill percentile.

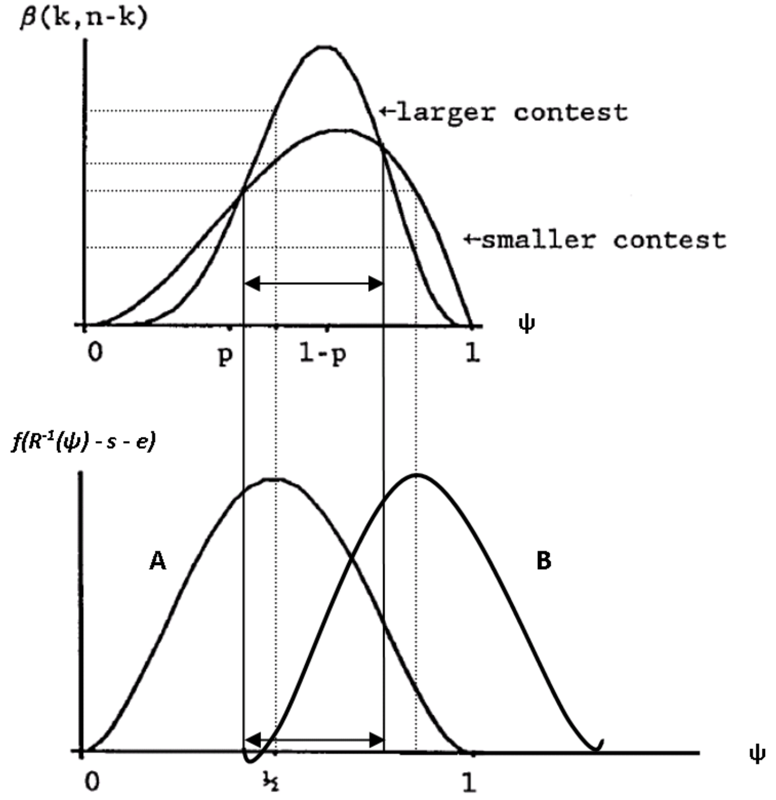
When either n changes (leaving k the same) or k changes (leaving n the same) p^* changes. When p^* changes, the teacher with the highest incentives changes as the marginal skill percentile is different. If n increases (leaving k the same) then p^* decreases. The marginal skill percentile is higher. The opposite is true when n decreases. Some teachers will have higher incentives and some will have lower incentives than before. However, it is still true that the marginal-skilled teacher will have the highest incentives and incentives will decrease monotonically as the skill percentile increases or decreases from that point.

When n increases but p^* stays the same, the effects are different. Gibbs (1989) finds that, in simulations with identical contestants, increasing n leaving p^* the same increases incentives when p^* is between 1/3 and 2/3. The intuition is that with more people, the variance in the order statistic that must be beaten to win a promotion is less. So for promotion rates in the middle, it is more worthwhile to put in effort to beat luck. But for promotions with very high (low) rates, it makes it even less likely that one will lose (win) so trying to beat luck is less productive. With identical contestants this effect is the same for all teachers. With heterogeneous contestants, the logic holds, but the effects are different for different teachers. For teachers with skill percentile at or near the margin, incentives increase. However, for those further from the margin, incentives decrease. Figure 4.3 below illustrates this point, showing that effort incentives increase for teachers whose skill percentile falls between the arrows, and decreases for those whose skill percentile falls outside of them [1]. Teacher A's skill percentile is closer to the marginal skill percentile than teacher B's skill percentile.

Thus, effort incentives are decreasing in the distance between where the teacher falls in the skill distribution and the probability of being promoted. Beliefs on skill are very important here. A teacher's incentives for effort depend on where they believe they fall

[1] Once again, if the promotion system were based on a standard, the number of teachers competing would not matter.

Figure 4.3: Heterogeneous competitors and number competing



in the skill distribution.

In sum, the one-period Gibbs model with skill added, predicts the following:

1. After the highest promotion, effort will cease ($p^* = 0$);

2. Incentives are highest for the teacher that believes she is on the margin of being promoted versus not being promoted. This teacher is the one for whom the skill percentile is equal to $1 - p^*$. Incentives are lower for those above or below that skill percentile and decrease monotonically;

3. When k (n) changes leaving n (k) the same, p^* changes so the marginal skill percentile changes. This is still where incentives are highest and incentives are lower for those with skill percentile above or below the new marginal skill percentile; and,

4. When n increases but p^* stays the same, increasing the number of teachers will increase incentives for those with skill percentile near the margin, and will decrease

incentives for those further away.

4.1.2 Multi-Period Model of Promotions

This section will add a dynamic component to the model. Assume teachers have careers lasting T periods. They are eligible for promotion after serving X years. In year 1, they start with a prior belief (s_1) about their relative skill rank, where $1 \leq s_1 \leq 1/N$, and a teacher's true relative rank is s. This belief can change over time (s_t) as teachers update their beliefs based on new information.

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 true 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 ($p_t(e_{t-1}, e_{t-2}, e_{t-3}, e_{t-4}, e_{t-5})$). Two simplifying assumptions are made: (1) that the impact of effort in one year on the probability of future promotion is independent of 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 ($dp_t/de_{t-1} = dp_t/de_{t-2} = ... = dp_t/de_{t-5}$).

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:

$$EV_j = -c(e_j) + Ep_{j+1} \sum_{t=j+1}^{T} \beta^{t-j} U_h + (1-Ep_{j+1})[Ep_{j+2} \sum_{t=j+2}^{T} \beta^{t-j} U_h + (1-Ep_{j+2})\{Ep_{j+3} \sum_{t=j+3}^{T} \beta^{t-j} U_h + (1-Ep_{j+3})(Ep_{j+4} \sum_{t=j+4}^{T} \beta^{t-j} U_h + (1-Ep_{j+4}) Ep_{j+5} \sum_{t=j+5}^{T} \beta^{t-j} U_h)\}] \tag{4.3}$$

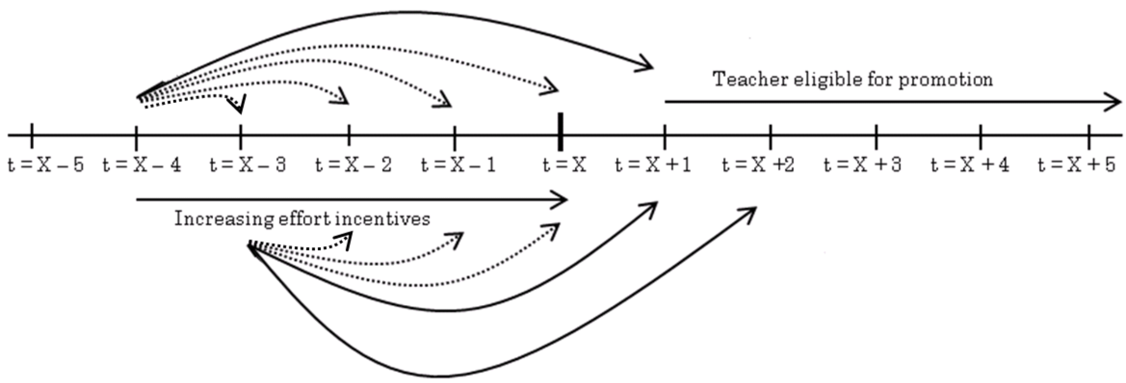
Here, $c(e_j)$ is the disutility of effort in year j, β is the discount rate, and Ep_t denotes expected promotion probability given the distribution of draws of luck. Current effort can alter the probability of being promoted in the subsequent five years, and once promoted the teacher enjoys higher utility 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 (4.3) with respect to e_j defines optimal effort by the teacher. If a teacher's beliefs about s are fixed, then the only thing that changes over time is that T becomes nearer so that there are fewer future years of higher wages to enjoy, reducing the expected benefits of promotion, thus leading to lower effort over time.

Given the simplifying assumptions, effort in the current year has an identical marginal effect on expected promotion probability in the subsequent five years unless the teacher is ineligible for promotion in those years ($t \leq X$). In those years, the expected probability of promotion is zero, as is the marginal probability of effort in all previous years. Thus, as long as $t \leq X-5$, there is no incentive to exert effort because in all of the subsequent five years, the teacher is ineligible for promotion. It is easy to see that in the years from $X-4$ to 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, and the first year of possible promotion is closer to the present (higher discount rate) increasing the discounted payoff from promotion. Figure 4.4 illustrates this point. 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 $t = X+2$). In year X and 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 longer an incremental increase in effort incentives over time associated with changes in the promotion eligibility window.

Next, we can return to the teacher's beliefs about where he or she lies in the skill distribution (s_t). As shown earlier in the one-period model, the marginal probability of effort is maximised when $s = 1-p^*$, or when one's relative skill position coincides precisely with the share of teachers promoted. As the absolute value of $s-(1-p^*)$ increases, the marginal probability of effort falls. Even if a teacher knows his or her own skill level perfectly, the teacher is unlikely to have full information on the skill of other teachers, so will not know for sure his or her relative skill level.

Figure 4.4: Effort incentives over time



If a teacher initially (in period 1) believes that her relative skill level is s_1, then new information about her relative performance enables the teacher to update beliefs about her skill level. I focus on one key piece of information about relative performance that is observable to both the teacher and the econometrician – whether the teacher was promoted or not in the past year. 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 $t > X$.

Of course, there may be a number of considerations which influence the extent to which the teacher believes the failure to be promoted is informative about her relative skill level. For example, if very few teachers can be promoted because of few quota positions available for higher ranks (the implied $1-p^*$ would then be very high) or if the teacher knows some other teachers are better performers and are likely to be promoted before her (meaning that she believes her skill percentile is further from $1-p^*$ than that of her competitors), then it may take repeated promotion failures before she significantly revises downward her assessment of her relative skill level. As additional years pass without promotion, the teacher's estimate of her relative skill level falls, further increasing the gap between estimated relative skill s_t and $1-p^*$. From the one-period model, we know that the larger this gap, the lower the marginal probability of effort. Thus, if the failure to be promoted is revealing information to teachers about their skill level, we would expect a teacher's effort to decline as she continues to be passed over for promotion year after year [2].

In summary, the multi-period model of teaching performance over time generates the following additional predictions[3]:

5. Effort incentives are expected to be zero, five and more years before a teacher is eligible for promotion (when $t \leq X - 5$);

6. Effort incentives are expected to increase over the five years preceding eligibility for promotion to the next rank (from $t = X - 4$ to $t = X$);

7. If a teacher continues to be passed over for promotion when she is eligible for promotion, then her effort will decline (many years after $t = X$); and,

8. A teacher's relative skill could increase over time with accumulated teaching experience.

4.2 Identifying Effort Incentives

This section will discuss the empirical model that will be used to test the theoretical model's predictions on individual teacher effort, and will discuss some strengths of, as well as challenges to identification.

In this chapter, I will focus on estimating the first of the static model's behavioural equations, but for individual levels of effort instead of average levels of effort, and will add to it the predictions derived from the multi-period model. The static model's basic behavioural equation was presented in Chapter 3 and is reproduced here, with both skill and the dynamic aspect added. Again, the higher the wage gap, the closer $1 - p^*$ is to the teacher's skill percentile, the larger the pool of competitors (as long as the teacher's skill percentile is close to $1 - p^*$), and the closer t is to the time a teacher becomes eligible to apply for a rank promotion ($t \to X$), the higher are incentives:

[2] The declining effort is the result of the promotions not being 'up or out' promotions.

[3] See Appendix 2 for proofs of these predictions

$$e = f(\Delta EV, MPE), MPE = h(n, p^*, s, t)$$
$$e = e(\Delta EV, n, p^*, s, t) \tag{4.4}$$

I can estimate (4.4) as follows:

$$ev_{irct} = \delta_1 \boldsymbol{D}_{irct} + \delta_2 \Delta w_{rc} + \delta_3 n_{ir\tau t} + \delta_4 p^*_{rc} + \delta_5 a_{ir\tau t} + v_{irct} \tag{4.5}$$

where ev_{irct} are evaluation scores of teacher i at rank r in county c at time t, Δw and p^* are the increase in wages and the promotion rate at rank r in county c, and a is ability. As a proxy for ability I will use an 'ability index'. The ability index was constructed using the teachers' scores on three teacher tests that were taken before the teacher began teaching (an education test, a language test, and a psychology test). Scores on these tests are included by first standardising the scores by the year in which they were taken, and then for each year of the data (2003-2006) the following measure is calculated: $[S - ave(S)]/ave(S)$ where S is the standardised score on the tests for a teacher, and $ave(S)$ is the average of the standardised scores of the teachers in the same rank in the same township in that year, to capture relative ability. The index was constructed using Principal Components Analysis (PCA) [4]. Since the model predicts that the effect of teachers varies with how close the skill percentile is to the promotion rate (Prediction 2), I include dummies for ability in the top and bottom 10% of the ability index. At the top and bottom of the ability index, we would expect that the skill percentile will be far from the promotion rate.

To measure n, I use the (log of) number of teachers in the same rank in the township [5]. Since the prediction on the number of teachers competing is also about the extremes of the skill distribution (Prediction 4), interactions between the number of teachers and

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

[5] The number of teachers in each rank is adjusted by the undersampling 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 to calculate the sampling rate and the undersampling rate. The number of teachers in each rank was then adjusted as follows: number interviewed * (1 + undersampling 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.

ability in the top and bottom 10% of the skill distribution are also included in addition to n.

In order to capture time ($\boldsymbol{D}$), I include a dummy variable for the years in which $t \leq X - 5$ in order to test Prediction 5, as well as separate dummies for the years $t = X - 4, t = X - 3, t = X - 2, t = X - 1$, and $t = X$ so that I can look specifically at the years leading up to eligibility for promotion (Prediction 6). I also include dummies for two to fifteen (or greater) years after eligibility so that I can look more closely at incentive effects many years into a rank (Prediction 7). The omitted category is the first year of eligibility to apply for a promotion, $t = X + 1$.

The evaluation scores measure a combination of effort and ability. To the extent that the ability index captures underlying skill, then the evaluation score can be more correctly interpreted as capturing effort when ability is included in the specification. However, this would not be able to correct for any unobserved aspects of skill. To correct for this (4.5) is estimated with teacher fixed effects, since there is four years of data for each teacher. As long as these unobserved aspects of skill do not change over time, then the teacher fixed effect can correct for this omitted variable. This leads to:

$$ev_{it} = \delta_1 \boldsymbol{D_{it}} + \delta_2 a_{10,it} + \delta_3 a_{90,it} + \delta_4 n_{it} + \delta_5 n_{it} * a_{10,it} + \delta_6 n_{it} * a_{90,it} + \phi_i + v_{it} \quad (4.6)$$

where $\boldsymbol{D_{it}}$ is the set of dummy variables for years before and after eligibility for promotion, $a_{10,it}$ is a dummy variable for whether the teacher falls into the bottom 10% of the ability distribution (10th percentile), and $a_{90,it}$ is a dummy variable for whether the teacher falls into the top 10% (90th percentile) of the ability distribution in that rank in the township, n is the (log) number of teachers in the same rank and township, ϕ_i is the individual teacher fixed effect, and v_{it} is the error term. The change in wages and the promotion rate can no longer be included, as these variables are at the county level and as these would not vary for individual teachers, they would be swept away by the fixed effect.

This measures within-teacher variation over time, and identification is off teachers that change from one category to another. The dummies for relative ability in the top or bottom 10% are identified off teachers moving into or out of these categories over time as the composition of their competitors changes. The fixed effect addresses concerns that there is a selection problem. Since we are looking at the number of years into a promotion but I have data only for 2003-2006, this means that I have information on people in the sample for different periods of time after promotion. The reasons we observe teachers for a specific time frame is not random; those teachers that have been in a rank for longer could be different from those that got promoted earlier.

I expect that the coefficient on the dummy variable on $t \leq X - 5$ will be zero. I expect the coefficients on the time dummies to be increasing as teachers near the year of eligibility and then to be decreasing many years after year X; I predict an inverted-U shape. The inclusion of n interacted with the ability dummies will test Prediction 4. I expect the coefficients on these dummies to be negative ($\delta_5 < 0$ and $\delta_6 < 0$). Finally, I focus on three rank levels for the effort regression: Primary 1, Middle 2 and Middle 1. This is because promotions are almost automatic between Primary 2 and Primary 1 and between Middle 3 and Middle 2, and there are no further promotions after Primary high and Middle high. As a result, Primary 1, Middle 2 and Middle 1 rank levels are the ones with promotion incentives. However, to see the incentive effects at the highest level (Prediction 1), I run a regression for Primary and Middle high levels, but instead of pre- and post-eligibility time dummies, I use dummies for the number of years into the promotion (from two years after promotion to Primary or Middle high to fifteen or more years after promotion to Primary or Middle high, leaving the first year after promotion as the omitted category once again).

There are some potential challenges to the estimation strategy. First, I do not include many teacher personal characteristics. Those that are time-invariant are swept away by the teacher fixed effect. However, if there are time-varying unobservables that are correlated with variables that are included in the model, then coefficient estimates would be biased. However, I do not expect teachers' personal characteristics to vary systematically with time into a rank or the number of teachers in the township for example, and thus their omission should not present a problem.

Second, measurement error could be another potential source of bias. Here, we would be most concerned with measurement error in the evaluation scores, or with systematic bias in the evaluation scores. The systematic bias could arise if, for example, principals tend to inflate teachers' performance on their evaluations (perhaps as a source of encouragement). For these regressions, measurement error in the evaluation scores would only be a problem if it were correlated with the regressors. This would be the case if, for example, principals tended to over-evaluate teachers' performance if the teacher had more experience or had been in a rank for longer, or was coming up for promotion. However, I do not find evidence for such effects in the data (see Section 4.3 for further details).

We would also be concerned with measurement error on the number of teachers. As mentioned previously, not all teachers were interviewed in all schools. As such, I have also tried restricting the sample to schools with a sampling rate of 25% and above and 35% and above to check that the results are robust to sampling rates and the adjustment described above.

The next section will present the results from estimating the empirical model.

4.3 Results

This section will present the results from estimating the empirical model of this chapter. I will first discuss the results from estimating equation (4.6) along with a discussion of the issues of selection and measurement error, and will then present some robustness checks that were performed in order to ensure the estimation was conducted appropriately.

4.3.1 Effort Incentives

The results from estimating equation (4.6) for effort in Primary 1, Middle 2 and Middle 1 rank levels are contained in Table 4.1. Figure 4.5 plots the coefficients on the time dummies for Primary 1, Middle 2 and Middle 1. The patterns of the time dummies are what were predicted: an inverted-U shape for Primary 1, Middle 2, and Middle 1 levels. The top panel of Figure 4.5 shows that effort is zero when $t \leq X - 5$. As predicted by the multi-period model (Prediction 5), since it is the last five evaluation scores that matter, in time periods five years and before eligibility for promotion, effort is zero. From $t = X - 4$ to $t = X$ effort is increasing, but the increase in effort perhaps starts a bit later than the model predicted. Teachers are getting closer to becoming eligible for promotion, and so are working harder. The coefficient on $t = X - 1$ is significant at the 5% level. The average evaluation score in the first year after eligibility ($t = X + 1$) is 2.65 (between a good and excellent score). In $t = X - 2$ and $t = X - 1$, the evaluation score is more likely to be an excellent score.

The bottom panel of Figure 4.5 shows that many years after eligibility, effort is decreasing. 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 effect is significant at the 1% level. This is evidence of Prediction 7; as time passes there are fewer periods in which one could enjoy the additional income from promotion, and teachers may be revising downwards the estimate of their skill percentile. As a result of these two factors, effort incentives are lower[6].

At the important ranks of Primary 1, Middle 2 and Middle 1, the number of competitors matters. We can see from Table 4.1 that the coefficients on the number of teachers interacted with ability in the top and bottom 10% of the ability distribution, where we expect ability and the promotion rate to differ substantially, are negative, and the coefficient on the number of teachers interacted with ability in the bottom decile is significant. This is evidence of Prediction 4 of the model: that more competitors increases effort incentives for those close to one minus the promotion rate, and decreases incentives for those further away. These results are statistically stronger than those of Chapter 3 but are still relatively weak.

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. Table 4.2 contains results of the effort regression for Primary and Middle high ranks, and Figure 4.6 plots the time dummy coefficients. The coefficients on the time after promotion dummies are steadily decreasing. The longer a teacher stays in this rank,

[6]The regressions were also conducted separately for male and female teachers (results not reported here). There were no differences for these groups.

Table 4.1: Primary 1, Middle 2, and Middle 1 rank effort incentives

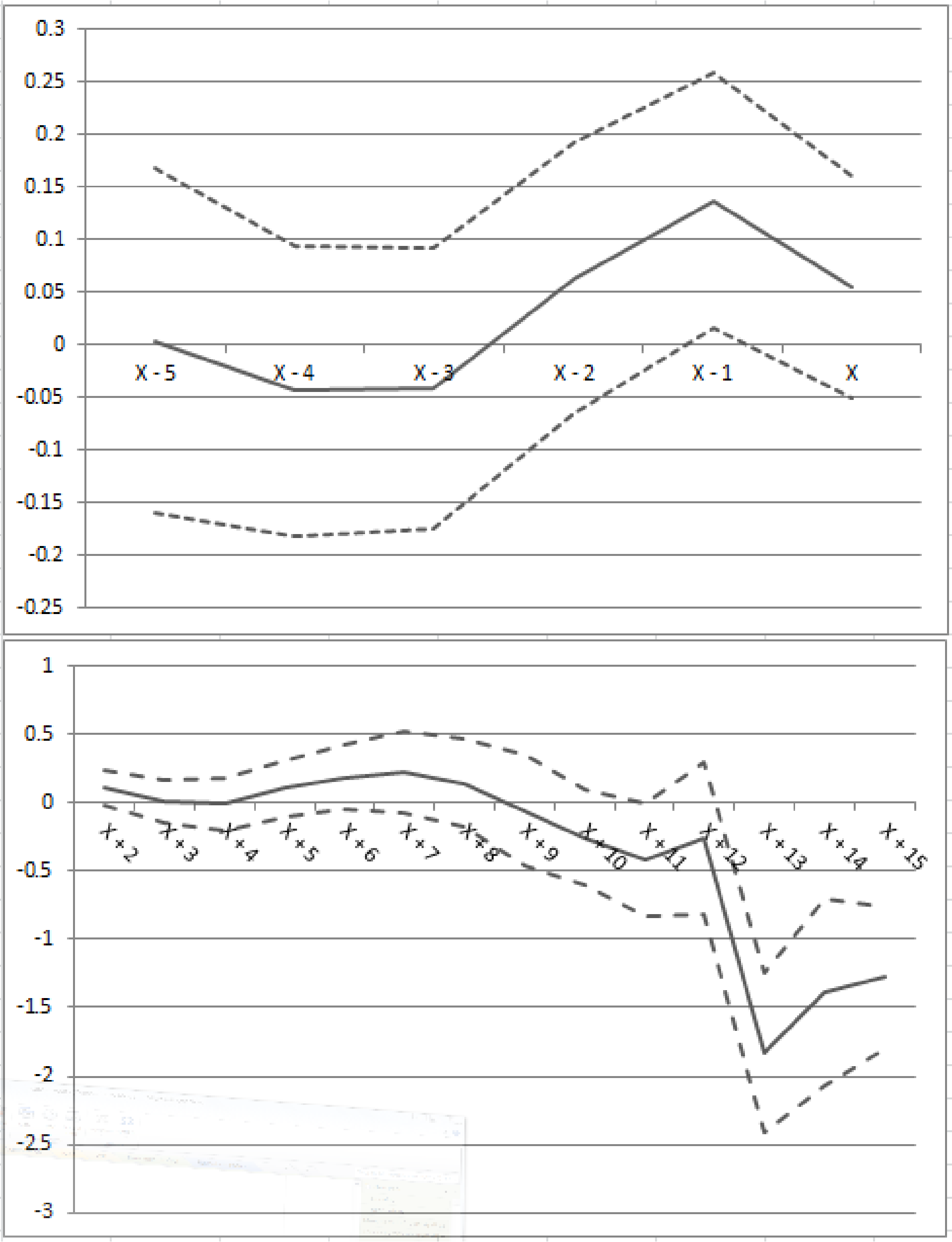
Outcome: evaluation score of teacher i in year t		
	(1)	
	Coef	Se
Relative age	0.555	(0.347)
Log number of teachers in rank in township	0.054	(0.081)
N * ability top 10%	-0.098*	(0.052)
N * ability bottom 10%	-0.050	(0.063)
Ability in top 10%	0.353*	(0.194)
Ability in bottom 10%	0.055	(0.200)
t $\leq X - 5$	0.004	(0.083)
t = X - 4	-0.044	(0.070)
t = X - 3	-0.041	(0.068)
t = X - 2	0.063	(0.066)
t = X - 1	0.137**	(0.062)
t = X	0.056	(0.054)
t = X + 2	0.117*	(0.064)
t = X + 3	0.010	(0.082)
t = X + 4	-0.004	(0.097)
t = X + 5	0.108	(0.103)
t = X + 6	0.191	(0.121)
t = X + 7	0.228	(0.156)
t = X + 8	0.146	(0.163)
t = X + 9	-0.051	(0.211)
t = X + 10	-0.256	(0.178)
t = X + 11	-0.420**	(0.212)
t = X + 12	-0.260	(0.282)
t = X + 13	-1.834***	(0.299)
t = X + 14	-1.385***	(0.348)
t $\geq X + 15$	-1.277***	(0.268)
Observations	3683	
R^2	0.020	

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.

Figure 4.5: Primary 1, Middle 2 and Middle 1, pre- and post-eligibility time dummies



Note: Dashed lines are 95% confidence intervals.

the less hard she works. This is not what is expected according to Prediction 1; that effort ceases after the final promotion. The model predicts a flat line. However, a reason for the downward trend could be that teachers at Primary and Middle high level are getting older and are 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 ability or the number of competitors, and as in Table 4.2, these coefficients do not matter.

Table 4.2: Primary and Middle high rank effort incentives

Outcome: evaluation score of teacher i in year t		
	(1)	
	Coef	Se
Relative age	-1.218	(0.763)
Log number of teachers in rank in township	-0.267	(0.309)
N * ability top 10%	0.139	(0.229)
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	

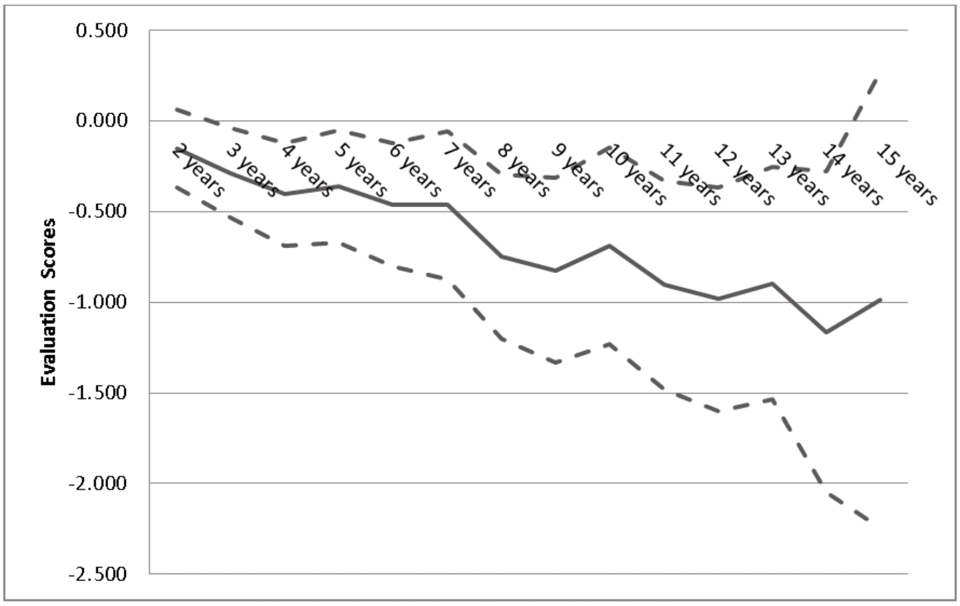
Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Regression includes teacher fixed effects and clustered standard errors at the teacher level.

In addition, the same pattern can be seen in the number of hours Primary and Middle high teachers spend on various tasks. The 2007 GSCF survey asks each teacher "in

Figure 4.6: Primary and Middle high time dummies



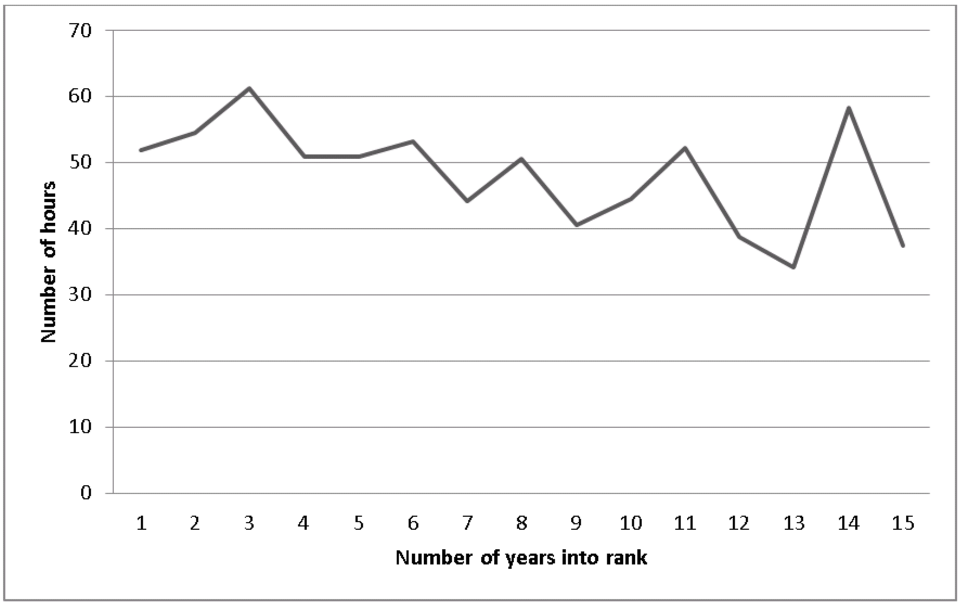
Note: Dashed lines are 95% confidence intervals.

a typical week, how many hours do you spend on...". The information collected is for: grading homework, preparing lesson plans, participating in teaching and research activities, coaching of students outside class, organising extracurricular activities, home visits, and disciplining students. I sum up the total number of hours spent on these activities in a given week, and plot this against the number of years into Primary and Middle high rank in Figure 4.7. Here as well, effort is declining over time.

The results from the Primary high level regressions also provide insight into the issue of selection. For example, one could argue that the downward slope we see a few years after eligibility for Primary 1, Middle 2 and Middle 1 levels is just an effect of poor teachers. However, we see that at Primary and Middle high level there is also a downward slope, and Primary and Middle high level teachers are certainly not poor teachers. Furthermore, the average age of Primary and Middle high teachers is 47 years old (and is 50 years old for teachers between 10 and 15 years into the rank). These are teachers nearing retirement, and they can be expected to slow down [7]. In Primary 1, Middle 2 and Middle 1 ranks, the teachers in the post-eligibility period are on average 10 years older than those in the pre-eligibility period. Table 4.3 reports the average age of teachers in each of the time

[7] The retirement age for teachers in China is 55 years for men and 50 years for women.

Figure 4.7: Primary and Middle high hours spent with students



dummy categories of Table 4.1. Ten years after eligibility, the average age is approximately 45. So it is likely that the teachers are getting older and are slowing down here as well. In the terminology of the multi-period model, they are nearing $t = T$, so the number of periods of potential additional income is decreasing, and teachers are likely revising downwards their skill percentile, thereby facing decreasing incentives. Anecdotal evidence also suggests that older teachers care less about promotions; they tend to focus on their responsibility to students and parents. It is the younger teachers that are more motivated by promotions. In addition, since this is a teacher fixed effects regression, we are looking at changes over time for the same teacher.

An argument could also be made that the increasing trend that we see before eligibility is simply teachers learning over time as they are in the first few years into the rank. 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 doing the same job as in previous rank levels. Although they are likely still learning, the upwards slope cannot solely be explained by a learning effect, as these teachers are likely to already have completed a large part of the learning curve.

Another challenge may be that principals manipulate evaluation scores of teachers that are nearing promotions, awarding teachers that are becoming eligible with high evaluation

Table 4.3: Average age of teachers in pre/post promotion eligibility years

Years pre/post eligibility	Mean	Sd	Minimum	Maximum	N
t = X - 5	30.03	5.87	22	58	622
t = X - 4	31.15	5.29	22	56	330
t = X - 3	32.45	5.34	23	57	322
t = X - 2	33.29	4.82	21	54	284
t = X - 1	34.1	5.09	22	54	263
t = X	35.57	5.28	23	55	256
t = X + 1	36.44	5.32	24	56	202
t = X + 2	37.67	5.33	27	57	175
t = X + 3	38.2	4.84	28	54	142
t = X + 4	38.5	4.62	29	54	111
t = X + 5	39.8	4.99	30	55	92
t = X + 6	41.15	4.61	34	56	68
t = X + 7	43.32	5.49	27	57	47
t = X + 8	45.41	5.24	28	55	32
t = X + 9	44.88	5.74	29	53	26
t = X + 10	45.65	4.93	38	54	20
t = X + 11	43.67	3.39	40	47	6
t = X + 12	44.67	3.39	41	48	6
t = X + 13	40.67	2.08	39	43	3
t = X + 14	43.2	2.59	40	46	5
t = X + 15	45.63	5.16	38	55	19

scores so that they can be promoted, and so the scores do not reflect true effort. Any such manipulation of evaluation scores would also produce an upward trend pre-eligibility, and would mean that there is systematic measurement error that is correlated with the time to eligibility for a promotion. In addition, one may question the use of evaluation scores as an accurate measure of effort, as they also measure ability. In order to investigate these points further, two different tests of the data are performed. First, I run a Logit regression with fixed effects of the probability of obtaining an 'excellent' or 'good' evaluation score [8] in 2006 on the log of the total time use variable, as well as many teacher controls. This regression can be carried out only in the cross section, since there was only one year of teacher time-use data collected. The coefficient estimates are contained in Table 4.4 [9]. Columns (1) and (3) include only the log of total time use and columns (2) and (4) add

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

[9]Marginal effects can not be computed for Logit models with fixed effects, as the marginal effects depend on the value of the fixed effects. As such the sign and significance of the effects can be estimated, but not much can be said of the magnitude of the effects.

the controls. Columns (1) and (2) include county fixed effects and columns (3) and (4) include township fixed effects. Total time use is positively and significantly related to the evaluation scores. So, it is not just principals manipulating scores, the evaluation scores are related to actual effort.

The second test is to first check which counties have the highest correlations between the evaluations and these measures of time use by running the above Probit model separately by county, and then re-run (4.6) restricting it to only those counties with strong correlations. If we find that the same pattern still holds, then this is another argument in favour of the evaluations reflecting true effort. In fact, this is what I find (see Figure 4.8 and Appendix table 4.5). When I restrict the regression to only counties with positive and significant correlations between evaluations and log of total time use, I find that the upward trend from $t = X - 4$ to $t = X$ is preserved, as is the downward trend for many years after eligibility. The results of these two tests indicate that the effects found in estimating (4.6) are indeed due to effort, and not due to principals manipulating the scores of teachers nearing promotion eligibility. They also indicate that the evaluation scores are a good measure of effort.

Other papers also provide evidence against the evaluation scores being solely manipulation. If principals were awarding high evaluation scores to those nearing promotion eligibility, all teachers would get promoted eventually and the ranks would have no meaning. There would be no correlation between ranks and performance. However, Park and Hannum (2001) showed that rank is a significant correlate in a regression of student test scores on various teacher characteristics. Ding and Lehrer (2001) find that teaching quality is being rewarded effectively in Jiangsu province (the evaluation and promotion system is the same across China). The fact that in each rank some teachers are never promoted, is also evidence that promotions are not automatic, and that evaluations are not manipulated to ensure that everyone is promoted. Finally, because evaluation scores are announced at teacher meetings and because principals are also evaluated and promoted based on the performance of their school, this should also reduce the incentive to inflate evaluation scores.

Another possible type of manipulation is principals assigning teachers to easier or

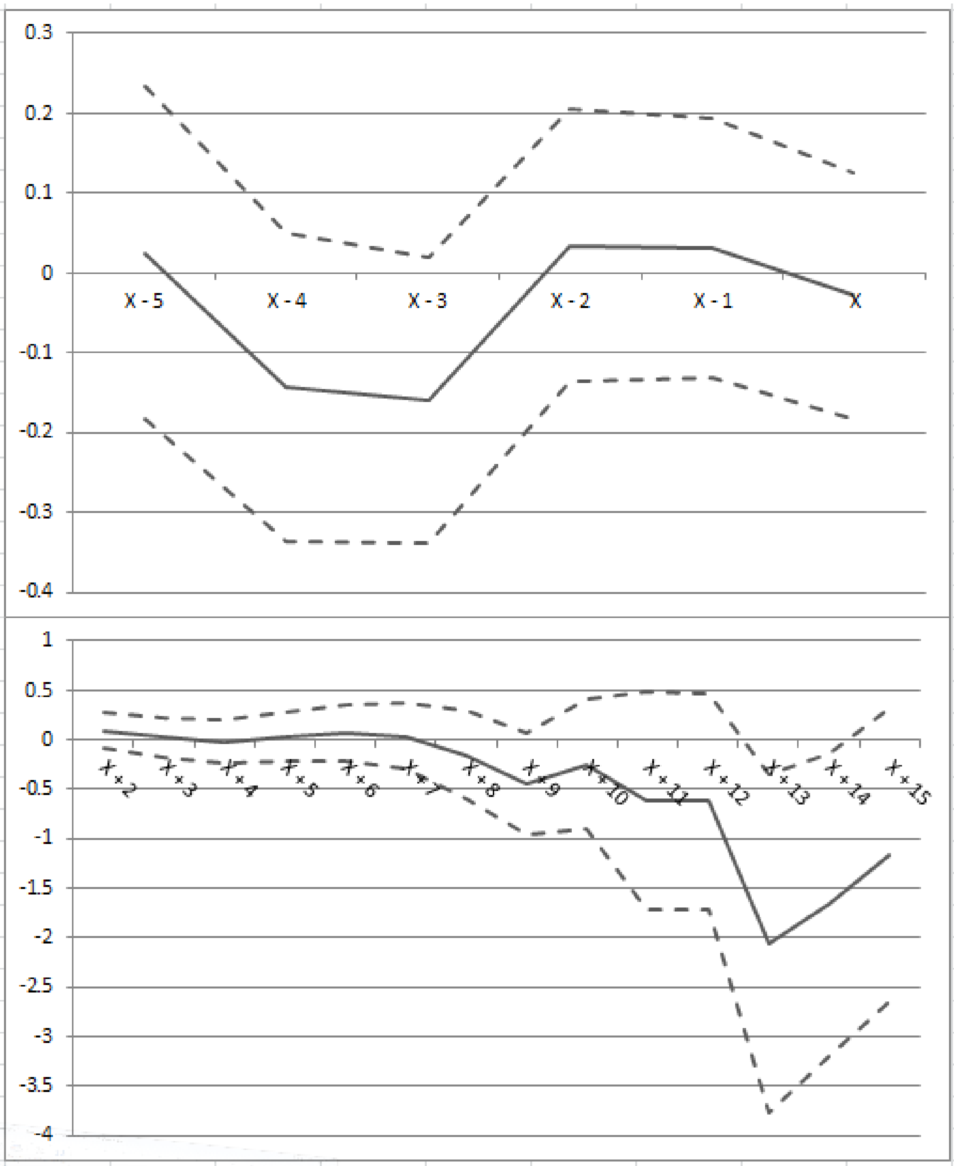
Table 4.4: Probability of Receiving an 'Excellent' or 'Good' evaluation score

	Outcome: excellent or good evaluation score							
	(1)		(2)		(3)		(4)	
	coef	se	coef	se	coef	se	coef	se
Log of total time use	0.322**	0.162	0.381**	0.189	0.344**	0.166	0.357*	0.195
Education - vocational college			0.024	0.201			-0.077	0.208
Education - college or higher			0.149	0.224			0.086	0.235
Female			-0.100	0.141			-0.073	0.148
Age			0.013	0.011			0.014	0.011
From same village			1.120	0.826			1.093	0.826
From same town			1.182	0.825			1.242	0.825
From same county			1.055	0.819			1.109	0.819
From same province			1.240	0.852			1.204	0.854
Married			0.261	0.381			0.319	0.383
Divorced			-1.676	1.129			-1.549	1.139
Widowed			0.402	0.846			0.519	0.867
Ability in bottom 10%			-0.146	0.162			-0.151	0.171
Ability in top 10%			0.130	0.161			0.109	0.169
County fixed effects	Yes		Yes		No		No	
Township fixed effects	No		No		Yes		Yes	
Number of observations	1,376		1,084		1,363		1,064	

Note: $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Standard errors clustered at the level of the fixed effect.

Figure 4.8: Counties with positive and significant correlations between time use and evaluation scores



Note: Dashed lines are 95% confidence intervals.

better classes or grades within the school so that their student's test scores are higher. In Gansu, there is generally only one class per grade, and teachers also tend to track students throughout their school cycle and follow them to higher grades (Park and Hannum, 2001). In addition, if this type of manipulation was occurring, we would not see a strong relationship between the evaluation scores and time-use measures of effort. Student test scores are also only one of four components of the evaluation, which is conducted by an entire committee of people. Teachers would need several people to inflate several measures of performance. As a result, I do not believe this type of manipulation should be a concern in this context.

4.3.2 Robustness Checks

Since the prediction on ability is arguably the most important of the one-period model, I check to make sure that the cutoffs I have chosen for the dummies are not sensitive to small changes and are not driving the results. As a result, I re-estimate the effort regression using the top 15% and bottom 15% instead of the top and bottom 10% of the ability distribution. The results are contained in Table 4.6, Column (1) in the Appendix. The pattern of the time dummies is preserved, and the interpretation of the main results is unaffected. The coefficients are almost exactly the same in magnitude, in fact. The coefficient on the number of teachers interacted with ability in the bottom 10% loses significance, but as we move closer to the middle of the ability distribution, we expect the effect of the number of teachers competing to be weaker.

I also check that the results are robust to the inclusion of the number of teachers variable, since this is a variable that has been adjusted. Equation (4.6) is run again omitting schools with low sampling rates (below 25% and below 35%) but again with ability in the top and bottom 10%. Here as well, the magnitude of the time before and after promotion eligibility dummy coefficients changes slightly but the signs remain unchanged (see Appendix, Table 4.6 columns (2) and (3)). The coefficients on the number of teachers interacted with the top and bottom 10% of the ability distribution are still negative. Most importantly however, the shape of the time dummies remains unchanged in all cases, and the interpretation of the effect of the number of teachers also remains consistent with the previous results. So, I believe that undersampling of teachers should not be a cause for concern, it does not appear to be systematic, and the procedure by which the number of teachers was adjusted seems to be innocuous.

4.4 Summary

This chapter extended the model of Chapter 3 to one in which teachers are no longer identical, but differ in ability, and it also added a dynamic model of teachers' effort incentives. This allowed the focus to centre on how individual teacher effort responds to the interaction between teacher characteristics and various components of the promotion system, in order to shed light on design effects. The model's equation from the FOC was used to specify an empirical model by which to test the model's predictions on the data. This empirical equation was a regression of the teachers' reported annual evaluation scores (which proxied for effort) on dummies for some years before eligibility to apply for promotion to the next rank level, and dummies for years after a teacher becomes eligible for promotion, the number of teachers in the competition, and an ability index, with dummy variables for teachers in the top and bottom 10% of the ability distribution. The regression was run for the three rank levels of interest: Primary 1, Middle 2, and Middle 1 rank. The model predicted that effort incentives would be higher when the teacher's estimate of their skill percentile was closer to one minus the promotion rate (as the competition would thus be close and marginal effort would make a difference), when teachers neared the time of eligibility to apply for a promotion, and when there were more teachers competing (as long as the skill percentile was close to one minus the promotion rate so that the contest was 'close'). Effort incentives would be lower when teachers had been repeatedly passed over for promotions (as they would then revise downward their estimate of their ability and may give up), when ability was in the extremes of the skill distribution, and when the number of teachers competing increased and ability was in the extremes of the skill distribution. Predicted incentives were zero when teachers reached the level after which there were no more promotions, and whenever the teacher is five or more years away from becoming eligible for promotion, because of the way that annual performance evaluation scores feed into the promotion decision.

The predictions of the model are largely affirmed by the data. Effort is zero when the teacher is five or more years from promotion eligibility. Plotting the coefficients on the time dummies reveals that effort increases during the five years before eligibility for rank promotions, and decreases several years afterwards. Effort is also decreasing after the last promotion (Primary high). Teachers are getting older at this point and are not working as hard. The inverted-U shape that is present at the Primary 1, Middle 2 and Middle 1 rank levels can be explained not by selection, but by effort. As we see a declining trend in effort for teachers at the Primary high level where the model predicts a flat trend, this can be taken as evidence that older teachers slow down. The teachers in later years of Primary 1, Middle 2 and Middle 1 are also getting older. The increasing trend in effort as a teacher comes closer to promotion eligibility is also not a learning effect; these teachers have already been teaching for over a decade. Furthermore, two tests are developed that indicate that evaluations do in fact reflect true effort, as they are strongly related to actual measures of teacher time use. The number of teachers competing for a promotion also matters, but the results are not very strong, with a larger number of teachers leading to increased incentives for those in the middle of the ability distribution, and decreased incentives for teachers at the top and bottom of the distribution, as their skill percentile is likely to be very far from the promotion rate.

4.5 Appendix 1 Chapter 4 - Additional Tables

Table 4.5: Counties with significant correlations between evaluation scores and time use

Outcome: evaluation score of teacher i in year t		
	(1)	
	coef	se
Relative age	0.722	(0.473)
Log number of teachers in rank in township	-0.052	(0.115)
N * ability top 10%	0.015	(0.095)
N * ability bottom 10%	0.117	(0.106)
Ability in top 10%	0.162	(0.302)
Ability in bottom 10%	-0.469	(0.340)
t $\leq X - 5$	0.025	(0.106)
t = X - 4	-0.143	(0.099)
t = X - 3	-0.160*	(0.092)
t = X - 2	0.034	(0.087)
t = X - 1	0.030	(0.083)
t = X	-0.029	(0.078)
t = X + 2	0.096	(0.089)
t = X + 3	0.023	(0.100)
t = X + 4	-0.023	(0.113)
t = X + 5	0.030	(0.124)
t = X + 6	0.063	(0.146)
t = X + 7	0.037	(0.171)
t = X + 8	-0.152	(0.224)
t = X + 9	-0.443*	(0.265)
t = X + 10	-0.255	(0.334)
t = X + 11	-0.617	(0.558)
t = X + 12	-0.623	(0.558)
t = X + 13	-2.060**	(0.872)
t = X + 14	-1.667**	(0.781)
t $\geq X + 15$	-1.176	(0.755)
Observations	1615	
R^2	0.027	

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Regression includes teacher fixed effects and clustered standard errors.

Table 4.6: 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	0.584*	(0.303)	0.818**	(0.322)	0.754**	(0.326)
Log number of teachers in rank in township	-0.012	(0.060)	0.024	(0.070)	0.013	(0.073)
Ability in top 15%	0.108	(0.079)				
Ability in bottom 15%	-0.094	(0.080)				
N * ability top 15%	-0.030	(0.024)				
N * ability bottom 15%	0.007	(0.022)				
Ability in top 10%			0.336*	(0.172)	0.169	(0.183)
Ability in bottom 10%			0.073	(0.189)	0.054	(0.199)
N * ability top 10%			-0.088*	(0.050)	-0.045	(0.052)
N * ability bottom 10%			-0.066	(0.060)	-0.052	(0.062)
t $\leq X - 5$	-0.003	(0.073)	0.028	(0.075)	0.015	(0.076)
t = X - 4	-0.047	(0.067)	-0.023	(0.070)	-0.050	(0.072)
t = X - 3	-0.045	(0.062)	-0.007	(0.064)	-0.035	(0.065)
t = X - 2	0.062	(0.060)	0.048	(0.061)	0.033	(0.062)
t = X - 1	0.141**	(0.056)	0.113*	(0.058)	0.097*	(0.059)
t = X	0.062	(0.053)	0.064	(0.055)	0.073	(0.055)
t = X + 2	0.119*	(0.061)	0.131**	(0.062)	0.149**	(0.062)
t = X + 3	0.010	(0.070)	0.026	(0.072)	0.043	(0.072)
t = X + 4	-0.000	(0.082)	-0.019	(0.084)	0.001	(0.085)
t = X + 5	0.113	(0.092)	0.076	(0.095)	0.120	(0.096)
t = X + 6	0.193*	(0.106)	0.156	(0.110)	0.190*	(0.111)
t = X + 7	0.235*	(0.132)	0.226*	(0.136)	0.302**	(0.138)
t = X + 8	0.149	(0.152)	0.057	(0.159)	0.143	(0.162)
t = X + 9	-0.050	(0.178)	-0.095	(0.179)	-0.003	(0.183)
t = X + 10	-0.262	(0.201)	-0.304	(0.202)	-0.224	(0.203)
t = X + 11	-0.425	(0.336)	-0.470	(0.333)	-0.382	(0.333)
t = X + 12	-0.263	(0.337)	-0.256	(0.355)	-0.166	(0.355)
t = X + 13	-1.828**	(0.798)	-1.832*	(0.970)	-1.674*	(0.968)
t = X + 14	-1.428*	(0.731)	-2.117***	(0.779)	-2.071***	(0.778)
t $\geq X + 15$	-1.297*	(0.706)	-1.217*	(0.713)	-1.181*	(0.712)
Observations	3683		3373		3279	
R^2	0.019		0.021		0.021	

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.

4.6 Appendix 2 Chapter 4 - Proof of Predictions in Multi-Period Model

Proof of Prediction 5:

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 effort in year $t = X - 5$ is such that equation (4.3) is maximised, where $j = X - 5$:

$$
\begin{aligned}
EV_{X-5} = -c(e_{X-5}) + Ep_{X-4} \sum_{t=X-4}^{T} \beta^{t-(X-5)} U_h + (1 - Ep_{X-4})[Ep_{X-3} \sum_{t=X-3}^{T} \beta^{t-(X-5)} U_h \\
+ (1 - Ep_{X-3})\{Ep_{X-2} \sum_{t=X-2}^{T} \beta^{t-(X-5)} U_h + (1 - Ep_{X-2})(Ep_{X-1} \sum_{t=X-1}^{T} \beta^{t-(X-5)} U_h \\
+ (1 - Ep_{X-1}) Ep_X \sum_{t=X}^{T} \beta^{t-(X-5)} U_h)\}]
\end{aligned}
\tag{4.7}
$$

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 (4.7) drop out except for $-c(e_{X-5})$. Thus, effort is zero.

Proof of Prediction 6:

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 (4.3).

In $j = X - 4$:

$$EV_{X-4} = -c(e_{X-4}) + Ep_{X-3} \sum_{t=X-3}^{T} \beta^{t-(X-4)} U_h + (1 - Ep_{X-3})[Ep_{X-2} \sum_{t=X-2}^{T} \beta^{t-(X-4)} U_h$$
$$(1 - Ep_{X-2})\{Ep_{X-1} \sum_{t=X-1}^{T} \beta^{t-(X-4)} U_h + (1 - Ep_{X-1})(Ep_X \sum_{t=X}^{T} \beta^{t-(X-4)} U_h$$
$$+ (1 - Ep_X) Ep_{X+1} \sum_{t=X+1}^{T} \beta^{t-(X-4)} U_h)\}] \quad (4.8)$$

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, (4.8) simplifies to:

$$EV_{X-4} = -c(e_{X-4}) + Ep_{X+1} \sum_{t=X+1}^{T} \beta^5 U_h \quad (4.9)$$

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

$$\max_{e_{X-4}} \left\{ Ep_{X+1} \sum_{t=X+1}^{T} \beta^5 U_h - c(e_{X-4}) \right\} \quad (4.10)$$

The first order condition is:

$$c'(e_{X-4}) = \frac{dEp_{X+1}}{de_{X-4}} \sum_{t=X+1}^{T} \beta^5 U_h \quad (4.11)$$

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:

$$EV_{X-3} = -c(e_{X-3}) + Ep_{X+1} \sum_{t=X+1}^{T} \beta^4 U_h + (1 - Ep_{X+1}) Ep_{X+2} \sum_{t=X+2}^{T} \beta^5 U_h \quad (4.12)$$

The first order condition is now:

$$c'(e_{X-3}) = \frac{dEp_{X+1}}{de_{X-3}} \sum_{t=X+1}^{T} \beta^4 U_h + \frac{dEp_{X+2}}{de_{X-3}} \sum_{t=X+2}^{T} \beta^5 U_h$$
$$- Ep_{X+2} \frac{dEp_{X+1}}{de_{X-3}} \sum_{t=X+2}^{T} \beta^5 U_h - Ep_{X+1} \frac{dEp_{X+2}}{de_{X-3}} \sum_{t=X+2}^{T} \beta^5 U_h \quad (4.13)$$

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:

$$c'(e_{X-3}) = \frac{dEp_{X+1}}{de_{X-3}} \left(\sum_{t=X+1}^{T} \beta^4 U_h + \sum_{t=X+2}^{T} \beta^5 U_h (1 - Ep_{X+1} - Ep_{X+2}) \right) \quad (4.14)$$

Now we can compare (4.14) and (4.11). Again, because of simplifying assumption (2), we can substitute $\frac{dEp_{X+1}}{de_{X-3}} = \frac{dEp_{X+1}}{de_{X-4}}$ in (4.14) and compare:

$$\frac{dEp_{X+1}}{de_{X-4}} \sum_{t=X+1}^{T} \beta^5 U_h < \frac{dEp_{X+1}}{de_{X-4}} \left(\sum_{t=X+1}^{T} \beta^4 U_h + \sum_{t=X+2}^{T} \beta^5 U_h (1 - Ep_{X+1} - Ep_{X+2}) \right) \quad (4.15)$$

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:

$$c'(e_{X-2}) = \frac{dEp_{X+1}}{de_{X-2}} [\sum_{t=X+1}^{T} \beta^3 U_h + \sum_{t=X+2}^{T} \beta^4 U_h (1 - Ep_{X+1} - Ep_{X+2}) +$$
$$(\sum_{t=X+3}^{T} \beta^5 U_h (1 - Ep_{X+1}(1 - Ep_{X+2} - Ep_{X+3}) - Ep_{X+2}(1 - Ep_{X+3}) - 2Ep_{X+3})] \quad (4.16)$$

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

Proof of Prediction 7:

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 $j + 1$:

$$EV_{j+1} = -c(e_{j+1}) + Ep_{j+2} \sum_{t=j+2}^{T} \beta U_h + (1 - Ep_{j+2})[Ep_{j+3} \sum_{t=j+3}^{T} \beta^2 U_h + (1 - Ep_{j+3})\{Ep_{j+4} \sum_{t=j+4}^{T} \beta^3 U_h + (1 - Ep_{j+4})(Ep_{j+5} \sum_{t=j+5}^{T} \beta^4 U_h + (1 - Ep_{j+5}) Ep_{j+6} \sum_{t=j+6}^{T} \beta^5 U_h)\}] \quad (4.17)$$

and compare it to the expression for j:

$$EV_j = -c(e_j) + Ep_{j+1} \sum_{t=j+1}^{T} \beta U_h + (1 - Ep_{j+1})[Ep_{j+2} \sum_{t=j+2}^{T} \beta^2 U_h + (1 - Ep_{j+2})\{Ep_{j+3} \sum_{t=j+3}^{T} \beta^3 U_h + (1 - Ep_{j+3})(Ep_{j+4} \sum_{t=j+4}^{T} \beta^4 U_h + (1 - Ep_{j+4}) Ep_{j+5} \sum_{t=j+5}^{T} \beta^5 U_h)\}] \quad (4.18)$$

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.

Chapter 5

Conclusion to Part I

Part I of this thesis has attempted to learn from the Chinese experience of incentivising public servants through promotions, with a focus on teachers. Although promotions remain the primary determinant of wages for teachers in low income countries, promotions are not generally used to motivate teacher effort. The Gansu Survey of Children and Families (GSCF) provides an ideal data set with which to investigate this topic as it includes detailed information on teachers' history of promotions and annual performance evaluation scores, and in many cases, all the teachers in a school have been interviewed. It is the first to empirically test the use of promotions as an incentive device in the public sector.

There are two main questions that Part I of this thesis has attempted to answer: firstly, can promotions motivate effort in the public sector? And second, what influences a teacher's choice of effort? A model of promotions as an incentive device that treats all teachers as identical is presented in Chapter 3. This model provides a starting point to think about how promotions could elicit effort, as well as how average levels of effort can be expected to vary in promotion contests with different parameters. The model predicted that higher effort would lead to an increased probability of promotion. It also predicted that higher wage gaps would elicit higher effort, and that contests in which the average age of teachers was lower, or in which the proportion of female teachers was lower, would have higher effort incentives. It also predicted that very high and very low promotion rates would elicit low effort. Finally, the model predicted that larger contests would increase effort incentives for contests with promotion rates between 1/3 and 2/3, and would decrease incentives in contests with promotion rates outside this interval. An empirical model of promotions was then developed, which noted that effort is endogenous to promotion because although effort is required to earn a promotion, it is the promotion that motivates effort. As such, instrumental variables techniques were used to estimate the model, using the change in the log of wages as an instrument for effort. This was argued to be an informative instrument, as higher wage gaps were found to be associated with higher effort, and to be a valid instrument, as teachers can only access higher wages through effort; wages only affect promotions through effort.

In Chapter 4, the promotion incentives model is extended to the case of heterogeneous teachers competing for promotions, and this allowed us to look at the individual level of effort rather than average levels. Teachers with very high and very low skill were predicted to have very low effort incentives. The model also predicted how effort would vary based on the interaction between the teacher's skill and the number of teachers competing for a promotion; effort was predicted to be low for teachers in the extremes of the skill distribution when the promotion contest was larger. A multi-period model of teacher effort was also developed, which predicted that teacher effort will be zero five and more years before becoming eligible for promotion because of the way that performance evaluations feed into promotion decisions, that teachers will increase their effort in the five years prior to becoming eligible to apply for a rank promotion, and that teachers will decrease their effort if they have been passed over for promotion several times after they are eligible. An effort equation was then estimated, which regressed the evaluation scores (again as a proxy for effort) on the parameters of the promotion system including dummies for the number of years into a rank, the number of teachers, and a measure of ability.

From Chapter 3 we learned that teachers are working hard to earn promotions. We also learned that contests with higher wage gaps were found to elicit more effort. Furthermore, average levels of effort are higher for promotion contests with promotion rates closer to one-half. From Chapter 4 we found that individual effort was highest when the skill percentile of the teacher was closer to the promotion rate. Those in the top and in the bottom ten per cent of the ability distribution were generally found to have low incentives. Effort was zero five and more years before a teacher was eligible for promotion. Teachers increase effort in the years prior to becoming eligible to apply for rank promotions, and teachers that are repeatedly passed over for promotion (those that stay in a rank for many years) also have low effort, and they may be revising their estimate of ability downwards. The number of teachers competing also matters in the way predicted by the model, but the results are weak. One reason for this could be that the promotion system here is a bit of a mix between a tournament and a standard model. Teachers have to achieve some minimum requirements (standards) and then compete. Finally, I also find that after the highest promotion, effort drops significantly as expected, since there are no more promotions to work for, and older teachers slow down.

The findings in Part I suggest that promotions are a viable way to motivate public servants, including teachers. These findings also suggest several recommendations for the design of incentive systems in the public sector. The most important is that incentivising teachers solely based on performance pay may not be necessary. The use of bonuses amongst this sample of teachers is very low, however, the system of promotions seems to be incentive enough for effort in many cases. The fact that effort drops steadily in the highest ranks may make a case for using performance pay at these levels to keep motivation high, but at lower levels it may not be necessary. If bonuses were introduced at lower rank levels, one would need to be mindful of the effect that a bonus could have on a teacher's information about her skill. Receiving a bonus will increase the teacher's estimate of her relative skill percentile, and before a teacher is promoted this could cause a decline in effort. The findings on salary levels and on the number of teachers in a contest would also provide policy makers with an idea as to how to mould the salary structure to optimally motivate effort, by providing insight into how to split teachers into contests. In addition, the results shed insight onto how the timing of promotions could be optimally structured. For example, if it is the past five evaluation scores that matter, then waiting times of more than five years before a teacher is eligible to apply for the next rank promotion are not optimal.

The promotion system described here is very specific to teachers in China. However, similar promotion-based incentives are used for all public servants in China, and as such, this study can serve to inform the design of public sector incentive systems in general. Finally, these results also help explain the success of the Chinese economic reforms. The reforms in the government in particular were very much tied to bureaucratic performance and merit-based promotions. The system is quite sophisticated but is understudied, and can provide examples of good practice for other contexts.

There are a couple of limitations to this work. First, a longer history of data on evaluation scores of teachers would enrich the analysis by further ameliorating the selection problem that was noted. Furthermore, better data on teachers' ability would provide better proxies. Second, the fact that the evaluations can be based on different aspects of teaching merits further study, which is outside of the scope of this thesis. It may be the case that putting greater emphasis on certain components of the performance evaluation would elicit either more or less effort, or different types of effort among teachers. The way that teachers substitute effort between various tasks when the performance measure puts more weight on certain aspects of effort would be very informative. Such a study would need to adequately control for how the choice of performance evaluation system is determined, perhaps through a randomised experiment or exogenous policy change. Finally, in order to make further recommendations on incentive system design in the public sector, such studies should be carried out in other contexts and/or countries so that external validity can be discussed, and policy implications made clearer.

Part II

Teacher labour markets, social distance, and learning outcomes

Chapter 6

Teacher labour markets, social distance, and learning outcomes

6.1 Introduction to Part II

Inequalities in educational opportunities are characteristic of many low-income countries, and this hinders growth and development by limiting growth in human capital (Mankiw et al., 1992; Barro, 2000, 2001; Castelló and Doménech, 2002). These inequalities in education then also translate into inequalities in the labour market. In South Asia, inequalities exist not only in terms of wealth and income, but also between social groups. In particular, in both India and in Pakistan, caste is a primary determinant of both income and social inequality (Mohmand and Gazdar, 2007; Gazdar, 2007; Zacharias and Vakulabharanam, 2011). Understanding the role of caste-based inequalities in education and labour markets is critical to finding ways to reduce gaps, increase the stock of human capital, and promote growth in South Asia. This chapter aims to identify the role of caste in education in Pakistan.

Teacher labour markets and teacher assignment systems in public schools determine which students are taught by which teachers, and hence, the characteristics of teachers that students are exposed to. The question of whether, and which teacher characteristics matter for student learning has received extensive attention. It is important to examine learning outcomes, as these are what determine the skills that children will take to the labour market. Although teacher quality is considered integral to learning outcomes, there is no consensus on what makes a high quality teacher (Glewwe, 2002; Clotfelter et al., 2006; Glewwe and Kremer, 2006; Hanushek and Rivkin, 2006a). A review of several recent studies in both the education and economics literature found that very few observable teacher characteristics contribute to learning (Glewwe et al., 2011). In the United States, studies have shown that the distribution of teachers' characteristics can affect student outcomes and can generate inequalities in learning (Rivkin et al., 2005; Hanushek and Rivkin, 2006a; Jackson, 2009). In the United States, the black-white achievement gap has received substantial treatment in the literature, but elsewhere social distance has received less study. In particular, the issue of caste in South Asia is a recent contribution to the literature.

Previous work has shown that caste is an important determinant of enrolment decisions in both India (Dreze and Kingdon, 2001; Borooah and Iyer, 2005; Dostie and Jayaraman, 2006) and in Pakistan (Jacoby and Mansuri, 2011). In India, there is also growing evidence that caste affects learning outcomes. There are a variety of reasons caste could affect learning, and it is not clear a priori whether there would be positive or negative effects, or both. There are three aspects to this: child caste, teacher caste, and the interaction between child and teacher caste. Child caste could be important if a school consists predominantly of either high or low caste children, and the other caste-type feels as if they are outsiders (Akerlof and Kranton, 2010). Alternatively, there may be differing returns to education for high and low caste children (Munshi and Rosenzweig, 2006; Luke and Munshi, 2007), or drawing attention to caste differences could affect children's confidence (Hoff and Pandey, 2006). Teacher caste could matter if high caste teachers are just higher quality teachers. The interaction between child and teacher caste could be important if either the same or different caste-type teachers can serve as role models to children (Kingdon and Rawal, 2010), or may discriminate against some children through biased grading practices (Hanna and Linden, 2012) or in the way in which they percieve children. Finally, since patronage in Pakistan is very closely tied to caste, it could also play a role in the education system and may even affect learning. Kingdon (1998) finds that conditional on enrolment, low caste children in India perform just as well as high caste children on cognitive tests. Unfortunately, there are no empirical studies in Pakistan of the effects of caste on learning outcomes, or of the impact of teacher caste on learning of students of different caste-types.

This chapter aims to identify the effect of caste on children's learning outcomes. First, I look at whether learning outcomes of high and low caste children differ. I then investigate whether teacher caste affects children's learning outcomes overall, and finally, whether high and low caste teachers differentially affect the learning outcomes of high and low caste students. I use a detailed longitudinal study from Punjab, Pakistan, that tracks learning outcomes of third grade children over four years, and also follows up with them five years after the study, by which time they are nearing the end of their schooling cycle. The initial four rounds of the survey allow me to identify teacher caste effects (as well as the interaction between teacher and child caste) using a child fixed effects specification, which ameliorates many concerns faced by studies attempting to identify the causal effects of teacher characteristics on student outcomes. In this case, the effect of a high caste teacher is identified from the same child switching between high and low caste teachers.

I find that in the cross-section, low caste children have higher test scores than high caste children. However, within schools there are no significant differences. Estimating the child fixed effects specification, I find that overall, having a high caste teacher has no impact on children's test scores. However, this result masks heterogeneity between caste types. I find that low caste children, and in particular low caste boys, have significantly higher test scores when they are taught by high caste teachers rather than by low caste teachers. This result is robust to the inclusion of many controls, as well as to tests that address the possibility of non-random switching of teachers. So, it is the interaction between child and teacher caste that matters in Pakistan.

Next, I investigate six possible reasons that I find these results: children feeling like outsiders, returns to education, teacher quality, role model effects, discrimination, and patronage. I do not find that the differences in test scores of high and low caste children varies between schools that are predominantly composed of high or low caste children. Using data on the children's parents, I also find that returns to education are highest for low caste men. I do not find that high and low caste teachers differ in their observable or unobservable quality, and I also do not find evidence of role model effects or discrimination. The patterns I find show that low caste children work harder when taught by high caste teachers. Low caste students taught by high caste teachers spend almost an extra half hour per day doing homework. This may be because high caste teachers are able to help children gain access to employment and/or educational opportunities later on, thereby further raising the returns to education for these children. This conjecture is supported by findings that show that the positive effect of high caste teachers on low caste boys is present in public but not in private schools, and also in high caste dominant villages but not in low caste dominant villages. It is in these cases that high caste teachers would have a greater ability to distribute such benefits. In addition, I find evidence that children who were taught by at least one high caste teacher are more likely to still be enrolled in school, and have completed more years of schooling when followed up five years after the main survey. Although this mechanism cannot be directly proven in this data, these results provide a starting point to thinking about the possible mechanisms of the caste effects that I identify.

The next section will describe the history and meaning of caste, as well as the context of education in Pakistan. Section 6.3 will discuss the ways in which caste could potentially affect learning outcomes. Section 6.4 will describe the data used in this study. The empirical model to be tested is described in Section 6.5, followed by the results from estimating the empirical model in Section 6.6. Section 6.7 will conclude Part II.

6.2 Caste and Education in Pakistan

The notion of caste in Pakistan is very different from that of India. Officially, caste does not exist in Pakistan, because caste is not recognised in Islam. However, people in rural Punjab are very cognizant of their caste group and that of others, and are also aware of the social conventions surrounding caste that have endured despite the official stance of the government. Several authors have noted its importance in relation to land ownership, employment, political disempowerment, health and education, and access to services (Gazdar, 2007; Mohmand and Gazdar, 2007; Jacoby and Mansuri, 2011).

In Pakistan, caste is a social, rather than a religious institution, and is interchangeable with clan, kin, and tribe (*zaat/biraderi*). It is also far more variable, and more difficult to define (Ibbetson, 1974). Castes are based on 'lineage endogamy', with patrilineal cousin marriage forming the basis for the kin networks. In addition, within each broad caste group there are sub-caste groups, and as is characteristic of endogamous kinship systems, men generally only marry women from the same sub-caste group within the broad caste group (Mohmand and Gazdar, 2007). These sub-castes are often associated with the region from which that particular sub-group originated.

Castes in Punjab are ranked according to historical land ownership and occupation. The hierarchy consists of land owners (*zamindars*), tenants/cultivators, service/artisan professions (such as weavers, iron smiths etc.), and finally outcaste groups (menial labour professions) (Ibbetson, 1974). Land owners and tenants are considered high caste, and service and menial labour professions are considered low caste. Within both the high and low caste categories, there is a range, with some castes considered to be of higher status than others. For example, the group '*Syed*' carries the highest status across Punjab, and other groups such as '*Jat*' and '*Rajput*' are considered of lower status than *Syeds*, but both are considered high caste, and which one is considered of higher status may differ by region and even by village. Here, I do not distinguish between the order of high and low caste groups, I distinguish pnly between those considered low caste and those not considered low caste. This is because a *Rajput* person will not treat a *Jat* person differently, but they may treat a low caste individual differently. Consequently, from now on any reference to 'caste' can be thought of as 'caste-type' (high or low).

People in a particular caste group tend to inherit the occupation of their predecessors. This is now becoming less prevalent as low caste groups are able to take up other professions, but still remains present. In Punjab, caste is closely connected to social status and dignity, and there is a large degree of social stereotyping, which is derogatory to lower caste groups. There are historical rules of social interaction that govern relationships between groups. For example, high caste groups will not eat with lower caste groups (Ibbetson, 1974). Furthermore, caste differences are much more salient in the north of

Punjab than in the south.

Groups assume a corporate structure that performs political and economic functions (Mohmand and Gazdar, 2007). Several caste groups generally reside in the same village. In each village, each group vests authority in a leader who is responsible for organising economic exchanges and support mechanisms between members, and plays a key role in marriage decisions. However, members of lower caste groups often defer to leaders of higher caste groups for dispute resolution and economic support, as they are the resident landlords and are key sources of credit and patronage for their tenants and labourers (Mohmand and Gazdar, 2007). Vyborny and Chaudhury (2012) have also found that in Punjab, caste is important for patronage networks that can provide individuals and families with access to assistance in times of need as well as preferential access to services. Since caste is based on endogamous marriage practices, it is extremely salient in the marriage market, with very few people marrying outside of their sub-caste group.

It is important to note that caste does not map perfectly onto wealth. It is possible to have more and less wealthy people in the same caste group, but they will have similar social standing in the village. Social mobility for low caste groups is predominantly through migration to urban areas, as well as through education, which provides access to formal sector (and public sector in particular) employment. Labour market opportunities are often mediated through social networks (Mohmand and Gazdar, 2007).

The education sector in Pakistan has grown substantially recently. From 2001 to 2005 enrolments jumped ten percentage points, and this was made possible largely by the rapid expansion in private provision of education, with one in three children studying in a private school by the end of 2005 (Andrabi et al., 2007). The Learning and Educational Attainment in Punjab Schools (LEAPS) study found that private schools provide better education (in terms of test score outcomes) than public schools and at a lower cost, but that there is much variation in the quality of both public and private schools. In each village in Punjab, there is an average of eight schools, which usually includes boys' and girls' public primary and middle schools, as well as a few private schools, which are generally mixed gender. As the next section will describe, caste does play a role in the education sector both in other South-Asian countries, and in Pakistan.

6.3 Caste and Learning Outcomes

Caste-based stigma and discrimination is very much present in schools throughout South Asia, and there are many possible ways in which caste could affect children's learning outcomes. Three overall groups of channels can be envisaged: child caste, teacher caste, and the interaction between child and teacher caste. I will consider each group in turn, beginning with children's caste.

One potential mechanism is children's confidence. Hoff and Pandey (2006) conduct two experiments in India and consider the issue of 'stereotype threat' whereby people act as if the stereotype about their 'type' is true. For example, framing a test as being about athletic ability rather than athletic knowledge can lead black students in the United States to perform better on the test. In Hoff and Pandey's first experiment, they have sixth and seventh graders in India solve mazes, and they experimentally vary whether children's caste is announced publicly at the beginning of the test. There are three groups in the experiment: a control group where no caste is announced, a treatment in which there is a mix of high and low caste children present and caste is announced, and a treatment in which there are only high/low caste children present and caste is announced. They find that announcing students' caste status affects their performance. In the treatment in which there are high and low caste children present and caste is announced, low caste children solve significantly fewer mazes than low caste children in the control group. The authors attribute this to lower confidence amongst low caste children when everyone is made aware that that they are of low caste. People expect them to perform less well and so they do. In the second treatment in which there are only high/low caste children present and caste is announced, low caste children fare the same as those in the control group but high caste children solve significantly fewer mazes than high caste children in the control group. The authors believe that this is because in the mixed groups, high caste children may want to distinguish themselves from low caste children and so want to perform better. In a second paper, the same experiment is conducted with tournament incentives instead of piece rate incentives, and the authors find that both high and low caste children solve more mazes under tournament incentives, but that revealing caste eliminates the positive effect of the tournament (Hoff and Pandey, 2011).

In the second experiment, the authors ask the children to bet on their success by asking children whether or not to accept a gamble of how many mazes they can solve. Here, they retained the same three groups but varied whether there was scope for judgment by a 'grader'. When there was scope for judgment, low caste children in the group where there were only low caste children present had much less confidence in their success. The authors explain that this is the result of low caste children believing that because they were segregated, the person in charge was more likely to discriminate (Hoff and Pandey, 2006). Outside of such experiments, it would be very difficult to identify whether confidence is driving caste effects on learning outcomes.

This second experiment also points to a potential effect of the distribution of caste groups in a school. Learning outcomes could differ substantially for high and low caste children depending on whether the school is predominantly composed of high or low caste children. Apart from peer effects, the likelihood of discrimination and stereotyping could be different depending on whether your caste-type is plentiful in the school. The identity economics literature would argue that the key factor is whether the school creates an inclusive atmosphere (Akerlof and Kranton, 2010). Each school has its own ideals, shaped and enforced by the principal and teachers, which includes norms on behaviour. These norms are often those of the predominant social group. The children who feel they do not fit these norms feel like 'outsiders' and tend to misbehave and not to perform well academically(Akerlof and Kranton, 2010). Successful schools create an identity that is inclusive for all children, and as a result, all children feel like 'insiders'. If a school is high caste dominant, low caste children may feel more like 'outsiders' and may have lower learning outcomes. This mechanism can be tested by comparing differences between high and low caste children in schools comprised predominantly of high or low caste students.

Another mechanism that could affect learning outcomes is the returns to education. If high and low caste groups have differential returns to education, then we can expect that they would invest differently in human capital. Some work has been done regarding the returns to education in Pakistan, but this work has focused on gender differentials (see Aslam et al., 2008; Aslam, 2009; Aslam and Kingdon, 2012). Munshi and Rosenzweig

(2006) study investments in education of children in South India in a unique setting in which high and low caste households live together on tea estates and do the same jobs for the same wages. Exogenously increasing female income, they find that low caste families invest more in their children's education than high caste families, and that low caste children have significantly higher educational attainment than high caste children. They show that this is because the returns to education are higher for low caste households than for high caste households. Low caste households have fewer opportunities in the villages from which they originally come. Cutting ties with their traditional homes and instead investing in human capital of their children on the tea estate will bring them a higher return. For high caste households, this is not the case. In a companion paper, Luke and Munshi (2007) find the same pattern for investments in the health of children. Low caste families may invest more in human capital than high caste families because the returns to doing so are higher for them. Estimating earnings functions for high and low caste groups would provide insight into whether returns to education differ for high and low caste children in Punjab.

It is also possible that teacher caste could affect learning outcomes of students. In the simplest case, it could be that high caste teachers are just higher quality teachers than low caste teachers. Because of the history of exclusion of low caste groups in Pakistan and their resulting disadvantage in education and the labour market, high caste groups could be more skilled and better trained. For teachers, this could translate into higher ability to teach students. Alternatively, because of the obstacles facing them, it could be the case that the low caste individuals that do select into the teaching profession are of very high quality, and so low caste teachers could be of equal or better quality than high caste teachers. Apart from better pedagogy, having a high quality teacher may also be motivating for students, and this could also result in higher test score outcomes. There could be observed and unobserved differences in quality. Observable differences can be controlled for in a regression framework. If high caste teachers improve the learning outcomes of all children, comparing children matched with high versus low caste teachers would tease out this effect. If high caste teachers do not equally affect high and low caste children, this would be evidence that it is not teacher quality that matters, but something else that affects only one group, perhaps something unobservable to the econometrician. In this case, unobserved differences in teacher quality can be differenced out by comparing high and low caste children in the same class.

Another possibility is that it is the interaction of child and teacher caste that affects learning outcomes. Some studies in the United States and other high-income countries have found that children perform better when taught by teachers of the same race (Dee, 2005; Hanushek and Rivkin, 2006a; Lindahl, 2007). These studies cite two possible mechanisms for how the interaction between teachers' and students' race could affect learning outcomes. The first is a role model effect. Minority teachers may inspire minority students by encouraging them to update their prior beliefs about their educational possibilities, and they may also feel more comfortable and more secure when taught by a person of their own race (Dee, 2005). Kingdon and Rawal (2010) find that the smaller the social distance between teachers and students in India, the better the learning outcomes of children. Students taught by teachers of the same caste status (low or high), gender, and religion, performed better in cognitive tests. They argue that this is because the teachers serve as role models to the children. However, their estimation strategy uses between-subject variation for identification of teacher effects, which relies on the absence of subject-specific teacher unobservables that are correlated with caste, religion or gender. It is easy to imagine that this may not be the case, especially for gender, where male teachers for example, may have less confidence in the mathematics ability of female students. According to these studies, if teachers are acting as role models, we would expect that children of a particular caste-type would perform better when taught by teachers of the same caste-type. Alternatively, it may be the case that teachers of a different caste-type could serve as role models to children. In this case, we would expect to see that children matched with teachers of the opposite caste-type have higher learning outcomes. However, finding this would not necessarily mean that children see teachers as role models. Data on whether a child wants to emulate his/her teacher by becoming a teacher later on, could also serve as a proxy for whether the child sees the teacher as a role model. These are passive effects of the interaction between teacher and child caste. Active mechanisms could also exist. Teachers may allocate more time and interact more with students who share the same racial or social background (Dee, 2005).

Another active mechanism that may differ by the interaction of teacher and student caste is discrimination. Discrimination could manifest itself in several ways. In Pakistan, Jacoby and Mansuri (2011) find that caste plays a role in the enrolment decisions of households, and availability of schools within settlements is crucial. Children normally attend the closest school to their house that is grade and gender appropriate. However, they find that high caste girls are less likely to enrol in school if they have to cross a settlement boundary to get to school regardless of the distance, and that low caste boys and girls are less likely to enrol if the closest school is one in which high caste groups are dominant, due to social barriers and potential maltreatment from teachers. An example from the qualitative data contained in the paper notes that teachers have high caste children sit inside classrooms while low caste children sit outside. Another form of discrimination could be biased grading practices. Hanna and Linden (2012) conducted an experiment in India in which they randomly assigned child characteristics including age, grade, and caste to exams, which were then marked by a sample of teachers. They found that low caste children were scored between 0.03 and 0.09 standard deviations below high caste children, and that it was actually low caste teachers that were driving these results. Comparing test scores from blind grading and non-blind grading, Botelho et al. (2010) also find evidence for discrimination in grading amongst black children in Brazil. Lower grades will affect both progression through the schooling system and then labour market outcomes, and also confidence. If teachers do grade in a biased manner, an intelligent child could have low learning outcomes even on tests not graded by their teacher if this leads them to believe they are not intelligent. In the case of Pakistan, it is not clear a priori who might be discriminating against whom. An experiment would be the ideal way to test for this type of discrimination. Discrimination could also manifest itself in the way in which teachers percieve a child. If a teacher dislikes a particular caste group, he/she may have a lower perception of the child's ability, and this could be a self-fulfilling prophecy for the child if the teacher either communicates this directly or indirectly, or if as a result, the teacher treats a child differently. To test this, one could compare teacher perceptions of children with children's actual ability.

Finally, there is another mechanism specific to the Pakistani context through which the interaction of child and teacher caste could affect learning outcomes. As noted above by Mohmand and Gazdar (2007) and by Vyborny and Chaudhury (2012), labour market opportunities and social benefits are distributed in society by high caste individuals as favours to low caste individuals. Studies have pointed to patronage in the public sector for public goods in general, food aid, employment in the public sector, and government benefits (including in Pakistan) (Besley et al., 2004; Caeyers and Dercon, 2012; Fafchamps and Labonne, 2012; Vyborny and Chaudhury, 2012). If high caste teachers have the ability to distribute employment or educational benefits through patronage networks, then this might differentially affect the behaviour of high and low caste children in the classroom. High caste children would not necessarily need a high caste teacher's help but a low caste child would. For example, teachers may be able to help them gain formal sector (and in particular, public sector) employment after they finish school. They may also be able to help them get access to government scholarships, or to get admission into a good high school. In order to form a social network link with the teacher, the child's parents may become more involved in the child's schooling by perhaps meeting with the teacher more often, or the child may work hard in order to impress the teacher. These efforts could lead to higher learning outcomes. Evidence of these things occurring would lend support to the patronage mechanism. Even though the child's entry into the labour market could be many years away, high caste teachers could still matter. The child and teacher most likely live in the same village, and the child and his/her parents could easily keep in touch with the teacher even if they do not since the teacher would come to the village for work every day. Public school teachers, and teachers in high caste dominant villages would have a greater ability to distribute benefits to students, and this can be tested by comparing public and private school teachers as well as high and low caste dominant villages. This can also be tested by looking at eventual outcomes of children and seeing whether low caste children that were taught by high caste teachers have better jobs or schooling completion.

All of these possible mechanisms may be at work here and this chapter will be able to test for six of them.

6.4 The LEAPS Survey and the Sample

The data come from the Learning and Education Attainment in Punjab Schools (LEAPS) survey, which is an extensive longitudinal data set collected from 2003-2006, with a follow up survey conducted in 2011. There are very few longitudinal studies in low-income countries that collect test score data as well as rich household level data that allows for the inclusion of many controls. 112 rural villages in the three districts of Attock (in the North of Punjab), Faisalabad (centre) and Rahim Yar Khan (south Punjab) were surveyed. There were two parts to the survey: a school survey, and a household survey.

The school survey tested 12,000 children in grade 3 in 2003 and attempted to follow them through grades 4-6. Tests were administered in mathematics, English and Urdu (the official language in Pakistan). These tests were meant to be general, and a similar test was administered each year. The tests were scored using Item Response Theory (IRT) in order to make the tests comparable over time, and to measure underlying ability. IRT does this by estimating different weights to correctly answered questions depending on the difficulty of the question. The knowledge scores generated by this process represent a student's underlying knowledge in a particular subject. The knowledge scores are then converted into standard deviations from the mean for ease of interpretation. Teacher, head teacher and school surveys, and teacher tests in the same three subjects were also administered along with the school survey.

The second component was a household survey, administered to a smaller set of 1,800 households that contained a child in grades 1 to 5. Many of the children in the household survey that were in grade 3 in 2003 were therefore also tested (an overlap of almost 1,200 children were administered both surveys and tests). It was in this smaller household survey that household caste was recorded, and this group of 1,200 children in the household survey that were also tested in the school survey is the sample I use here.

Both teachers and students are grouped into high and low caste groups, based on the caste reported by the household or teacher [1]. High caste groups are those historically

[1] Cassan (2012) has noted that caste groups in Pakistan have manipulated their identity in the past. However, in order for this type of manipulation to affect the results, it would need to be the case that households mis-report their caste in response to the switching of their child's teacher over the years of the survey. This is not the case in the data.

associated with land ownership, and low castes are those historically associated with service professions or castes that were considered outcasts. Whether caste group names are associated with land ownership or service professions was determined by consulting two sources. The first was H. A. Rose's *A Glossary of the Tribes and Castes of Punjab and North-West Frontier Province*, which is the most extensive work detailing all castes in Punjab (Ibbetson and Maclagan, 1911). The second source was the District Gazetteers for the three districts in the sample, which were compiled by the British during the colonial era and include detailed descriptions of the occupations, customs, and land ownership of castes and sub-castes ((India), 1932, 1996; Dīn, 2001). Table 6.20 in the appendix lists the caste groups reported in the survey and whether they are classified as high or low caste.

Table 6.1 provides descriptive statistics of the 1,194 children in the sample, who are observed for between one and four years from 2003-2006 (2,642 child-year observations in total). High caste children make up 75% of the sample [2], and there are also more boys than girls (since these are all enrolled children by grade 3). High caste children tend to be enrolled in private school more than low caste children, and they also tended to still be enrolled in school by the fifth round of the survey (2011), and had completed more schooling. High caste households are much more likely to own land and have more assets, and parents are likely to be more educated.

Some interesting patterns emerge with regards to test scores. Boys tend to perform better in math compared to girls, and girls tend to outperform boys in English and in Urdu. Low caste children also tend to have higher test scores than high caste children. This may be because high caste children generally help their parents with the family business or farm, and can also use their families' connections for future opportunities. For low caste children, education is the only means for social mobility. This is consistent with the findings of Munshi and Rosenzweig (2006) and Luke and Munshi (2007) in India described earlier: that that the perceived returns to education are higher for low caste households. In addition, the higher test scores of low caste children likely also reflect the

[2] This is in line with other studies, both in Pakistan and in India. In Pakistan, Jacoby and Mansuri (2011) also find that 75% of their sample of children is high caste. In India, Kingdon and Rawal (2010) find that 76% of students are 'high caste' (not Scheduled Caste or Scheduled Tribe).

Table 6.1: Descriptive statistics of children

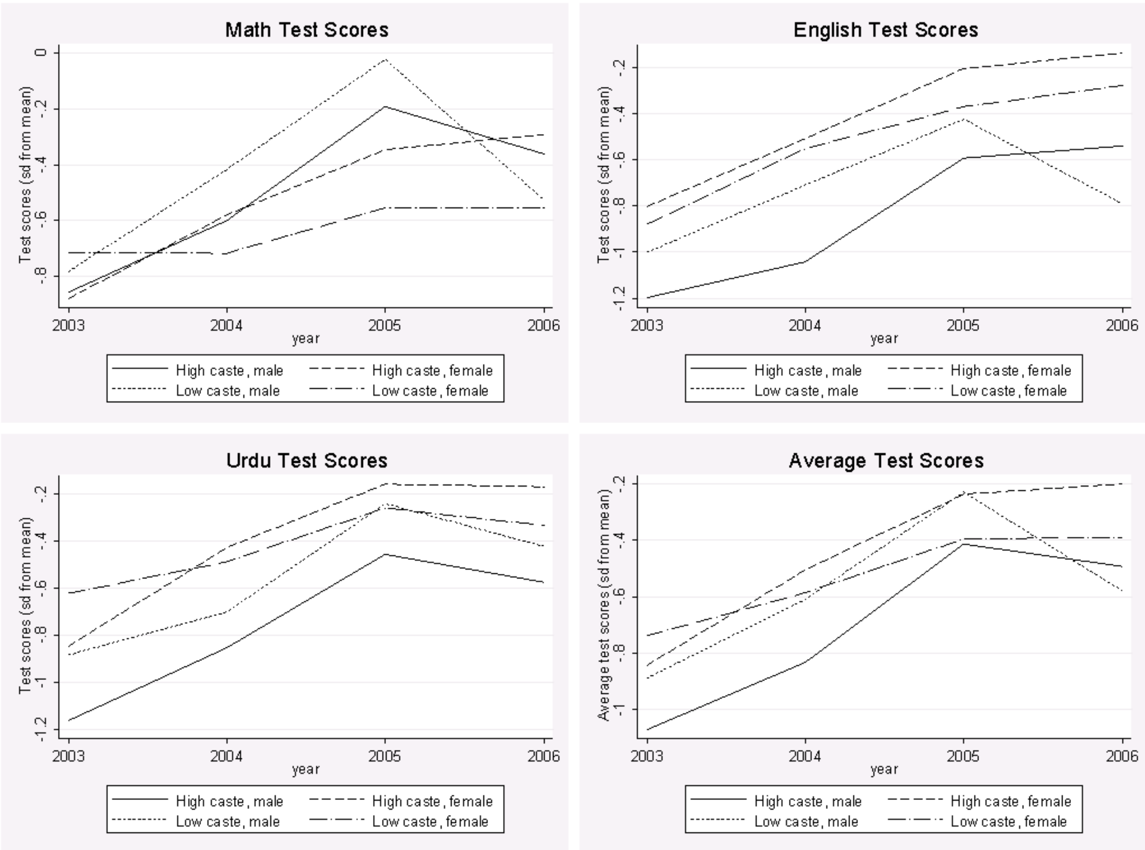
	High caste				**Low caste**			
Children	**Boys**		**Girls**		**Boys**		**Girls**	
Number of children*time	1,068		932		393		249	
Proportion of children	0.404		0.353		0.149		0.094	
Characteristics	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd
Average age	10.15	1.67	10.30**	1.65	10.12	1.64	10.45**	1.65
Grade	3.91	0.82	3.89	0.82	3.92	0.79	3.82	0.80
Education	Mean	Sd	Mean	Sd	Mean	Sd	Mean	Sd
Enrolled in private school	0.25	0.44	0.29	0.45	0.21	0.41	0.21	0.41
Enrolled in school (2011)	0.49	0.50	0.45	0.50	0.41	0.49	0.26**	0.44
Highest grade (2011)	7.89	1.97	8.15*	2.06	7.84	2.21	7.84	2.26
Households	Mean		Sd		Mean		Sd	
Household size	8.10		3.42		8.38*		2.96	
Owns land	0.56		0.50		0.35***		0.48	
Monthly educ. expend.	235		160		215***		147	
Asset index	-0.09		2.29		-0.25		2.21	
Mother uneducated	0.61		0.49		0.74***		0.44	
Father uneducated	0.37		0.48		0.51***		0.50	

Asterisks denote significant differences between boys and girls, and between high and low caste households.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

fact that these are the low caste children that have selected into enroling in school, so they may be of higher ability than low caste children on average. Figure 6.1 graphs the test scores (measured in standard deviations from the mean) of children in Math, English and Urdu over the four years of the main survey, as well as the average of the three test scores, which is the outcome that will be used here.

Table 6.2 contains descriptive statistics on teachers in the sample by caste type, as well as average test scores of high and low caste children that are taught by them. Most teachers are high caste. High caste teachers tend to be female, teaching in the same village that they are originally from, and more experienced. However, they tend to have lower levels of education than low caste teachers. Test scores of high and low caste teachers do not differ in any of the three subjects, nor are there large differences in comparing absenteeism or time use for high and low caste teachers. The test scores of low caste children are always higher than those of high caste children, regardless of whether they are taught by a high or low caste teacher. However, low caste children have higher test scores when taught by low caste teachers. If high caste teachers are not considerably

Figure 6.1: Children's Test Scores



different from low caste teachers in observable characteristics, but in the cross section low caste children appear to be performing better when taught by low caste teachers, this provides a justification to study the potentially differential effects of high and low caste teachers on high and low caste children in more detail.

Table 6.2: Descriptive statistics of teachers

	High caste		Low caste		All teachers	
Number of teachers*time	2,357		285		2,642	
Proportion of teachers	0.892		0.108		1.00	
	Mean	Sd	Mean	Sd	Mean	Sd
Age	34.59	9.49	35.49	7.73	34.68	9.32
Proportion female	0.55	0.5	0.37	0.48	0.53	0.5
Originally from same village	0.39	0.49	0.35	0.48	0.38	0.49
Proportion married	0.67	0.47	0.73	0.45	0.68	0.47
Teaching experience	6.69	18.18	3.96	10.9	6.4	17.56
Education: matric	0.4	0.49	0.24	0.43	0.38	0.49
Education: FA	0.28	0.45	0.26	0.44	0.28	0.45
Education: BA	0.25	0.43	0.45	0.5	0.27	0.44
Education: MA	0.07	0.25	0.06	0.23	0.07	0.25
Any training	0.8	0.4	0.88	0.32	0.81	0.4
Teaching in a private school	0.26	0.44	0.24	0.43	0.25	0.44
Math test	2.88	1.01	2.83	1.01	2.88	1.01
English test	2.38	0.93	2.59	0.85	2.4	0.92
Urdu test	2.68	0.85	2.65	0.76	2.67	0.84
Minutes/week grading	32.03	36.17	34.72	34.87	32.33	36.03
Minutes/week preparation	32.27	40.45	30.08	37.14	32.03	40.09
Minutes/week private tutoring	12.38	45.03	12.37	52.52	12.38	45.93
Proportion mother uneducated	0.3	0.46	0.32	0.47	0.3	0.46
Proportion father uneducated	0.69	0.46	0.65	0.48	0.69	0.46
Absences last month	1.98	2.56	2.08	3.12	1.99	2.63
Math test scores of high caste kids	-0.55	1.1	-0.56	1.16	-0.55	1.11
English test scores of high caste kids	-0.71	1.18	-0.59	1.12	-0.7	1.18
Urdu test scores of high caste kids	-0.64	1.15	-0.63	1.25	-0.64	1.16
Average test scores of high caste kids	-0.63	1.01	-0.6	1.03	-0.63	1.01
Math test scores of low caste kids	-0.5	1.03	-0.44	1.01	-0.49	1.02
English test scores of low caste kids	-0.68	1.07	-0.49	0.98	-0.65	1.06
Urdu test scores of low caste kids	-0.56	1.02	-0.34	1.18	-0.52	1.06
Average test scores of low caste kids	-0.58	0.9	-0.42	0.93	-0.55	0.9

Table 6.3 provides summary data on matches and switches between caste types of children and teachers. Panel A gives the number of male/female children of high/low caste-type matched with male/female teachers of high/low caste-type. Panel B provides the number of children who switch from a teacher of a certain caste/gender type to a

teacher of either the same or a different caste/gender type. Panel C is simply Panel B in proportions. Public schools are segregated by gender, so gender matches are very common. If a child starts with a high caste teacher, he/she tends to stay with a high caste teacher. One point to note is that the number of switches of caste types is very low (only 120). Only 3.2% of total switches between all teachers are from a high caste teacher to a low caste teacher. Switches from low caste teachers to high caste teachers are also relatively rare; only 5% of the total switches between all teachers in the sample. Table 6.4 provides the proportions of high and low caste students and teachers observed in the data in each of the four years.

Table 6.3: Matches and switches between teachers and students

Panel A: Matches			Teacher			
			High caste		Low caste	
			Male	Female	Male	Female
Child	High caste	Male	765	216	79	8
		Female	27	823	21	61
	Low caste	Male	254	59	76	4
		Female	11	202	3	33
Panel B: # switches			To Teacher			
			High caste		Low caste	
			Male	Female	Male	Female
From Teacher	High caste	Male	525	11	23	0
		Female	35	672	6	17
	Low caste	Male	47	5	54	0
		Female	5	17	0	40
Panel C: Proportion of switches			To Teacher			
			High caste		Low caste	
			Male	Female	Male	Female
From Teacher	High caste	Male	0.36	0.008	0.016	0
		Female	0.024	0.461	0.004	0.012
	Low caste	Male	0.032	0.003	0.037	0
		Female	0.003	0.012	0	0.027

An obvious concern would be non-random matching of teachers and students based on caste. In order to check whether the data resemble random matching of students and teachers, I perform two sets of tests based on Monte Carlo simulations, the first within villages and the second within schools. Within villages, I randomly assign teachers to classes in a particular year [3]. This takes the class structure within schools as given and

[3] I omit the fourth year of data because very few teachers were surveyed in the fourth round. Since the

Table 6.4: Proportions of high and low caste teachers and students, by year

Matches					
2003	High caste	Low caste	**2004**	High caste	Low caste
Teachers	0.87	0.13	Teachers	0.90	0.10
Children	0.77	0.23	Children	0.74	0.26
2005	High caste	Low caste	**2006**	High caste	Low caste
Teachers	0.89	0.11	Teachers	0.98	0.02
Children	0.76	0.24	Children	0.77	0.23

the number of classes within villages as given [4]. I can then calculate the frequencies of each of the four types of caste interactions (teacher high caste, child high caste; teacher high caste, child low caste; teacher low caste, child high caste; teacher low caste, child low caste) in each village that is actually observed in the data and compare the observed matches to the matches from the simulated data in two ways. I first use the Chi-squared goodness of fit test for each village-year. This generates 314 Chi-squared test statistics that I use to test the hypothesis that in a particular village and year, the distribution of the observed frequencies of matches is not too different from the distribution of the frequencies in the simulated data. Appendix Table 6.21 shows the means and standard deviations of the four categories of caste interactions with the observed and simulated data. In only 1.7% of the village-year cases (four village-years) is the Chi-squared p-value less than 0.05 [5], suggesting that the data do resemble random matches of teachers and students. Figure 6.2 plots the distribution of the p-values from this test. It is not the case that the test is only just being rejected. The p-value is one in most village-years and there is no concentration of cases that only just pass the test. In the second test, I use the simulated matches to construct 95% confidence intervals for each of the four match types in each village and year. Table 6.5 reports the proportions of village-year cases for which the observed frequency in each category falls within the 95% confidence interval. This tests points to slightly less random looking data, especially for high caste children. 31% of village-years (97) do not fall within the confidence intervals in at least one category.

children were by then in middle school, only children who remained in primary school had their teachers interviewed in the fourth round. Random matching is performed 1,000 times.

[4] This is done because we cannot have third and sixth grade children in the same class, for example. In addition, the number of children in a class is then held constant.

[5] The average p-value is 0.8931 and the lowest p-value is 0.0032.

Figure 6.2: P-values from Chi-squared test within villages

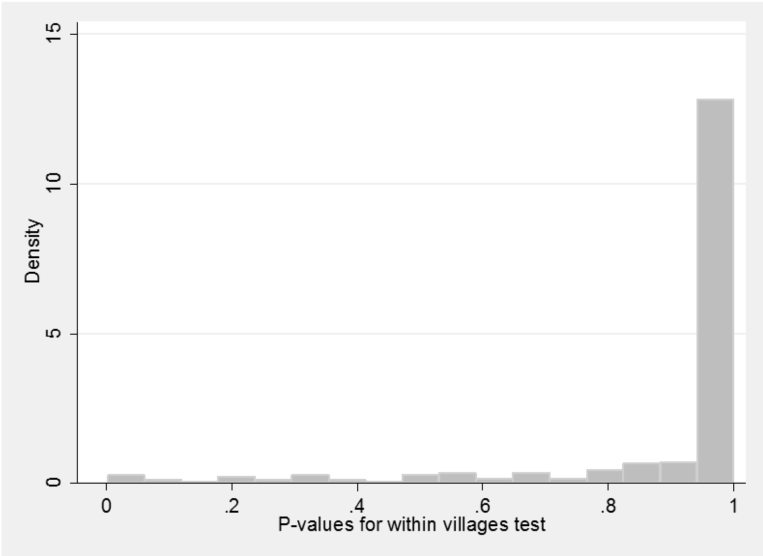
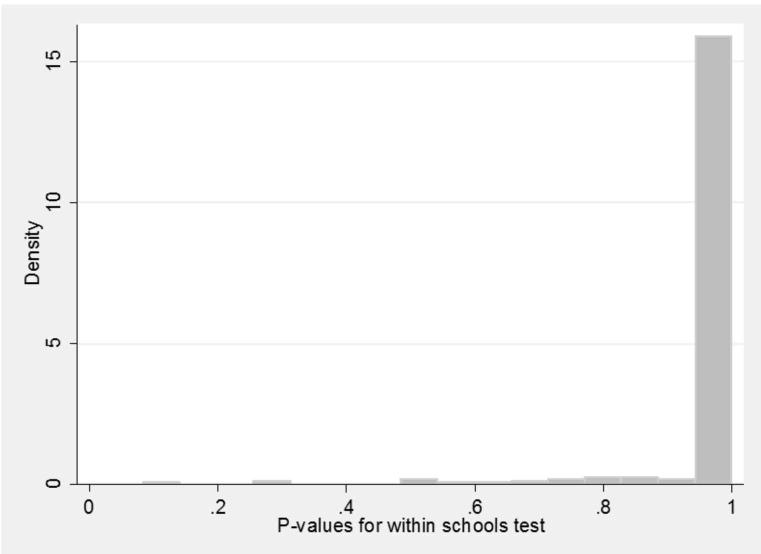


Table 6.5: Proportion of villages and schools that fall within the 95% confidence interval

Village Level	
Category	**Proportion**
Teacher high caste, Child high caste	72.93%
Teacher high caste, Child low caste	78.66%
Teacher low caste, Child high caste	73.25%
Teacher low caste, Child low caste	87.58%
School Level	
Teacher high caste, Child high caste	86.47%
Teacher high caste, Child low caste	92.48%
Teacher low caste, Child high caste	86.84%
Teacher low caste, Child low caste	96.62%

Within schools, I follow the same procedures but pool all years and randomly assign teachers to classes within the school [6]. I again calculate the observed frequencies of each of the four types of caste interactions and compare the observed matches to the matches from the simulated data. The Chi-squared goodness of fit test generates 266 Chi-squared test statistics that I use to test the hypothesis that in a particular school, the observed frequencies of matches are not too different from the frequencies in the simulated data. For all 266 schools, the Chi-squared p-value is above 0.05 [7], suggesting that the data does resemble random matches of teachers and students within schools. Figure 6.3 plots the distribution of p-values from this test, and here as well, it is not the case that schools are only just passing the test. The bottom panel of Table 6.5 reports the proportions of schools for which the observed frequency in each category falls within the 95% confidence interval. For all four categories, this proportion is above 85%, and for low caste children matched to both high and low caste teachers, the proportion of schools passing the test is above 90%. 41 schools (15%) fail the test in at least one category. These results show that although students were not randomly allocated to teachers in the sense of an experiment, the data are pseudo-random at least within schools.

Figure 6.3: P-values from Chi-squared test within schools



[6] I omit schools for which I have data on only one teacher. This leaves 266 schools within which children and teachers are randomly assigned to one another. The procedure amounts to asking, 'what if this teacher taught class x this year instead of last year'. This random matching is performed 1,000 times.

[7] The smallest p-value is 0.0834, and the average p-value is 0.983.

6.5 Testing for Caste Effects

In order to test the way in which child caste, teacher caste, and teacher-student caste interactions affect test score outcomes, I estimate an education production function of the form:

$$\begin{aligned} T_{ijst} = \beta_1 \boldsymbol{D_{ijst}} + \beta_2 \boldsymbol{XF_i} + \beta_3 \boldsymbol{XC_{it}} + \beta_4 \boldsymbol{XF_j} + \beta_5 \boldsymbol{XC_{jt}} + \beta_6 \boldsymbol{XF_s} + \\ \beta_7 \boldsymbol{XC_{st}} + t + \alpha_i + \alpha_j + \alpha_s + \epsilon_{ijts} \end{aligned} \tag{6.1}$$

where T_{ijst} is the test score of child i, taught by teacher j, in school s at time t. $\boldsymbol{D_{ijst}}$ represents the dummy variable(s) that will be included for child caste, or teacher caste, or for the interaction between teacher and student caste in the various specifications. $\boldsymbol{XF_i}$, $\boldsymbol{XF_j}$, and $\boldsymbol{XF_s}$ are fixed child, teacher and school observables, respectively. $\boldsymbol{XC_{it}}$, $\boldsymbol{XC_{jt}}$, and $\boldsymbol{XC_{st}}$ are time-variant child, teacher and school observables, respectively. t are time dummies, α_i, α_j and α_s are the child-specific, teacher-specific and school-specific time-invariant unobserved components, respectively, and ϵ_{ijts} is the error term, which allows clustering at the individual child level.

Non-random matching of children and teachers to schools, and non-random matching of children and teachers to one another within schools, are the main threats to the validity of OLS estimation of equation (6.1). To address these issues, I employ a number of techniques to control for non-random sorting. As the Monte Carlo tests showed, sorting of teachers and children to one another based on caste is quite random within schools, but not always across schools. So at the very least, controlling for school fixed effects is important. As such, I estimate (6.1) first with school fixed effects, and then with child fixed effects, since there are multiple years of data available for each child.

With school fixed effects, the regression relates the difference between the child's test score in a particular year and the average of all children's test scores in all years at a particular school, to deviations of the child, teacher or school characteristics in one year from their average in all years at the same school:

$$(T_{ijst} - \overline{T}_{i,s}) = \beta_1(\boldsymbol{D_{ijst}} - \overline{\boldsymbol{D}}_s) + \beta_2(\boldsymbol{XF_i} - \overline{\boldsymbol{XF}}_{i,s}) + \beta_3(\boldsymbol{XC_{it}} - \overline{\boldsymbol{XC}}_{i,s}) + \beta_4(\boldsymbol{XF_j} - \overline{\boldsymbol{XF}}_{j,s})$$
$$+ \beta_5(\boldsymbol{XC_{jt}} - \overline{\boldsymbol{XC}}_{j,s}) + t + (\alpha_i - \overline{\alpha}_{i,s}) + (\alpha_j - \overline{\alpha}_{j,s}) + (\epsilon_{ijst} - \overline{\epsilon}_s) \tag{6.2}$$

Any time-invariant school-level observed ($\boldsymbol{XF_s}$) and unobserved (α_s) factors are swept away since they are common to all students and do not change over time. Sorting of children and teachers to schools now will not bias the results. However, time-invariant unobserved child and teacher factors (α_i and α_j) remain as deviations from the average at the school. Any sorting of children to teachers within schools due to unobserved time-invariant factors that are correlated with caste, will bias β_1. The results of the Monte Carlo simulation-based tests in the previous section suggest that caste-based matching of teachers and students within schools is not occurring, but I include measures to correct for it anyway.

One potential difficulty is the existence of multiple classes within each grade. If schools organise either advanced or remedial classes, and teachers are allocated to them in a manner related to caste, this will also create bias in the estimated coefficient. In Punjab, this concern is ameliorated because most schools have only one class per grade. Another potential difficulty is non-random matching across grades. For example, schools may place high quality, high caste teachers in higher grades and high and low caste students that perform well will get promoted to these grades. Schools could alternatively have a policy of matching high quality teachers to low ability cohorts, for example. It is also possible that teachers and children (or parents) can lobby for reshuffling of teachers within schools based on teacher characteristics that they prefer, including caste, or they may strategically time enrolment or arrange for their children to skip grades or be held back. Any type of positive sorting within schools (high quality teachers matching to high quality students) would overestimate the effect of teachers. For children, within schools, any non-random sorting would have to be due to assignment to grades in the absence of multiple classes per grade. As a result, I include grade dummies, as well as the child's age and a dummy variable for whether the child was held back in the grade. This will control for differential age at enrolment, grade assignment and non-promotion based on caste. The grade dummies also control for the possibility of high caste teachers being allocated to higher grades, as long as any non-random sorting of teachers across grades is common across schools.

In addition, school fixed effects do not control for children or teachers switching schools [8]. If low caste children that perform well switch to a better school, or to a school with more high caste teachers, then this will overestimate the effect of high caste teachers. There are nine cases in which the switch between high and low caste teachers is because of switching schools (6.9% of total switches between high and low caste teachers). I omit these observations from the analysis. I also check whether teachers match to schools based on caste. Not all teachers in a school were surveyed, but the sample of teachers surveyed in each school is random. A regression of a dummy variable for a high caste teacher on over fifty school controls produces only six significant correlates (four at the 10% level, one at the 5% level, and one at the 1% level, see Appendix Table 6.22). This finding further alleviates concerns of unobserved teacher factors influencing caste-based matching to schools.

Child fixed effects estimation can control for non-random sorting of children within schools. Such estimation relates the difference between a child's test score in year t and the average test scores of the child over all years, to the deviation of the child, teacher or school characteristics of the child in time t from the average for the child in all years:

$$\begin{aligned}(T_{ijst} - \overline{T}_i) &= \beta_1(\boldsymbol{D_{ijst}} - \overline{\boldsymbol{D}}_i) + \beta_2(\boldsymbol{XC_{it}} - \overline{\boldsymbol{XC}}_i) + \beta_3(\boldsymbol{XF_j} - \overline{\boldsymbol{XF}}_{j,i}) \\ &+ \beta_4(\boldsymbol{XC_{jt}} - \overline{\boldsymbol{XC}}_{j,i}) + t + (\alpha_j - \overline{\alpha}_{j,i}) + (\epsilon_{ijst} - \overline{\epsilon}_i)\end{aligned} \tag{6.3}$$

Now, time-invariant observed and unobserved child factors also drop out. With child fixed effects, non-random switches between high and low caste teachers is the concern. A key assumption for child fixed effects to be valid is that past error terms are not correlated with the switching of teachers. For example, if children that perform better (or worse)

[8]Only 2.6% of children in the sample switch schools between rounds, and only 2.7% of teachers switch schools between rounds.

last year switch to high caste teachers, this would be a problem. As a result, I run a regression of switching between high and low caste teachers on lagged test scores, as well as other controls. The results are contained in Appendix Table 6.23. Lagged test scores do not predict switching between high and low caste teachers, however, lagged test scores can be included in the regressions. Child fixed effects also requires that unobserved inputs such as motivation, preferences for an own caste-type (or opposite caste-type) teacher, and any others, be constant over time and not affect switching of teachers. Finally, the fixed effect (for example, child ability) also must be constant over time for child fixed effects to difference it out.

I will begin by estimating a regression of average test scores on a dummy variable (D) for being a high caste child. This specification will identify the effect of child caste on learning outcomes. This effect can be identified only with school fixed effects, of course. I will also add to this specification a dummy variable for having a high caste teacher in that year. The dummy variable for having a high caste teacher can be identified with both school and child fixed effects, so I will identify the effect of teacher caste mainly through child fixed effects regressions. Identification here is off the same child switching between high and low caste teachers.

Then, I will look at differential effects by caste-type of the child. Here as well, I will use both school and child fixed effects. Within schools there are four possible matches of high/low caste students and teachers. In the specifications with school fixed effects, the dummy variables included in $\boldsymbol{D}$ are:

- child high caste, teacher high caste;
- child high caste, teacher low caste; and,
- child low caste, teacher high caste.

The coefficients on these dummies represent the effect of that particular match relative to the omitted category of low caste students matched with low caste teachers. In the specifications with child fixed effects, the dummies included in D are:

- child high caste, teacher high caste; and,

- child low caste, teacher high caste.

Since this is a within-child specification and a child can only be either high or low caste, we can only include two dummy variables, and the coefficients represent, for that particular caste-type of child, the effect of being taught by a high caste teacher relative to being taught by a low caste teacher. Identification of the two variables is off children who switch between high and low caste teachers over time.

Non-random attrition in the sample could bias coefficient estimates. If the lowest ability low caste children drop out of school, then the remaining ones will induce positive matching of teachers and students by caste (within schools). Most children are observed in at least three rounds of the survey, and there are no significant differences between high and low caste children. See Table 6.24 in the appendix for details.

Another potential concern is that caste may be proxying for other characteristics. Consequently, I control for some child and household-level characteristics that could be correlated with caste. As child controls I include the age of the child, a dummy for whether the child is female (in the school fixed effects specifications), a household asset index, dummies for the grade level of the child, and a dummy for whether the child is repeating the grade. High and low caste teachers may also differ systematically in their observable and unobservable characteristics, and so it is important to control for observable characteristics. I will also employ methods in the next section to address unobservable characteristics. As teacher controls I include the teacher's age, number of years of teaching experience, dummies for the highest level of education, and dummies for whether the teacher is female and is originally from the same village in which the school is located. As school controls I include the number of children enrolled in the school, and a school facilities index [9]. Finally, time dummies are included to control for the fact that children tend to perform better on the test over time.

[9]This index is constructed from a number of questions on the survey that ask about whether the school has the following facilities: library, computer, sports ground, hall, surrounding wall, fans, electricity, the type of toilet, whether potable water is available, and the seating arrangements for students (desks, mats, etc.). Principal Components Analysis was used to aggregate these measures.

6.6 Results

This section will discuss the results of estimating the empirical models presented in the previous section. I will begin with identifying the effects of child caste, teacher caste, and the interaction of teacher caste with student caste, and then will look into some potential mechanisms by which the findings could be occurring.

6.6.1 Child caste, teacher caste, and the interaction between child and teacher caste

Table 6.6 presents basic results on the effect of child caste, as well as the effect of high caste teachers on all children. The outcome is the average of math, English and Urdu test scores for each child. Columns (1)-(3) present school fixed effects results. A dummy variable for high caste children is included, as is a dummy variable for a high caste teacher. Column (1) does not include controls and thus provides a basic correlation, column (2) adds child, teacher and student controls (equation (6.2)). Column (3) adds the lagged test score to the right hand side to control for past learning. It is meant to capture the contribution of all previous inputs and any past unobservable endowments or shocks, and is becoming increasingly common to include in test score specifications (Todd and Wolpin, 2007; Andrabi et al., 2011). Here, one could also worry that children are being matched to high/low caste teachers based on their previous test scores (for example, if children performing poorly were placed with high caste, high quality teachers). Columns (4)-(6) present child fixed effects results, which identify only teacher caste effects (as child caste is time-invariant and drops out). Column (4) once again has no controls. Column (5) adds child, teacher and school controls (equation (6.3)). Column (6) presents results using the Arellano-Bond (A-B) estimator, which properly includes the lagged test score as well as accounts for the panel dimension of the data [10].

From columns (1)-(3) we can see that although in the cross section, low caste children have higher test scores than high caste children, this is not the case within schools. This

[10] Lagged dependent variable models with fixed effects are appropriately estimated with the A-B estimator, as simply including the lag within a fixed effects model induces correlation between the error term and the lag (Cameron and Trivedi, 2009).

result suggests that low caste children are attending better schools. Table 6.25 in the appendix investigates this point further. The first column presents the difference in test scores between high and low caste children, for all schools, public schools and private schools from a simple t-test. The second column presents the coefficient on a dummy variable for high caste children from a regression of average test scores on this dummy as well as child, teacher and school controls, with school fixed effects. This is done for all schools, and for public schools and private schools separately. The coefficients from the school fixed effects regressions are all larger than the differences from the t-tests. These results suggest that low caste children are attending better schools in terms of test score outcomes, and that this is happening to a greater extent in public schools. In all specifications in Table 6.6, the estimated coefficient for switching to a high caste teacher is not statistically different from zero.

Table 6.6: Child caste and teacher caste

Outcome: average of math, English and Urdu test scores						
	(1)	(2)	(3)	(4)	(5)	(6)
		School FE			Child FE	
	Basic	Controls	+ lag	Basic	Controls	+ lag
High caste child	0.025	-0.004	-0.058			
	(0.078)	(0.073)	(0.056)			
High caste teacher	0.048	0.012	0.064	0.030	0.017	-0.077
	(0.122)	(0.094)	(0.122)	(0.069)	(0.070)	(0.140)
Lagged test scores			0.503***			0.297*
			(0.057)			(0.154)
Observations	2642	2642	1492	2642	2642	538
R^2	0.115	0.229	0.709	0.409	0.419	

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

All regressions include child, teacher and school controls, and time dummies.

Standard errors clustered at the level of the fixed effect.

In Table 6.7, I disaggregate by caste type of the child to identify the effect of the interaction between child and teacher caste. Columns (1)-(3) are school fixed effects regressions, and columns (4)-(6) are child fixed effects regressions. Columns (3) and (6) also include the lagged test score. The school fixed effects regressions show no effect of the caste interaction dummies on test score outcomes, but the child fixed effects regressions show a positive effect of high caste teachers for low caste children. However, the coefficient is not significant in specification (6), which includes the lagged test score. Including the lagged test score would be preferable, but it removes much of the variation in child-teacher caste combinations. In the regressions that include the lagged test score, the first year of data is lost so instead of three changes, there are only two. Although there are four years of data on children, very little teacher data was collected in the fourth year as most children were in middle school by then, and middle school teachers were not surveyed. As a result, there are only 100 observations for this year, so losing the first year removes many of the switches in the data. With 2,642 child*year observations over all four years, there are only 120 switches. Without the first year there are 1,492 observations and only 46 switches. This is not sufficient to precisely estimate these coefficients. Again, it is not the case that children are switching to high/low caste teachers based on their previous test scores (see Appendix Table 6.23). Lagged test scores do not predict switching between high and low caste teachers.

Table 6.7: Interaction between child and teacher caste

Outcome: average of math, English and Urdu test scores						
	(1)	(2)	(3)	(4)	(5)	(6)
		School FE			Child FE	
	Basic	Controls	+ lag	Basic	Controls	+ lag
Child high caste,	0.057	-0.007	-0.029	-0.041	-0.057	-0.242
teacher high caste	(0.188)	(0.162)	(0.166)	(0.093)	(0.095)	(0.167)
Child low caste,	0.034	-0.006	0.022	0.170**	0.167**	0.241
teacher high caste	(0.181)	(0.148)	(0.169)	(0.085)	(0.082)	(0.189)
Lagged test scores			0.503***			0.298*
			(0.057)			(0.153)
Observations	2692	2642	1492	2692	2642	1492
R^2	0.119	0.229	0.709	0.415	0.420	

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

All regressions include child, teacher and school controls, and time dummies.

Standard errors clustered at the level of the fixed effect.

Consequently, (5) in Table 6.7 is the preferred specification with all child, teacher and school controls, and child fixed effects. High caste children perform equally well when taught by high or low caste teachers. However, low caste children perform significantly better when taught by high caste teachers. Switching from a low caste teacher to a high caste teacher increases the test scores of low caste children by 0.17 standard deviations on average, and this effect is significant at the five per cent level. That the coefficient from the school fixed effects regression differs so much from that of the child fixed effects regression shows the importance of controlling for this child-level unobserved heterogeneity. The results suggest that it is 'worse' low caste children that are switching to high caste teachers. The interaction between child and teacher caste matters for learning outcomes in Pakistan.

As a robustness check, I re-run specification (5) in Table 6.7 in two ways. I restrict the sample first to schools for which the observed frequencies of all four possible caste matches fell within the 95% confidence interval for the caste match groups, which was constructed using the simulated data. I then restrict the sample to village-years that passed the Chi-squared distribution test [11]. Appendix Table 6.26 contains the results with the dummies for the interaction between child and teacher caste. When the sample is restricted in these ways, the results are stronger. High caste teachers once again do not matter for high caste children, and low caste children perform significantly better when taught by high caste teachers. For low caste children, switching to a high caste teacher increases test scores by 0.26 standard deviations (significant at the 10% level) when the sample is restricted to schools that pass the confidence interval test, and increases test scores by 0.22 standard deviations (significant at the 5% level) when the sample is restricted to village-years that pass the Chi-squared distribution test.

Table 6.8 investigates the heterogeneity of this effect further, and shows how the results differ by gender of the child and for the type of school the child attends (public or private). Again, for each type of child in terms of caste and gender, the omitted category is being matched with a low caste teacher. Column (1) includes only child fixed effects and no controls, column (2) adds time dummies, column (3) adds child controls, column (4) adds teacher controls, and column (5) adds school controls, to make the full set of controls. Column (6) contains results for children in public schools, and column (7) for children in private schools. The results are relatively stable over specifications. The effect of a low caste child performing better when matched with a high caste teacher appears

[11] No schools failed the Chi-squared distribution test. For village-years that pass the 95% confidence interval test, there are no switches between high and low caste teachers within children remaining once the sample is restricted and so the coefficients cannot be estimated.

to be stronger for boys than for girls, and seems to be stronger in public schools [12]. For low caste boys, switching from a low caste to a high caste teacher increases test scores by 0.283 standard deviations, and this is significant at the one per cent level. This effect is not significant in private schools, but is significant in public schools [13]. It is noteworthy that Andrabi et al. (2011) and Das et al. (2006) have found that learning outcomes are significantly higher in private schools than in public schools in Punjab. This is true for both high and low caste children (see Appendix Table 6.27). Others have also noted that private schools deliver higher learning outcomes than public schools in Pakistan (Alderman et al., 2001; Andrabi et al., 2008; Aslam, 2009). Perhaps in private schools, learning outcomes are more equal, so having a high caste teacher does not contribute an additional benefit for low caste boys. There is certainly an effect to explain in public schools.

6.6.2 Mechanisms

In the cross section, low caste children have higher test scores than high caste children. Within schools, there are no differences, but low caste children are selecting better (especially public) schools. Children on average do not have differential learning outcomes when taught by high versus low caste teachers. It is low caste boys in particular that appear to benefit from being taught by high caste teachers. There are several possible reasons for these findings, and I explore six of them here. Perhaps low caste children have high returns to education. It is also possible that the finding of no difference in learning outcomes between high and low caste children within schools masks heterogeneity depending on whether the school is high or low caste dominant. Teacher caste may still matter, but high caste teachers may differ from low caste teachers in terms of unobservable characteristics and may be of higher quality. The result on high caste teachers and low caste boys could be because these high caste teachers serve as role models to low caste boys. Alternatively, perhaps low caste teachers discriminate against low caste

[12] The coefficient for low caste girls in public schools is identified off only 8 switches. I will not focus on this result since so few children are driving it.

[13] However, the p-value in a test for the equality of the coefficient on child low caste male, teacher high caste between the public and private school specifications is 0.802, so they are not significantly different.

Table 6.8: Child and teacher caste interaction by gender and public/private school

	Outcome: average of math, English and Urdu test scores						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Basic	+ Time	+ Child	+ Teacher	+ School	Public	Private
C: high, female. T: high	-0.147	0.014	0.014	-0.008	0.004	0.135	-0.076
	(0.157)	(0.098)	(0.099)	(0.097)	(0.097)	(0.169)	(0.151)
C: high, male. T: high	0.061	-0.088	-0.094	-0.113	-0.104	-0.207	-0.047
	(0.185)	(0.148)	(0.150)	(0.150)	(0.150)	(0.218)	(0.133)
C: low, female. T: high	-0.121	-0.126	-0.113	-0.162	-0.163	-0.398**	-0.125
	(0.248)	(0.132)	(0.135)	(0.134)	(0.134)	(0.192)	(0.195)
C: low, male. T: high	0.563***	0.291***	0.288***	0.281***	0.283***	0.224**	0.172
	(0.110)	(0.097)	(0.097)	(0.092)	(0.091)	(0.110)	(0.179)
Observations	2642	2642	2642	2642	2642	1970	672
R^2	0.009	0.411	0.414	0.421	0.422	0.415	0.480

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

All regressions include child fixed effects and standard errors clustered at the child level.

C denotes child, T denotes teacher.

boys, or high caste teachers favour low caste boys. A sixth possibile mechanism is that high caste teachers can provide access to certain benefits for low caste boys through their patronage networks if they perform well in school.

High versus low caste dominant schools

I begin with looking further into the effects of child caste. It is easy to imagine that low caste children may be treated differently if they are surrounded by predominantly high caste peers. This was suggested in Jacoby and Mansuri (2011). From the perspective of the identity economics literature, if a school is high caste dominant, low caste children may feel more like 'outsiders' and may have lower learning outcomes. Table 6.9 displays results from a school fixed effects regression of average test scores on dummies for the interaction between child caste and whether a school is high caste dominant. A school is considered high caste dominant if the proportion of children enrolled in grades 1 to 5 who are high caste is greater than the proportion of children enrolled in grades 1 to 5 who are low caste. Since a school is either high or low caste dominant, the coefficients reflect the difference in test scores between high and low caste children in high or low caste dominant schools. Table 6.9 shows that there are no differences between high and low caste children in either high or low caste dominant schools.

Table 6.9: High and low caste dominant schools

Outcome: average of math, English and Urdu test scores	
	(1)
High caste child, high caste dominant school	-0.013
	(0.074)
Low caste child, high caste dominant school	0.036
	(0.019)
Observations	2642
R^2	0.102

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Regression includes child, teacher and school controls, and time dummies.

Standard errors clustered at the level of the fixed effect (school).

Returns to Education

Next, I test whether low caste children have high returns to education. If they do, this can help explain why we see that low caste children have higher test scores than high caste children in the cross section, and why they select better schools. I look at the parents of the children in the sample and compare returns to education for high caste and low caste men in the same village, and also compare women in general [14]. Table 6.10 provides the results from regressing the log of the monthly wage on dummy variables for having completed primary school, middle school, high school, or more education than high school. The omitted category is having completed either no education or not having finished primary school. I include as many individual and household level controls as possible (age, marital status, household size, deaths in the household, whether the household owns land, whether anyone left the household, whether the household receives remittances, total expenditures, and the asset index). Although the coefficients for women are large, they are very imprecisely estimated because so few women actually report earning a positive income. Returns to education for the women who select into wage earning jobs may be high. However, many authors have noted that in general, returns to education are very low for women in Pakistan and that positive returns only accrue after beginning high school (Aslam et al., 2008; Aslam, 2009; Aslam and Kingdon, 2012). Returns to completing primary school are identical (and indistinguishable from zero) for both high and low caste men. Returns to completing middle school, high school, and more than high school are higher for low caste men than for high caste men (see also Figure 6.4) [15]. This shows the importance of education for economic (and possibly social) mobility for the low caste male children in the sample.

Teacher Quality

The argument could also be made that high caste teachers are simply high quality teachers. Although we do not see that teacher caste matters for learning outcomes of children on average, perhaps there are unobserved aspects of teacher quality that differ

[14] So few women report positive wages that there are insufficient observations to run separate regressions for high and low caste women.

[15] However, the coefficients in for high and low caste men are not individually or jointly significantly different from one another.

Table 6.10: Returns to Education

	Outcome: log of the monthly wage					
	(1)		(2)		(3)	
	High caste men		Low caste men		Women	
	coef	se	coef	se	coef	se
Primary school	0.099	0.072	0.099	0.136	-0.636	0.599
Middle school	0.141*	0.084	0.234	0.195	1.048	0.961
High school	0.266**	0.115	0.602**	0.302	1.099	0.996
More than high school	0.562***	0.151	0.791**	0.353	1.853	1.171
Constant	7.627***	0.291	7.448***	0.47	6.649***	1.54
Number of observations	1,020		314		105	
R^2	0.27		0.34		0.686	

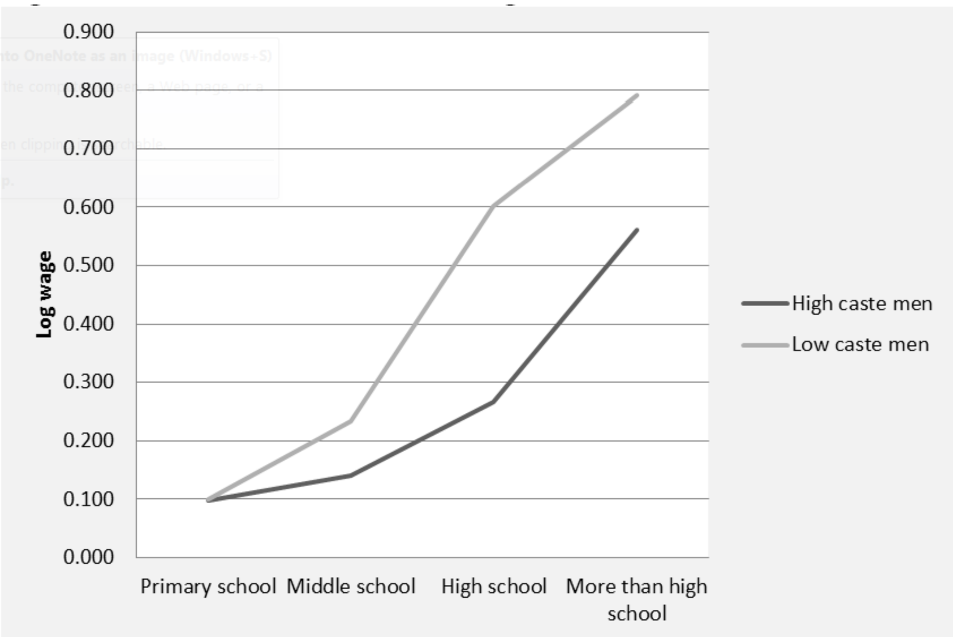
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Omitted category: less than primary school.

Regressions include individual and household controls, as well as time dummies and village fixed effects.

Standard errors clustered at the level of the fixed effect.

Figure 6.4: Returns to Education



by caste. To investigate this further, I run a regression of male children's average test scores on child, teacher and school controls, but this time including teacher*time fixed effects. I also include a dummy variable for high caste students. The coefficient on this dummy shows the differences in test scores for high and low caste boys taught by the same teacher in the same year (children in the same class). This differences out time-invariant unobserved aspects of teacher quality. Table 6.11 provides the results. High caste students have significantly lower test scores than low caste students in the same class. This effect is stronger in public schools, with high caste boys scoring 0.182 standard deviations lower than low caste boys, and this effect is significant at the five per cent level. In addition, the effect is not present for low caste teachers in public schools (column (3)) but is for high caste teachers in public schools (column (4)). Another indication that high caste teachers are not simply higher quality teachers is that they do not similarly benefit high caste children. High caste children perform equally when taught by high and low caste teachers, as the results of Table 6.8 show. Table 6.8 also shows that including teacher observables in the regression does not change the coefficient on the caste dummy variables very much at all (comparing columns (3) and (4)). This indicates that teacher observable characteristics that could be related to quality also do not vary systematically by caste. So it is not the case that high caste teachers are just better teachers; it is that some children perform significantly better when taught by them.

Table 6.11: Test scores with teacher fixed effects

	Outcome: average of math, English and Urdu test scores			
	(1)	(2)	(3)	(4)
	All Schools	Public Schools	Low Caste (Public)	High Caste (Public)
Child high caste	-0.147*	-0.182**	-0.343	-0.143*
	(0.080)	(0.087)	(0.325)	(0.087)
Observations	1461	1101	133	968
R^2	0.055	0.065	0.076	0.073

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

All regressions include child, teacher and school controls, time dummies, and teacher*time fixed effects.

Standard errors clustered at the teacher*time level.

Discrimination

Now I will return to the interaction between child and teacher caste to try to understand why low caste boys benefit from being taught by high caste teachers. I cannot test whether teachers are biased when they are grading children. However, the way in which teachers perceive and judge the performance of children could potentially be important for learning outcomes. To test this, I construct a measure that can be thought of as measuring biased perceptions. Teachers were asked to rate children on their academic performance on a scale of 1 to 10. I rank children in the same class according to these ratings. I also rank children in the same class according to an average of their three test scores. I then subtract the teacher's ranking from the test ranking, and create a dummy variable for whether this difference is negative (meaning that the teacher overrates the child's performance). This outcome is regressed on the same four caste interaction dummies and controls as in Table 6.8, along with child fixed effects and time dummies. The results are contained in column (1) of Table 6.12. The coefficient on the caste interaction dummy for low caste boys is not significantly different from zero. It does not look like high caste teachers are over- or under-rating the academic performance of low caste boys. Interestingly, high caste teachers appear to be overrating high caste boys. This does not seem to translate into higher test scores for them, however.

Teachers may actually be able to rate children's performance quite accurately if they have more information on the child's ability than one year of test scores can provide. Perhaps they are aware of previous performance, or of family circumstances that may affect a child's current performance. The test score measure is noisy, so if teachers do have more knowledge on a child's true ability, then the outcome variable above may not reflect discrimination; it would reflect the deviation from true ability. A teacher's assessment of a child's ability may be a function of true ability of the child and some degree of error. Knowledge about children's true ability may differ by caste of the child and teacher. If this is the case, the coefficients on the caste interaction dummies in the discrimination regression will be biased, as there will be systematic measurement error in the oucome variable that is correlated with caste. If teachers of the same caste-type as a child have better information (so deviation of the ranking from true ability is lower) than

Table 6.12: Discrimination Effects

Outcome: difference between actual test score rank and teacher's rank of child	(1)	(2)
	One year	Cumulative
C: high, female. T: high	-0.260	-0.045
	(0.162)	(0.220)
C: high, male. T: high	0.250**	0.240**
	(0.122)	(0.107)
C: low, female. T: high	0.421	0.341
	(0.442)	(0.250)
C: low, male. T: high	-0.162	-0.081
	(0.186)	(0.169)
Observations	1149	1149
R^2	0.040	0.025

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

All regressions include child, teacher and school controls, and time dummies.

Standard errors clustered at the level of the fixed effect (child).

teachers of the opposite caste-type as a child, then the coefficient for low caste children taught by high caste teachers will be biased downwards (upwards) if true ability is higher (lower) than ability demonstrated by the test score. Although the coefficient is negative and insignificant, the presence of this type of bias may affect the interpretation of the above result. As a check, I construct a test score measure that is the average of not just the three current test scores, but also all test scores that occurred in previous years for which I have data [16]. I then rank the students in the same class again according to this new test measure and compare the difference between the ranking of the teachers rating and the ranking of this cumulative test score. If teachers are using information on past test scores in their assessment, then including this information in the discrimination measure should reduce the measurement error in their ranking (possibly by less for low caste children, however). The results are contained in column (2) of Table 6.12. The results are actually quite similar to those of column (1), so I do not believe that this type of systematic measurement error is a problem here.

[16] In the first year, only the three current-year tests are used.

Role model effects

Next, I focus on role model effects. One interpretation of role model effects that was discussed in Section 6.3 was that teachers who are more like the child could elicit better performance. However, this does not seem to be the case, as it is high caste teachers that produce high test score outcomes for low caste boys. Another way to look at role model effects is that a child may like to emulate his/her teacher. To check whether high caste teachers could be serving as role models to low caste children, I use data on what the children report wanting to do when they grow up. In the third and fourth rounds of the survey, the children were asked to state what profession they would like to eventually go into, and I construct a dummy variable for whether the child would like to be a teacher (27% of the sample overall), which could be a sign that the child would like to be like their teacher, and so may consider the teacher as a role model. I also construct a dummy variable for whether the child reports wanting to have a skilled job apart from teaching (doctors, government jobs, politicians, engineers, or private sector jobs). Perhaps high caste teachers could inspire children to have high aspirations. I regress these outcomes on child and teacher caste and gender type dummies within schools [17]. The omitted category is a low caste male child matched with a low caste teacher. The results are reported in Table 6.13. It appears that low caste boys report wanting to become teachers significantly less when they are taught by high caste teachers than low caste teachers. Interestingly, high caste boys appear not to want to go into teaching, regardless of whether they are taught by high or low caste teachers (the p-value for the difference in coefficients for high caste boys taught by high and low caste teachers is 0.625). Low caste boys taught by high caste teachers are marginally more likely to report wanting to have a high skilled job than low caste boys taught by low caste teachers. Again, high caste boys are very likely to report wanting a high skilled job, regardless of the caste of their teacher. It does not look like low caste boys would like to emulate their teachers by also becoming teachers. Instead, they would like to go into other high skilled professions, many of which pay better than teaching (for example, doctors and engineers). This result points to the next mechanism.

[17] There are not enough observations to include child fixed effects since the question was only asked in the third and fourth rounds of the survey, and so I include school fixed effects instead.

Table 6.13: Role Model Effects

Outcome	(1) Teaching Career	(2) Skilled Job
C: high, female. T: high	-0.042	0.203
	(0.265)	(0.255)
C: high, female. T: low	0.043	0.083
	(0.297)	(0.287)
C: high, male. T: high	-0.520**	0.375
	(0.253)	(0.237)
C: high, male. T: low	-0.436**	0.540***
	(0.182)	(0.178)
C: low, female. T: high	-0.230	0.239
	(0.272)	(0.264)
C: low, female. T: low	-0.169	0.108
	(0.380)	(0.366)
C: low, male. T: high	-0.552**	0.402*
	(0.251)	(0.235)
Observations	743	743
R^2	0.136	0.086

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

All regressions include child, teacher and school controls, and time dummies.

Standard errors clustered at the level of the fixed effect (school).

Patronage

Finally, I explore the role of patronage in the education sector. I examine the possibility that high caste teachers are able to help low caste boys with educational or employment opportunities, thereby increasing their already high returns to education when taught by them. That the positive effect of high caste teachers on low caste boys' learning outcomes is present in public but not in private schools may be important. High caste public school teachers are the ones that often have access to patronage networks in the government and would be in a better position to be able to help children get access to school-related benefits and public sector jobs. Private school teachers would not have these connections with local officials. Teachers in public schools will have interaction with local government officials at the district education offices, and also possibly other government officials located nearby. They would be more aware of government programs (for example scholarships and other educational benefits) and could also rely on this network for information on the availability of public sector jobs. Since private school teachers have less direct interaction with local governments regarding their schools, they would be expected to have weaker networks with the government and less information on government programs and public sector jobs.

The interpretation of the results on public and private schools could depend on selection of teachers and students into public and private schools. In order to generate the above result, one or a combination of the following would need to be the case: high ability low caste children select into public schools; low ability low caste children select into private schools; high ability or better connected high caste teachers select into public schools; low ability or worse connected high caste teachers select into private schools; low ability low caste teachers select into public schools; or high ability low caste teachers select into private schools. In terms of observable characteristics, it does not look like higher ability low caste children are selecting into public schools and lower ability low caste children are selecting into private schools; the low caste children in public schools have lower test scores than those in private schools (of course this could also reflect school quality), and no household characteristics predict attending a public versus a private school (see Appendix Table 6.28, column (1)). For both high *and* low caste teachers, there are observable differences between those that teach in public and private schools. Both high and low caste teachers in public schools tend to be older, male, from the same village, and have more experience. However, they have lower levels of education (see Appendix Table 6.28, columns (2) and (3)). In addition, the teachers' test scores do not differ between public and private schools, and many of the observable characteristics such as experience and education have been found not to matter for learning outcomes in the literature. It is possible that younger, female teachers from within the village make better teachers, for example. In fact, learning outcomes of children of high caste teachers are higher in private schools than in public schools. Of course, there could be other things specific to private schools such as more monitoring or accountability as well. However, there is no compelling evidence to suggest that high caste teachers in public schools are better teachers than those in private schools, or that low caste teachers in public schools are worse teachers than those in private schools.

Another interpretation of the sorting of teachers between public and private schools is not that it is higher ability high caste teachers but high caste teachers with better connections that end up teaching in public schools. After all, the salaries are higher at public schools and a teacher would likely have needed a connection in the government in order to secure a public sector teaching job. The ability to use networks to secure a job may be, but is not necessarily related to teaching ability. This interpretation of selection is consistent with the results for public versus private schools, and with the conjecture that one mechanism by which high caste teachers affect the learning outcomes of low caste boys is patronage, because teachers with better connections to get themselves jobs are also in a better position to help their students.

If teachers are helping students through their patronage networks, we would also expect the effect of high caste teachers on low caste boys to be larger in cases in which the village is high caste dominant. In these cases, high caste teachers are likely to have greater influence. High caste dominant villages are those in which the proportion of high caste people that own land is greater than the proportion of low caste people that own land (85% of villages) [18]. I focus on low caste boys in public schools and regress the

[18]Before the main four-year survey was carried out, a village-level household census was conducted in

average test score outcome on a dummy variable for a high caste teacher in a high caste dominant village and a dummy variable for a high caste teacher in a low caste dominant village. This is also a child fixed effects regression, and since children either live in a high or low caste dominant village, the omitted category is having a low caste teacher in that village type. I also include child, teacher and school controls as before, as well as time dummies. The results from a regression for low caste boys only are contained in Table 6.14. From Table 6.8 we saw that for low caste boys overall, being taught by a high caste teacher produces test scores that are 0.283 standard deviations higher than being taught by a low caste teacher. In public schools, the effect was 0.224 standard deviations (and was small and insignificant for private schools). From Table 6.14, in public schools in high caste dominant villages, test scores are 0.354 standard deviations higher when the child is taught by a high caste teacher. This effect is significant at the five per cent level. In public schools in low caste dominant villages, the effect is statistically indistinguishable from zero.

Table 6.14: High versus Low Caste Dominant Villages

Outcome: average of math, English and Urdu test scores	
	(1)
T: high. High caste dominant village.	0.387***
	(0.146)
T: high. Low caste dominant village.	0.109
	(0.234)
Observations	384
R^2	0.535

Standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
All regressions include child, teacher and school controls, and time dummies.
Standard errors clustered at the level of the fixed effect (child).

Here as well, we could worry about differences between low caste children and high and low caste teachers in high versus low caste dominant villages. It is possible that low caste children in high caste dominant villages are better in some way than low caste children in low caste dominant villages. It is also possible that high caste teachers in

all the LEAPS villages. Land ownership and caste data was recorded for all households. Whether a village is high or low caste dominant is determined using this data.

high caste dominant villages are either better teachers or have a preference for teaching in high caste dominant villages, compared to high caste teachers in low caste dominant villages. Low caste teachers in high caste dominant villages may also be poor teachers and/or prefer not to be in high caste dominant villages, compared to those in low caste dominant villages. These could also produce the result that high caste teachers have a stronger effect on low caste boys in high caste dominant villages. I check whether low caste children, and high and low caste teachers differ in observable characteristics between high and low caste dominant villages. Low caste children in high caste dominant villages have a higher household asset index than those in low caste dominant villages, but they are also more likely to have been held back and there are no differences in test scores (see Appendix Table 6.29, column (1)). This is not compelling evidence of the type of selection described above. There are no significant differences between high caste teachers in high and low caste dominant villages. The only difference between low caste teachers in high and low caste dominant villages is that low caste teachers in high caste dominant villages are more likely to have a bachelor's degree (see Appendix Table 6.29, columns (2) and (3)). So I do not find support for the argument that the greater effect of high caste teachers in high caste dominant villages is due to such differences in children and teachers between high and low caste dominant villages.

To investigate further the possibility that high caste teachers may be able to provide low caste boys with educational or employment benefits, I use the follow up survey conducted in 2011. It was still too soon to have data on employment outcomes, but by 2011, the children were of high school age (and the enrolled children were on average in grade ten). I find that low caste boys in public schools that were previously taught by high caste teachers are significantly more likely to be enrolled in school compared to boys that were not taught by any high caste teachers. They have also completed more schooling even if they have dropped out. Table 6.15 presents results from a regression for low caste children in public school. The outcome of being enrolled in school in 2011 is regressed on a dummy for having had at least one high caste teacher during the four years of the main survey in column (1), and the outcome of the highest grade completed by 2011 is regressed on the same dummy in column (2). Both specifications also include child level controls, as well as average characteristics of the teachers they had. These characteristics include average age, experience, test scores, and education, as well as dummies for the number of female teachers and number of teachers from the same village as the school. Low caste children that attended public schools [19] and had at least one high caste teacher between 2003 and 2006 are significantly more likely to be enrolled in school in 2011, and had completed almost one additional grade by 2011 compared to children that were not taught by any high caste teachers. Since I cannot include school or child fixed effects here, these results should not be taken as causal [20]. Finally, another interpretation of the results on the children's preferred career is that the low caste children taught by high caste teachers are actually aiming for better careers because they believe the high caste teacher they currently have may be able to help them attain it either through educational benefits or connections to employment opportunities, and this increases the returns to education for low caste children.

Table 6.15: Enrolment and highest grade in 2011

Outcome	(1) Enrolment	(2) Highest Grade
At least one high caste teacher	0.985**	0.875**
	(0.461)	(0.421)
Number of observations	85	83
R^2	0.728	0.986

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Regressions include child controls as well as controls of previous teachers, and village fixed effects.

Only low caste children that attended public schools during 2003-2006 are included.

Standard errors clustered at the village level.

It is also interesting to note that the returns to education for high and low caste men differ in high and low caste dominant villages. Table 6.16 presents results of a regression of the log of the monthly wage of the children's parents on education dummies, interacted with dummy variables for high and low caste dominant villages. I also include individual

[19]The effect for private schools cannot be identified. There are only 28 low caste children observed in round 5 that went to private school. Two did not have a high caste teacher during the main survey period, one child is enrolled in round 5 and the other is not. 26 children did have a high caste teacher during the main survey period. 13 are enrolled in round 5 and 13 are not enrolled in round 5. As a result, there is no variation, even between villages.

[20]In addition, attrition between 2004 and 2011 is high. Only 55% of children could be located, 57% of high caste children and 47% of low caste children.

and household controls, time dummies, and village fixed effects. For each of high and low caste dominant village types, the omitted category is less than primary education. Separate regressions are run for high and low caste men. For low caste men, returns to education are higher in high caste dominant villages. This could mean that in low caste dominant villages, low caste men do not need education for social and economic mobility to the same extent that they need it in high caste dominant villages. Perhaps in low caste dominant villages they can just rely on their own connections for employment, or perhaps since they own land they engage in farming rather than wage labour. However, it could also be the case that 'better' low caste individuals select into high caste villages in a way that is related to both educational attainment and income. Interestingly, for high caste men, returns to education are higher in low caste dominant villages. This could indicate that in low caste dominant villages, high caste men cannot just rely on their connections but need education in order to improve their earnings potential.

Table 6.16: Returns to education in high versus low caste dominant villages

Outcome: log of the monthly wage				
	(1)		(2)	
	High caste men		Low caste men	
	coef	se	coef	se
High caste dominant village: primary school	0.102	0.076	0.150	0.133
High caste dominant village: middle school	0.141	0.09	0.227	0.197
High caste dominant village: high school	0.254**	0.117	0.700**	0.282
High caste dominant village: high school+	0.493***	0.159	0.864***	0.332
Low caste dominant village: primary school	0.300	0.441	-0.089	0.211
Low caste dominant village: middle school	0.165	0.285	0.359	0.243
Low caste dominant village: high school	0.428*	0.232	0.330	0.357
Low caste dominant village: high school+	0.901***	0.173		
Constant	7.835***	0.339	7.459***	0.464
Number of observations	985		312	
R^2	0.276		0.357	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Regressions include individual and household controls, as well as village fixed effects.

Standard errors clustered at the village level.

I also look at what channel the higher test scores for children taught by high caste teachers is coming through - how are these higher test scores are achieved? It could be that parents invest more in their children's education because they know that the high caste teacher could help the child later on. Parents can do this by increasing educational expenditures, by helping the child with their schoolwork, or by making it a point to get to know the child's teacher by meeting with them regarding their child's performance in order to reinforce the social network connection and signal that they care about their child's future. It could also be that children work harder, and this could be of their own accord or with encouragement from parents.

The school survey asked teachers whether they had met the child's parents in the last month, and if so, whether the meeting had been to discuss the child's performance in school. The household survey asked parents about monthly education expenditures for each child, and how many hours per week were spent helping the child with their homework. The household survey also collected extensive data on child time use, including the number of minutes per day spent on housework and paid work, homework, leisure (including sleep/rest, play, and media), and learning (time at school and on tuition). Table 6.17 displays regressions of the household investment and child time use variables on the same specification as (6.3) with the full set of child, teacher and school controls, as well as time dummies and child fixed effects, for children in public schools. Column (2) shows that low caste boys in public schools spent significantly more time per day on homework when taught by high caste teachers, on average 29 minutes more per day. 29 additional minutes per day is a substantial amount; if the child does homework every week day, this amounts to almost two and a half hours extra per week. The mean number of minutes per day spent on homework is 143 minutes in the sample, so an additional 29 minutes is a 20% increase. None of the other outcomes show significant effects for low caste boys. So what is happening is not that households are investing more, but that children are working harder.

Patronage effects may also differ depending on whether the school consists predominantly of high or low caste children. It may be the case that the positive benefits of high caste teachers for low caste children are present only when either high or low caste children are the majority in the school. Teachers could have limitations on how many children they can help, and so it may be easier for them to help low caste children when there are not too many enrolled in the school. Alternatively, high caste teachers surrounded by many low caste students may be more sympathetic to helping them.

Table 6.17: Teacher-student caste, child time use and household investments

Outcome	(1) House/paid work	(2) Homework	(3) Leisure	(4) Learning	(5) Parents met teacher	(6) Met re: performance	(7) Hrs helping w/ hwork	(8) Education Expenditure
C: high, female. T: high	-13.982	-13.195	53.034	5.165	-0.227*	0.060	0.287	16.175
	(28.194)	(15.510)	(47.358)	(26.131)	(0.134)	(0.140)	(4.139)	(35.221)
C: high, male. T: high	4.342	16.853	-33.040	-39.533	-0.319***	-0.392***	-1.754	-9.473
	(9.427)	(17.086)	(49.384)	(28.330)	(0.114)	(0.109)	(1.957)	(29.147)
C: low, female. T: high	-56.828**	-25.972	30.712	45.088	-0.123	-0.422	.	-19.068
	(27.703)	(31.570)	(55.826)	(29.995)	(0.484)	(0.314)	.	(57.245)
C: low, male. T: high	**-13.392**	**29.707****	**-8.521**	**3.754**	**0.135**	**0.119**	**0.711**	**-16.412**
	(11.230)	**(14.230)**	**(51.582)**	**(22.684)**	**(0.170)**	**(0.176)**	**(1.687)**	**(27.799)**
Observations	2144	2093	2144	2144	1202	1202	658	2117
R^2	0.050	0.035	0.077	0.024	0.062	0.045	0.119	0.195

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

All regressions include child, teacher and school controls, time dummies, and child fixed effects.

Standard errors clustered at the child level.

Table 6.18 presents results for the child-teacher caste interaction in low caste and high caste dominant schools. The regression includes child, teacher and school controls, as well as time dummies and child fixed effects. Identification is off children of a certain caste-gender type taught by a high/low caste teacher, switching between the caste composition of the school being predominantly high (low) caste to being predominantly low (high) caste. For low caste boys, the effect of being taught by a high caste teacher is the same in high and low caste dominant schools. An F test confirms that the two coefficients on low caste boys matched with high caste teachers in high versus low caste dominant schools are not significantly different from one another (p-value 0.96).

Table 6.18: High and Low Caste Dominant Schools

Outcome: average of math, English and Urdu test scores				
Child	Teacher	School	(1)	
			coef	se
High caste, female	High caste	Low caste dominant	-0.006	(0.122)
High caste, female	Low caste	High caste dominant	-0.043	(0.114)
High caste, female	Low caste	Low caste dominant	0.253	(0.291)
High caste, male	High caste	Low caste dominant	0.144	(0.093)
High caste, male	Low caste	High caste dominant	0.088	(0.184)
High caste, male	Low caste	Low caste dominant	0.254	(0.160)
Low caste, female	High caste	High caste dominant	0.043	(0.193)
Low caste, female	High caste	Low caste dominant	-0.079	(0.188)
Low caste, female	Low caste	High caste dominant	0.181	(0.260)
Low caste, male	High caste	High caste dominant	0.279**	(0.117)
Low caste, male	High caste	Low caste dominant	0.285**	(0.117)
Low caste, male	Low caste	High caste dominant	0.064	(0.112)
Number of observations				2,642
R^2				0.424

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Regressions include child, teacher and school controls, time dummies, child fixed effects.

Standard errors clustered at the child level.

The results in this section help to shed some light on the reasons for which low caste children have higher test scores and choose better schools, high caste teachers do not matter on their own, and low caste boys benefit from being taught by high caste teachers. The mechanism cannot be proven, but the data does point to the possibility that high caste teachers increase the already high returns to education of low caste boys because they are able to help them later on, thereby leading these boys to work harder. All other explanations cannot be completely ruled out, but the data certainly point to patronage networks as one possibility.

What is Caste?

Since there are important effects of caste on learning outcomes, it is important to understand what these social differences between castes mean. It is possible that this measure of social distance is capturing something else that is related to caste. As discussed in the Introduction, there are gaps in learning between households that are poorer, and with lower levels of education. Are these inequalities the same as caste inequalities? I focus on children and look at two alternatives: household wealth and parental education. Instead of splitting children into high and low caste, I split them into 'not poor' and 'poor', and educated and uneducated father [21]. I use the household asset index and classify households as poor if they fall into the bottom quintile of this index. The average test scores of the children are then regressed on dummies analogous to the teacher-student caste interaction dummies, but with wealth and parental education for children, and high and low caste as before for teachers. Table 6.19 contains the results for wealth in column (1) and for parental education in column (2). The coefficients do not resemble those of the original caste interaction dummies. There appear to be no effects of differences in wealth or parental education levels. This shows that children's caste does not map completely onto wealth, or onto household education. It is important to understand what caste means for teachers, as well. As shown previously, caste does not mean observed or unobserved quality for teachers.

In summary, social distance is important for child learning outcomes. Low caste children have high returns to education, and invest in their human capital. Although teacher caste does not affect learning outcomes on its own, the interaction between the caste type of teachers and students does matter. Low caste boys benefit from being taught by high caste teachers. This seems to be particularly true in public schools, rather than in private schools. The effects are also stronger when the village is high caste dominant.

[21] The results do not differ if mother's education is used instead.

Table 6.19: Wealth and parental education

Outcome: average of math, English and Urdu test scores			
(1)		(2)	
Wealth	coef/se	Education	coef/se
Child: not poor, female.	-0.09	Child: educated father, female.	-0.014
Teacher: high caste.	(0.084)	Teacher: high caste.	(0.096)
Child: not poor, male.	0.026	Child: educated father, male.	0.091
Teacher: high caste.	(0.104)	Teacher: high caste.	(0.139)
Child: poor, female.	0.021	Child: uneducated father, female.	-0.098
Teacher: high caste.	(0.086)	Teacher: high caste.	(0.151)
Child: poor, male.	0.079	Child: uneducated father, male.	0.022
Teacher: high caste.	(0.102)	Teacher: high caste.	(0.135)
Constant	-0.43	Constant	0.523
	(0.443)		(0.560)
Number of observations	2,642	Number of observations	1,735
R^2	0.422	R^2	0.428

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Regressions include child, teacher and school controls, time dummies, child fixed effects.

Standard errors clustered at the child level.

I cannot find any evidence of discrimination on the part of teachers, nor for high caste teachers serving as role models to low caste boys. Low caste male children in public schools that are taught by high caste teachers spend more time on homework compared to when the child is taught by a low caste teacher. These boys are also significantly more likely to still be enrolled in school, and had completed more years of schooling when they had at least one high caste teacher in primary school. These results are suggestive of high caste teachers increasing the returns to education for low caste students, as low caste male children work harder when taught by them in order to get access to educational and employment benefits that teachers are able to provide for them through patronage networks.

6.7 Conclusion to Part II

Part II of this thesis has sought to understand what role caste plays in the learning outcomes of children in Punjab, Pakistan, and contributes to the literature on social distance in education. It attempts to identify the effect of child caste, teacher caste, and the interaction between child and teacher caste on learning outcomes. Many measures are employed in order to reduce bias from non-random sorting of children and teachers both to schools and within schools. Six possible mechanisms of how caste differences between teachers and students may affect learning outcomes are then explored: social dynamics in schools, returns to education, teacher quality, discrimination, role model effects, and patronage networks.

I find that although in the cross-section, low caste children have higher test scores than high caste children, within schools this is not the case. However, low caste children attend better schools (and public schools in particular) because they have high returns to education. In addition, differences between high and low caste children do not vary with the caste composition of the school. Teacher caste on its own does not matter for learning outcomes. It is not the case that high caste teachers are higher quality teachers; low caste children outperform their high caste peers in the same class. The interaction between teacher and child caste does matter for learning outcomes. I find that low caste male children perform significantly better when taught by high caste teachers compared to low caste teachers, and that this effect is only present in public schools. In addition, this effect is stronger in high caste dominant villages. It is not the case that parents are investing more in children taught by high caste teachers. It is the children that are working harder. Low caste boys spend significantly more time on homework when taught by a high caste teacher, and low caste boys that had at least one high caste teacher are more likely to still be enrolled in school, or if they drop out, have completed almost one additional grade, compared to boys that were not taught by a high caste teacher. The paper finds no evidence of discrimination by teachers in perceptions of child ability, or of role model effects. Furthermore, the positive effect of high caste teachers on low caste boys does not differ by caste composition of the school. Finally, caste does not map perfectly either onto inequalities in wealth or in parental education levels. The results of this study suggest that caste is a characteristic of children and teachers that matters for learning in rural Punjab. The distribution of teachers and students across schools, which determines which teachers students are exposed to, can have important implications for reducing gaps in educational opportunities in low-income countries.

Another implication of the results in this chapter regards labour markets in rural Punjab. There are two points to note. First, almost 90% of teachers in the data are high caste, and most low caste children (85%) are exposed to at least one high caste teacher. As a result, it is not clear that reallocating high caste teachers to low caste children would make much of a difference to learning outcomes, even though teacher caste does not matter for high caste children and does matter for low caste children. Second, private schools have significantly higher learning outcomes than public schools, including for low caste children, but low caste children are more likely to attend public school. If it is the case that low caste children (or their parents) are choosing public schools so that their teachers can help them in the future and/or because they cannot afford private education, then they may be compromising on a better education in exchange for a better social network. Subsidies to private schooling (such as vouchers or conditional or unconditional cash transfer programmes) may help with credit constraints. However, measures to help low caste children access government eduational benefits and find formal sector work could also be useful. These two points could indicate that the way in which both teaching and other formal sector jobs are sought and allocated in rural Punjab could be more efficient.

6.8 Appendix Chapter 6 - Additional Tables

Table 6.20: List of high and low caste groups

Caste Name	High/Low
Aarain	High
Abbasi	High
Ansari	Low
Awaan	High
Baloch	High
Butt	High
Chachar	High
Gujjar	High
Jat	High
Laar	High
Mohana	Low
Mughal	High
Muslim Sheikh	Low
Naich	High
Pathan	High
Qureshi	High
Rajput	High
Rehmani	Low
Samija	High
Sheikh	Low
Solangi	Low
Syed	High

Table 6.21: Observed and Simulated Data

Within Villages					
Observed	Mean	Std. Dev.	Min	Max	Observations
Teacher high caste, child high caste.	5.299	2.999	0	14	314
Teacher high caste, child low caste.	1.511	1.511	0	8	314
Teacher low caste, child high caste.	0.511	1.024	0	5	314
Teacher low caste, child low caste.	0.354	1.256	0	11	314
Simulated	Mean	Std. Dev.	Min	Max	Observations
Teacher high caste, child high caste.	5.253	2.948	0	14	314
Teacher high caste, child low caste.	1.537	1.487	0	8	314
Teacher low caste, child high caste.	0.546	0.973	0	4.352	314
Teacher low caste, child low caste.	0.323	1.116	0	11	314
Within Schools					
Observed	Mean	Std. Dev.	Min	Max	Observations
Teacher high caste, child high caste.	5.034	4.151	0	21	266
Teacher high caste, child low caste.	1.368	2.101	0	9	266
Teacher low caste, child high caste.	0.485	1.377	0	9	266
Teacher low caste, child low caste.	0.331	1.35	0	12	266
Simulated	Mean	Std. Dev.	Min	Max	Observations
Teacher high caste, child high caste.	5.008	4.165	0	21	266
Teacher high caste, child low caste.	1.351	2.072	0	9.347	266
Teacher low caste, child high caste.	0.435	1.153	0	7.526	266
Teacher low caste, child low caste.	0.301	1.194	0	12	266

Table 6.22: Teachers matching to schools

	(1)		(2)	
	Outcome: High caste teacher			
	OLS		Probit	
	coef	se	coef	se
Private school	-0.097	0.059	-0.819	0.789
Facilities index	-0.474	0.34	-5.736	4.246
Medium of teaching - Punjabi	-0.018	0.054	0.575	0.775
Medium of teaching - Pashto	-0.101	0.092	-0.459	0.985
Medium of teaching - Sindhi	-0.037	0.059	0.228	0.736
Medium of teaching - Seriaki	-0.049	0.103	-0.209	1.049
Medium of teaching - Other	-0.135	0.12	-0.556	1.245
Number of male teachers	0.002	0.007	0.089	0.074
Number of female teachers	-0.003	0.008	0.054	0.07
Building - Rented	0.038	0.072	0.867	0.64
Building - Government	-0.186***	0.066	-1.962***	0.677
Building - Donated	-0.06	0.056	(dropped)	
Building - Other	-0.219*	0.119	0.929	1.265
Number of students	0.000	0.000	-0.001	0.002
Number of govt. schools within 5 mins	-0.037*	0.021	-0.476**	0.202
Number of Islamic schools within 5 mins	-0.018	0.027	-0.065	0.21
Number of private schools within 5 mins	0.004	0.026	0.018	0.229
Number of govt. schools within 5-15 mins	0.004	0.013	-0.152	0.212
Number of Islamic schools within 5-15 mins	-0.011	0.013	-0.519***	0.173
Number of private schools within 5-15 mins	0.002	0.002	0.258	0.187
Number of govt. schools greater than 15 mins	-0.002	0.001	-0.104	0.072
Number of Islamic schools greater than 15 mins	0.001	0.001	0.094	0.075
Number of private schools greater than 15 mins	0.001	0.001	0.000	0.018
Time from school to nearest phone	0.059	0.036	0.695*	0.378
Time from school to nearest bank	-0.012	0.025	-0.152	0.291
Time from school to nearest health facility	-0.015	0.019	-0.315	0.241
Time from school to nearest transport facility	-0.024	0.023	-0.374	0.289
Time from school to nearest government office	-0.001	0.001	-0.083	0.203
School has a library	-0.339	0.236	-4.268	2.948
School has a computer	-0.500*	0.297	-5.748	3.698
School has a place for sports	-0.4	0.291	-5.069	3.586
School has a hall	-0.316	0.226	-4.522	2.835
School has walls surrounding it	-0.338	0.243	-4.191	2.936
School has fans	-0.574	0.4	-6.941	5.123
School has electricity	-0.605	0.42	-7.66	5.311
Toilets - latrine	-0.081	0.086	-1.528	0.975
Toilets - flush	-0.145	0.12	-2.031	1.541
Toilets - septic tank	-0.354	0.222	-3.781	2.408
Toilets - other	-0.354	0.233	-4.562	3.087

Continued on next page

Table 6.22 – *Continued from previous page*

Outcome: High caste teacher				
	(1)		(2)	
	OLS		Probit	
	coef	se	coef	se
Water - outside school	-0.1	0.119	-1.381	1.243
Water - tap/well inside school	-0.169	0.175	-2.028	2.129
Water - pump inside school	-0.314	0.249	-3.816	3.002
Water - official water supply	-0.472	0.34	-5.014	3.986
Sitting arrangement - mats	-0.066	0.085	-0.904	1.197
Sitting arrangement - desks and chairs	-0.178	0.153	-1.968	1.911
Sitting arrangement - desks/chairs/mats	-0.225	0.202	-2.437	2.635
Sitting arrangement - other	-0.255	0.268	(dropped)	
School is high caste dominant	0.104*	0.057	1.055**	0.47
Contracts for teachers	-0.04	0.05	-0.599	0.585
School awards teachers bonuses	0.091**	0.038	0.846**	0.426
School has an SMC	0.051	0.048	0.915	0.557
Constant	3.146**	1.552	28.585	19.284
Number of observations	1,662		627	
R^2	0.115		0.455	

Notes: $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Regressions include time dummies and village fixed effects.

Standard errors clustered at the village level.

1.0

Table 6.23: Lagged test scores and switching between high and low caste teachers

	All children			High caste male			High caste female			Low caste male			Low caste female		
	coef/se	**coef/se**	**coef/se**	**coef/se**	**coef/se**	**coef/se**	**coef/se**	**coef/se**	**coef/se**	**coef/se**	**coef/se**	**coef/se**	**coef/se**	**coef/se**	**coef/se**
Lagged average test scores	-0.003	-0.026	-0.011	0.009	-0.046	-0.019	-0.026*	0.003	0.002	0.009	-0.054	-0.026	-0.015	-0.032	0.011
	(0.009)	(0.021)	(0.012)	(0.013)	(0.038)	(0.020)	(0.014)	(0.027)	(0.018)	(0.028)	(0.048)	(0.042)	(0.033)	(0.024)	(0.023)
Age	-0.003	-0.005	-0.004	0.001	-0.038	-0.01	-0.003	0.02	0.002	-0.002	-0.002	0.001	-0.031*	-0.007	-0.012
	(0.005)	(0.014)	(0.007)	(0.008)	(0.024)	(0.010)	(0.007)	(0.019)	(0.011)	(0.014)	(0.031)	(0.025)	(0.016)	(0.038)	(0.013)
Female	0.100***		0.038												
	(0.025)		(0.029)												
Child high caste	-0.047***		-0.008												
	(0.018)		(0.016)												
Grade 3	0.054	-0.167	0.952***	-0.094		0.596	-0.015	3.308	-0.025	-0.012	-12.687**	0.121	0.323**	-0.244	0.761***
	(0.071)	(0.211)	(0.196)	(0.107)		(0.418)	(0.050)	(9.315)	(0.119)	(0.078)	(6.095)	(0.313)	(0.161)	(0.370)	(0.206)
Grade 4	0.053	-0.054	0.970***	-0.022	-0.145	0.643	-0.122	4.765	-0.059	-0.005	-0.042	-0.012	0.287**		0.886***
	(0.052)	(0.121)	(0.151)	(0.069)	(0.129)	(0.395)	(0.079)	(13.974)	(0.148)	(0.023)	(0.032)	(0.039)	(0.117)		(0.177)
Grade 5	-0.006	0.019	0.985***	-0.042	0.005	0.665*	-0.128	4.89	0.005	-0.073		0.49	0.225		0.218
	(0.039)	(0.081)	(0.131)	(0.053)	(0.142)	(0.393)	(0.087)	(13.971)	(0.135)	(0.116)		(0.535)	(0.267)		(0.364)
Grade 6	-0.025	0.149*	1.038***	-0.06	0.1	0.796**	0.003	0.017	0.008	-0.006*	-0.015*	0	0.015	-0.045**	-0.026
	(0.054)	(0.089)	(0.136)	(0.069)	(0.151)	(0.381)	(0.011)	(0.017)	(0.012)	(0.003)	(0.008)	(0.013)	(0.028)	(0.021)	(0.028)
Household asset index	-0.001	0	-0.001	-0.008	-0.001	0	-0.098*	1.681	-0.01	-0.183	0.136	-0.587*	0.302*	-0.034	0.927***
	(0.007)	(0.010)	(0.008)	(0.009)	(0.012)	(0.011)	(0.051)	(4.663)	(0.115)	(0.123)	(0.268)	(0.341)	(0.166)	(0.537)	(0.289)
Child was held back	-0.104***	0.054	0.028	-0.058	0.1	0.024	0.001	0.012**	0.011*	-0.109**	0.058	-0.019	0.001	-0.002	0.007
	(0.029)	(0.107)	(0.051)	(0.048)	(0.124)	(0.071)	(0.002)	(0.006)	(0.006)	(0.042)	(0.049)	(0.087)	(0.005)	(0.008)	(0.013)
Teacher's age	0	0.009***	0.010*	0	0.013***	0.012	-0.227**	-0.181	-0.082	0.003*	0.141**	-0.002	-0.114		0.69
	(0.001)	(0.003)	(0.005)	(0.002)	(0.004)	(0.008)	(0.089)	(0.153)	(0.141)	(0.001)	(0.068)	(0.002)	(0.122)		(0.441)
Teacher female	-0.162***	-0.089	-0.101	-0.154***	-0.042	-0.167	-0.039*	0.031	0.024	0.057	-0.169*	0	-0.01	0.031	0.029
	(0.028)	(0.082)	(0.124)	(0.042)	(0.121)	(0.169)	(0.022)	(0.061)	(0.048)	(0.068)	(0.098)	(0.143)	(0.052)	(0.099)	(0.149)
Teacher from same village	-0.067***	-0.006	-0.004	-0.066***	0.002	0.026	-0.002*	-0.019	-0.001	-0.081	-0.321***	-0.303	0	0.189**	0.001
	(0.014)	(0.033)	(0.041)	(0.019)	(0.051)	(0.087)	(0.001)	(0.056)	(0.001)	(0.061)	(0.118)	(0.287)	(0.002)	(0.078)	(0.003)
Teaching experience	-0.001	0.001*	0	0	0.003**	0.002	0.011	0.049	0.006	0.272*	-0.252	0.16	0.074	0.580***	0.397
	(0.001)	(0.001)	(0.001)	0.000	(0.001)	(0.002)	(0.024)	(0.070)	(0.093)	(0.163)	(0.156)	(0.206)	(0.051)	(0.149)	(0.266)
Teacher education - FA	0	-0.022	-0.028	-0.039	-0.04	-0.034	0.070*	0.199**	0.118	0.033	-0.079	-0.127	0.301***	0.508***	0.332
	(0.017)	(0.041)	(0.058)	(0.026)	(0.054)	(0.093)	(0.038)	(0.099)	(0.167)	(0.145)	(0.148)	(0.102)	(0.094)	(0.136)	(0.272)
Teacher education - BA	0.023	0.06	0.001	-0.011	0.063	-0.066	0.014	0.124	0.07	0	0	0.001	0.03	0.211	0.178

Continued on next page

Table 6.23 – *Continued from previous page*

	All children			High caste male			High caste female			Low caste male			Low caste female		
	coef/se	**coef/se**	**coef/se**	**coef/se**	**coef/se**	**coef/se**	**coef/se**	**coef/se**	**coef/se**	**coef/se**	**coef/se**	**coef/se**	**coef/se**	**coef/se**	**coef/se**
	(0.023)	(0.067)	(0.117)	(0.037)	(0.135)	(0.180)	(0.035)	(0.131)	(0.178)	0.000	0.000	(0.002)	(0.062)	(0.187)	(0.284)
Teacher education - MA	0.065*	-0.007	0.089	0.076	0.056	0.22	0.082**	0.12	-0.03	0.004	-0.016	-0.006	0.143*	-0.065	0.075
	(0.038)	(0.079)	(0.123)	(0.067)	(0.139)	(0.233)	(0.039)	(0.111)	(0.044)	(0.018)	(0.036)	(0.081)	(0.086)	(0.103)	(0.150)
Private school	0.123***	0.151**	-0.003	0.140**	0.143	-0.01	0	-0.002***	-0.001	(dropped)	(dropped)	(dropped)	-0.000*	0	0
	(0.033)	(0.069)	(0.027)	(0.068)	(0.089)	(0.072)	0.000	0.000	(0.001)				0.000	(0.001)	(0.002)
School size	0	-0.001***	0	0	0	0	0.008	0.038	0.046	-0.134	12.760**	-0.229	-0.042***	0.005	0.018
	0.000	0.000	0.000	0.000	0.000	0.000	(0.010)	(0.024)	(0.037)	(0.108)	(6.149)	(0.247)	(0.015)	(0.014)	(0.024)
School facilities index	0.004	0.025**	0.027	0.01	0.033**	0.034	(dropped)	(dropped)	(dropped)	-0.378***		-0.106	(dropped)	(dropped)	(dropped)
	(0.006)	(0.012)	(0.024)	(0.008)	(0.015)	(0.029)				(0.127)		(0.202)			
Constant	0.15	0.036	0.073	0.16	0.112	0.192	0.235	-3.117	-0.056	0.477**	-0.114	-0.127	0.161	-1.919***	-0.997
	(0.105)	(0.287)	(0.309)	(0.175)	(0.387)	(0.486)	(0.148)	(8.518)	(0.447)	(0.212)	(0.689)	(0.632)	(0.223)	(0.723)	(0.664)
Fixed effects	None	Child	School	None	Child	School	None	Child	School	None	Child	School	None	Child	School
Number of observations	1,473	1,473	1,473	600	600	600	504	504	504	229	229	229	140	140	140
R2	0.081	0.122	0.649	0.088	0.209	0.678	0.109	0.236	0.688	0.167	0.315	0.756	0.334	0.67	0.919

Notes: $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Outcome is switching between a high and low caste teacher. All specifications include time dummies, and standard errors clustered at the level of the fixed effect.

Table 6.24: Child attrition

	Low caste		**High caste**	
Round	Number	Percent	Number	Percent
1	60	9.35	186	9.3
2	230	35.83	684	34.2
3	348	54.21	1,098	54.9
4	4	0.62	32	1.6
Total	**642**	**100**	**2,000**	**100**

Table 6.25: Children sorting across public and private schools

	T-test	School FE Regression
All schools	-0.076*	-0.004
Public schools	-0.143***	-0.044
Private schools	0.104	0.18*

Table 6.26: Robustness Checks

	Outcome: average of math, English and Urdu test scores	
	(1)	(2)
	Schools that pass Monte Carlo CI test	Village-years that pass the Monte Carlo χ^2 test
Child high caste, teacher high caste	0.155	-0.081
	(0.142)	(0.101)
Child low caste, teacher high caste	0.261*	0.218**
	(0.156)	(0.085)
Observations	1550	2362
R^2	0.425	0.413

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

All regressions include child, teacher and school controls, and time dummies.

Standard errors clustered at the level of the fixed effect (child).

Table 6.27: Test scores and private schools

Outcome: average of math, English and Urdu test scores		
	(1)	(2)
	High Caste Children	Low Caste Children
Private school	0.432***	0.415**
	(0.104)	(0.162)
Observations	2000	642
R^2	0.244	0.249

Standard errors in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

All regressions include child, teacher and school controls, and time dummies.

Standard errors clustered at the level of the fixed effect (village).

Table 6.28: Teachers' and Children's likelihood of being in a private school

	Outcome: private school						
	(1)			(2)		(3)	
	Low caste children			High caste teachers		Low caste teachers	
	coef	se		coef	se	coef	se
Average test scores	0.092***	0.035	Average test scores	-0.017	0.017	0.053	0.105
Age	0.028*	0.017	Age	-0.016***	0.002	-0.003	0.017
Female	-0.05	0.059	Female	0.070**	0.028	-0.109	0.184
Grade 4	-0.113	0.174	From village	0.110***	0.032	0.232**	0.101
Grade 5	-0.245**	0.098	Teaching experience	0.000	0.001	-0.058	0.069
Household asset index	-0.015	0.013	Education: FA	0.059*	0.034	0.097	0.069
Held back	0.109	0.106	Education: BA	-0.118***	0.036	-0.15	0.108
Mom uneducated	-0.022	0.097	Education: MA	-0.224***	0.056	-0.289**	0.116
Dad uneducated	0.026	0.049	Married	-0.346***	0.041	-0.289	0.273
Year=2004	0.046	0.178	Year=2004	0.025	0.018	-0.032	0.113
Year=2005	0.135	0.094	Year=2005	0.023	0.02	0.081	0.195
Year=2006	0.037	0.229	Year=2006	0.059	0.06	5.173	6.538
Constant	0.074	0.18	Constant	1.027***	0.079	0.561	0.768
Number of observations	407		Number of observations	2,268		277	
R^2	0.064		R^2	0.556		0.297	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

All regressions include village fixed effects, as well as standard errors clustered at the village level.

Table 6.29: Teachers' and Children's likelihood of being in a high caste dominant village.

	Outcome: high caste dominant village						
	(1)			(2)		(3)	
	Low caste children			High caste teachers		Low caste teachers	
	coef	se		coef	se	coef	se
Average test scores	0.046	0.039	Average test scores	-0.015	0.019	-0.056	0.07
Age	0.003	0.041	Age	0.002	0.002	-0.011	0.008
Female	0.038	0.089	Female	0.024	0.023	-0.137	0.111
Grade 4	0.026	0.154	From village	-0.02	0.035	-0.201	0.139
Grade 5	0.096	0.165	Teaching experience	0.001	0.001	0.111	0.074
Household asset index	0.060**	0.03	Education: FA	0.03	0.051	0.042	0.157
Held back	0.247***	0.09	Education: BA	0.023	0.043	0.340**	0.154
Mom uneducated	-0.023	0.089	Education: MA	-0.007	0.077	0.157	0.197
Dad uneducated	-0.011	0.056	Married	-0.038	0.033	-0.127	0.103
Year=2004	-0.105	0.14	Year=2004	-0.007	0.018	0.042	0.071
Year=2005	-0.379**	0.173	Year=2005	-0.004	0.014	0.062	0.089
Year=2006	-0.200	0.183	Year=2006	-0.051	0.068	-10.632	6.96
Constant	0.873**	0.418	Constant	0.845***	0.1	1.019***	0.299
Number of observations	399		Number of observations	2,177		274	
R^2	0.044		R^2	0.007		0.288	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

All regressions include standard errors clustered at the village level.

Chapter 7

Conclusion

This thesis has sought to contribute to two strands of the literature on teachers in low-income countries: incentivizing teacher effort, and identifying the effect of a particular teacher characteristic in the education production function. Both of these have important impacts on student learning. Part I investigates the first of these questions by studying the promotion and annual performance evaluation system for teachers in Gansu province in rural China. It is the first study of promotion-based incentive systems in the public sector. Part II focuses on a particular teacher characteristic in South Asia: caste. In Pakistan, caste influences many aspects of the economy, including education. This is the only study of the effects of caste on learning in Pakistan. Studying both of these issues contributes to our understanding of how to manage one of the world's largest workforces.

In Part I, I investigate whether the highly sophisticated system of teacher performance evaluations and promotions affects teacher effort. I use a theoretical model of promotions as an incentive device to derive testable predictions on whether and how a promotion system can be expected to affect teacher effort. Chapter 3 lays out a basic model that shows how promotions can motivate effort. Data from the Gansu Survey of Children and Families (GSCF) is then used to test whether effort is exerted by teachers in order to win promotions. The results show that teachers are indeed working hard for promotions. They respond to promotion contests in the manner predicted by the model, and average levels of teacher effort are higher in promotion contests in which wage gaps are larger, when promotion rates are neither very high nor very low, and when competition for promotions is more intense. What this shows is that promotions can be an effective way to motivate teachers. The current focus in the literature is on bonuses; promotions can be an alternative.

Chapter 4 focuses on the effort of individual teachers. It looks at how three aspects of the promotion system interact to influence effort: relative teacher ability, the number of teachers in a promotion contest, and the rules on timing of eligibility for promotion. The model of Chapter 3 is extended to allow teachers to differ by ability, and a multi-period model is added in order to allow testing for dynamic effects. Teachers need to wait a certain number of years before they are eligible for promotion to the next rank, and it is the past five performance evaluation scores that count towards the promotion decision. I find that effort is zero five and more years before a teacher is eligible, that teachers work hard in the five years leading up to promotion eligibility, and that the effort of teachers that have been eligible for many years and not promoted, suffers. I also find that larger promotion contests induce lower incentives for teachers in the extremes of the skill distribution.

These results offer many insights into how to optimally structure promotion contests. Firstly, because the performance evaluations are based on four well-defined aspects of teaching quality, and because they are conducted by a group of people, they are not easy to manipulate. Gaming of incentive systems has been a problem that has occurred in virtually all of the bonus systems for teachers with which researchers have experimented. In Gansu, the annual performance evaluation scores are strongly related to actual measures of effort. However, because the evaluation and promotion system is so complex, it is very time consuming, and may be difficult to replicate elsewhere. Perhaps there exists a middle ground that is simpler, but includes some of the strengths of the Chinese system. Secondly, waiting times of more than five years before a teacher is eligible for promotion are not optimal when only the past five evaluation scores count. As the data shows, teachers exert no effort during this time. Third, for contests with very low or very high promotion rates (for example, from intern to the first rank in Primary or Middle school, or promotion to the highest rank level in Middle school) it may be more effective to have smaller contests; perhaps at the school level, rather than at the township level.

Finally, since teachers in the highest rank levels do not exert as much effort as they used to, there may be a case for using bonuses at these levels. If bonuses are also used at lower levels, the decrease in effort incentives from a teacher receiving a positive signal of her skill needs to be kept in mind. In a similar vein, it could be useful not to make teacher evaluation scores public until teachers are eligible for promotion. Since a teacher needs either one excellent evaluation score or two good evaluation scores, teachers that achieve this goal early on in the five-year period are likely to slack off afterwards. These lessons generalise to teacher incentive systems regardless of location, and also to public sector incentive systems in general.

In Part II, I study the effect of caste on learning outcomes in Punjab, Pakistan. The unique data I have on both caste and learning is rare, and it is rich. A four-year panel of students and teachers from the Learning and Education Attainment in Punjab Schools (LEAPS) survey allows me to convincingly identify the causal effect of child caste, teacher caste, and the interaction between child and teacher caste on learning outcomes. I find that low caste children have higher test scores than high caste children, and that they attend better (especially public) schools. This is because they have higher returns to education than low caste children. The effects of child caste do not differ depending on whether a school consists mainly of high or low caste children. Teacher caste does not matter on its own. It is not the case that high caste teachers differ from low caste teachers in terms of observable or unobservable quality; low caste children outperform their high caste peers in the same class. The interaction between child and teacher caste matters. I find that low caste boys benefit significantly when they are taught by a high caste teacher compared to when they are taught by a low caste teacher. This effect is stronger in public schools, and in high caste dominant villages. This result is not due to discrimination, or to role model effects. I find that low caste boys work significantly harder when taught by high caste teachers, and that this could be due to one of the salient features of caste in Pakistan: that labour market opportunities and government benefits are mediated through patronage networks. It is possible that low caste boys work hard in order to form a social link with their teacher who could help them with educational and employment opportunities later on. What we learn from this chapter is that social distance in the form of caste matters for learning in Pakistan, but perhaps in unexpected ways. Once low caste children are enrolled in school, they outperform their high caste peers, and take steps to take secure future labour market opportunities. If this is the mechanism that is driving such differences in learning, these results can also tell us something about labour markets in rural Pakistan. Perhaps there is scope for increasing efficiency in this respect, with the ultimate goal of reducing inequality between social groups.

Further work on incentivising teachers through promotions will be essential to the design and implementation of successful incentive structures for both teachers and other civil servants. Further work on social distance in Pakistan will help to shed light on the mechanisms behind caste-based and other social inequalities in education, and on how to reduce them. Both of these lines of work are important in understanding how to manage teachers in low-income countries.

Bibliography

Akerlof, G. A. and R. E. Kranton (2010). *Identity economics: How our identities shape our work, wages, and well-being.* Princeton University Press.

Alderman, H., P. F. Orazem, and E. M. Paterno (2001). School quality, school cost, and the public/private school choices of low-income households in Pakistan. *Journal of Human Resources*, 304–326.

Andrabi, T., J. Das, and N. Karachiwalla (2012). Scope of Work, Education Fund for Sindh voucher evaluation. Technical report.

Andrabi, T., J. Das, A. Khwaja, T. Vishwanath, and T. Zajonc (2007). The Learning and Educational Attainment in Punjab Schools (LEAPS) Report. *World Bank, Washington, DC*.

Andrabi, T., J. Das, and A. I. Khwaja (2008). A dime a day: The possibilities and limits of private schooling in Pakistan. *Comparative Education Review 52*(3), 329–355.

Andrabi, T., J. Das, A. I. Khwaja, and T. Zajonc (2011). Do value-added estimates add value? Accounting for learning dynamics. *American Economic Journal: Applied Economics 3*(3), 29–54.

Aslam, M. (2009). The relative effectiveness of government and private schools in Pakistan: Are girls worse off? *Education Economics 17*(3), 329–354.

Aslam, M. and G. G. Kingdon (2012). Can education be a path to gender equality in the labour market? An update on Pakistan. *Comparative Education 48*(2), 211–229.

Aslam, M., G. G. Kingdon, and M. Söderbom (2008). Is female education a pathway to gender equality in the labor market? Some evidence from Pakistan. *Girls Education in the 21st Century*, 67.

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

Baker, G., M. J. Gibbs, and B. Holmstrom (1994b). The wage policy of a firm. *Quarterly Journal of Economics 109*(4), 921–955.

Baker, G. P., M. C. Jensen, and K. J. Murphy (1988). Compensation and incentives: Practice vs. theory. *The Journal of Finance 43*(3), 593–616.

Barro, R. J. (2000). Inequality and growth in a panel of countries. *Journal of Economic Growth 5*(1), 5–32.

Barro, R. J. (2001). Human capital and growth. *American Economic Review 91*(2), 12–17.

Besley, T., R. Pande, L. Rahman, and V. Rao (2004). The politics of public good provision: Evidence from Indian local governments. *Journal of the European Economic Association 2*(2-3), 416–426.

Bognanno, M. L. (2001). Corporate tournaments. *Journal of Labor Economics 19*(2), 290–315.

Borooah, V. K. and S. Iyer (2005). Vidya, Veda, and Varna: The influence of religion and caste on education in rural India. *The Journal of Development Studies 41*(8), 1369–1404.

Botelho, F., R. Madiera, and M. Rangel (2010). Discrimination goes to school? Racial differences in performance assessments by teachers. *Working Paper, Department of Economics, University of Sao Paolo*.

Bourguignon, F., F. H. Ferreira, and M. Menendez (2007). Inequality of opportunity in Brazil. *Review of Income and Wealth 53*(4), 585–618.

Bruns, B., D. Filmer, and H. A. Patrinos (2011). *Making schools work: New evidence on accountability reforms.* World Bank Publications.

Burgess, S., C. Propper, M. Ratto, and E. Tominey (2012). Incentives in the public sector: evidence from a government agency. *IZA Discussion Paper*.

Caeyers, B. and S. Dercon (2012). Political connections and social networks in targeted transfer programs: Evidence from rural Ethiopia. *Economic Development and Cultural Change 60*(4), 639–675.

Cameron, A. C. and P. K. Trivedi (2009). *Microeconometrics using Stata*, Volume 5. Stata Press College Station, TX.

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

Cassan, G. (2012). Identity based policies and identity manipulation: evidence from Colonial Punjab. Technical report.

Castelló, A. and R. Doménech (2002). Human capital inequality and economic growth: some new evidence. *The Economic Journal 112*(478), C187–C200.

Chaudhury, N., J. Hammer, M. Kremer, K. Muralidharan, and F. H. Rogers (2006). Missing in action: teacher and health worker absence in developing countries. *The Journal of Economic Perspectives 20*(1), 91–116.

Clotfelter, C. T., H. F. Ladd, and J. L. Vigdor (2006). Teacher-student matching and the assessment of teacher effectiveness. *Journal of Human Resources 41*(4), 778–820.

Corts, K. S. (2007). Teams versus individual accountability: solving multitask problems through job design. *RAND Journal of Economics 38*(2), 467–479.

Das, J., P. Pandey, and T. Zajonc (2006). Learning levels and gaps in Pakistan. *World Bank Policy Research Working Paper* (4067).

Dee, T. S. (2005). A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review 95*(2), 158–165.

Deininger, K. and L. Squire (1998). New ways of looking at old issues: inequality and growth. *Journal of Development Economics 57*(2), 259–287.

Dīn, M. M. (2001). *Bahawalpur State Gazetteer with Map 1904*. Sang-e-Meel Publications.

Ding, W. and S. F. Lehrer (2001). The optimal policy to reward the value added by educators: Theory and evidence from China. In *International Conference on Education Reform in China, Harvard University, Cambridge, Mass. Processed.*

Dixit, A. (1997). Power of incentives in private versus public organizations. *American Economic Review 87*(2), 378–382.

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

Dostie, B. and R. Jayaraman (2006). Determinants of school enrollment in Indian villages. *Economic Development and Cultural Change 54*(2), 405–421.

Dreze, J. and G. G. Kingdon (2001). School participation in rural India. *Review of Development Economics 5*(1), 1–24.

Duflo, E. (2001). Schooling and labor market consequences of school construction in Indonesia: Evidence from an unusual policy experiment. *American Economic Review 91*(2), 785–814.

Duflo, E., R. Hanna, and S. P. Rya (2012). Incentives work: Getting teachers to come to school. *American Economic Review 102*(4), 1241–1278.

Fafchamps, M. and J. Labonne (2012). Nepotism and punishment: The (mis-) performance of elected local officials in the Philippines. Technical report, Working Paper, Oxford University.

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

Figlio, D. N. and L. S. Getzler (2006). Accountability, ability and disability: Gaming the system? *Advances in Applied Microeconomics 14*, 35–49.

Figlio, D. N. and L. W. Kenny (2007). Individual teacher incentives and student performance. *Journal of Public Economics 91*(5), 901–914.

Figlio, D. N. and J. Winicki (2005). Food for thought: the effects of school accountability plans on school nutrition. *Journal of Public Economics 89*(2), 381–394.

Gazdar, H. (2007). Class, caste or race: veils over social oppression in Pakistan. *Economic and Political Weekly*, 86–88.

Gibbons, R. and M. Waldman (1999). Careers in organizations: Theory and evidence. *Handbook of Labor Economics 3*, 2373–2437.

Gibbons, R. and M. Waldman (2003). Enriching a theory of wage and promotion dynamics inside firms. *Journal of Labor Economics 89*(1), 59–107.

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

Gibbs, M. J. (1991). An economic approach to process in pay and performance appraisals. *manuscript, University of Chicago, Graduate School of Business*.

Gibbs, M. J. (1994). Testing tournaments? An appraisal of the theory and evidence. *Labor Law Journal 45*, 493–493.

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

Gibbs, M. J. (1996). Promotions and incentives. *Unpublished working paper, University of Chicago*.

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 M. Kremer (2006). Schools, teachers, and education outcomes in developing countries. *Handbook of the Economics of Education 2*, 945–1017.

Glewwe, P. W., E. A. Hanushek, S. D. Humpage, and R. Ravina (2011). School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010. Technical report, National Bureau of Economic Research.

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

Hanna, R. N. and L. L. Linden (2012). Discrimination in grading. *American Economic Journal: Economic Policy 4*(4), 146–168.

Hanushek, E. A. and S. G. Rivkin (2006a). School quality and the black-white achievement gap. Technical report, National Bureau of Economic Research.

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

Hanushek, E. A. and L. Woessmann (2011). How much do educational outcomes matter in OECD countries? *Economic Policy 26*(67), 427–491.

Heckman, J., C. Heinrich, and J. Smith (1997). Assessing the performance of performance standards in public bureaucracies. *American Economic Review 87*(2), 389–395.

Heckman, J., C. Heinrich, and J. Smith (2002). The performance of performance standards. *Journal of Human Resources 37*(4), 778–811.

Hoff, K. and P. Pandey (2006). Discrimination, social identity, and durable inequalities. *American Economic Review 96*(2), 206–211.

Hoff, K. and P. Pandey (2011). Making up people: Experimental evidence on identity and development from caste India. *World Bank mimeo*.

Hölmstrom, B. (1979). Moral hazard and observability. *The Bell Journal of Economics*, 74–91.

Holmström, B. (1999). Managerial incentive problems: A dynamic perspective. *The Review of Economic Studies 66*(1), 169–182.

Holmstrom, B. and P. Milgrom (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *JL Econ. & Org. 7*, 24.

Ibbetson, D. B. and E. Maclagan (1911). Glossary of the tribes and castes of the Punjab and North-West Frontier Province,(ed.) Rose, H. A.

Ibbetson, S. D. (1974). *Punjab castes.* Sh. Mubarak Ali.

(India), P. (1932). *Gazetteer of the Attock District, 1930.* Sang-e-Meel Publications.

(India), P. (1996). *Gazetteer of the Chenab Colony.* Sang-e-Meel Publications.

Jackson, C. K. (2009). Student demographics, teacher sorting, and teacher quality: Evidence from the end of school desegregation. *The Journal of Labor Economics 27*(2), 213–256.

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.

Jacoby, H. and G. Mansuri (2011). Crossing boundaries: gender, caste and schooling in rural Pakistan. *World Bank Policy Research Working Paper Series*.

Kingdon, G. G. (1998). Does the labour market explain lower female schooling in India? *The Journal of Development Studies 35*(1), 39–65.

Kingdon, G. G. and S. Rawal (2010). Akin to my teacher: Does caste, religious and gender distance between student and teacher matter? Some evidence from India. *Department of Quantitative Social Science Working Papers*, 10–18.

Knoeber, C. R. and W. N. Thurman (1994). Testing the theory of tournaments: An empirical analysis of broiler production. *Journal of Labor Economics*, 155–179.

Kremer, M., N. Chaudhury, F. H. Rogers, K. Muralidharan, and J. Hammer (2005). Teacher absence in India: A snapshot. *Journal of the European Economic Association 3*(2-3), 658–667.

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.

Lavy, V. (2002). Evaluating the effect of teachers group performance incentives on pupil achievement. *Journal of Political Economy 110*(6), 1286–1317.

Lavy, V. (2003). Paying for performance: The effect of individual financial incentives on teachers productivity and students scholastic outcomes. *Jerusalem, Hebrew University*.

Lazear, E. and S. Rosen (1981). Rank- order tournaments as optimum labour contracts. *Journal of Political Economy 89*(5), 841–864.

Lazear, E. P. (1995). *Personnel Economics*, Volume 1993. The MIT Press.

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.

Li, X., A. Ahluwalia, and D. Quiang (2012). Lessons in public sector reform in China.

Lindahl, E. (2007). Gender and ethnic interactions among teachers and students–evidence from Sweden. Technical report.

Liu, S. and C. Teddlie (2005). A follow-up study on teacher evaluation in China: historical analysis and latest trends. *Journal of Personnel Evaluation in Education 18*(4), 253–272.

Luke, N. and K. Munshi (2007). Social affiliation and the demand for health services: Caste and child health in south India. *Journal of development economics 83*(2), 256–279.

MacLeod, W. B. and J. M. Malcomson (1988). Reputation and hierarchy in dynamic models of employment. *Journal of Political Economy*, 832–854.

Malcomson, J. M. (1984). Work incentives, hierarchy, and internal labor markets. *Journal of Political Economy*, 486–507.

Mankiw, N. G., D. Romer, and D. N. Weil (1992). A contribution to the empirics of economic growth. *Quarterly Journal of Economics 107*(2), 407–437.

Marrero, G. A. (2009). Inequality of opportunity and growth. *Documentos de Trabajo (FEDEA)* (24), 1–40.

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

Medoff, J. L. and K. G. Abraham (1981). Are those paid more really more productive? The case of experience. *Journal of Human Resources*, 186–216.

Milgrom, P. and J. Roberts (1988). An economic approach to influence activities in organizations. *American Journal of Sociology*, S154–S179.

Milgrom, P. R. (1988). Employment contracts, influence activities, and efficient organization design. *Journal of Political Economy*, 42–60.

Mirrlees, J. (1974). Notes on welfare economics, information and uncertainty. *Essays on economic behavior under uncertainty, North-Holland, Amsterdam*, 243–261.

Mirrlees, J. A. (1976). The optimal structure of incentives and authority within an organization. *The Bell Journal of Economics*, 105–131.

Mohmand, S. and H. Gazdar (2007). Social structures in rural Pakistan. *Thematic paper prepared under TA4319, Determinants and Drivers of Poverty Reduction and ADBs Contribution in Rural Pakistan. ADB, Islamabad.*

Munshi, K. and M. Rosenzweig (2006). Traditional institutions meet the modern world: Caste, gender, and schooling choice in a globalizing economy. *American Economic Review 96*(4), 1225–1252.

Muralidharan, K. (2011). Long-term effects of teacher performance pay: Experimental evidence from India. *Society For Research On Educational Effectiveness*.

Muralidharan, K. and V. Sundararaman (2009). Teacher performance pay: Experimental evidence from India. Technical report, National Bureau of Economic Research.

OECD (2010). *PISA 2009 Results: What Students Know and Can Do: Student Performance in Reading, Mathematics and Science*. 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.

Prendergast, C. and R. Topel (1996). Favoritism in organizations. *Journal of Political Economy 104*(5), 958–978.

Rivkin, S. G., E. A. Hanushek, and J. F. Kain (2005). Teachers, schools, and academic achievement. *Econometrica 73*(2), 417–458.

Rosen, S. (1982). Authority, control, and the distribution of earnings. *The Bell Journal of Economics*, 311–323.

Rosen, S. (1986). Prizes and incentives in elimination tournaments. *American Economic Review 76*, 701–715.

Sattinger, M. (1975). Comparative advantage and the distributions of earnings and abilities. *Econometrica*, 455–468.

Sen, A. (2001). *Development as Freedom.* Oxford Paperbacks.

Shavell, S. (1979). Risk sharing and incentives in the principal and agent relationship. *The Bell Journal of Economics*, 55–73.

Staiger, D. and J. Stock (1997). Instrumental variables regression with weak instruments. *Econometrica 65*, 557–586.

Todd, P. E. and K. I. Wolpin (2007). The production of cognitive achievement in children: Home, school, and racial test score gaps. *Journal of Human capital 1*(1), 91–136.

Vyborny, K. and A. Chaudhury (2012). Patronage and the poor: Evidence from rural Punjab, Pakistan. *mimeo*.

Waldman, M. (1984). Worker allocation, hierarchies and the wage distribution. *The Review of Economic Studies 51*(1), 95–109.

Wilson, J. Q. (1989). *Bureaucracy.* Arizona: Basic Books.

Zacharias, A. and V. Vakulabharanam (2011). Caste stratification and wealth inequality in India. *World Development 39*(10), 1820–1833.

Zhao, M. and P. Glewwe (2010). What determines basic school attainment in developing countries? Evidence from rural China. *Economics of Education Review 29*(3), 451–460.