10-3-2022 2:00 PM

Essays on the Economics of Education

Enrique Martin Luccioni, *The University of Western Ontario*

Supervisor: Nirav Mehta, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Economics

Abstract

My thesis consists of three chapters on the Economics of Education. In the first chapter, I take a structural approach to studying the extent to which teacher behavior and teacher interactions with students determine teaching contribution to test score growth in the classroom. Teachers' contribution may differ across classrooms, as it may depend on the types of students taught, and because teachers may adjust their effort to new contexts. The estimated model suggests that teacher and student efforts play a significant role in determining student knowledge. My findings indicate that teachers who are effective in teaching low-performing students may not be as effective teaching high-performing students. In a counterfactual, I assigned teachers assessed as highly effective according to value-added estimates, to classrooms with a high proportion of low-achieving students. The results suggest that the value-added measures overstate the expected performance of certain teachers in the reassigned classrooms. Additionally, I quantify the benefits of reassigning teachers based on their comparative advantages, effort choices, and endowments. Compared to the assignment based on the value-added specification, the new reassignment produces higher gains for the low-performing students.

In my second chapter, I estimate the potential gains of paying teachers according to varying, optimally-designed linear schemes. The empirical literature considers a variety of schemes that have been evaluated under randomized controlled trials. However, there is little evidence about their relative performance. To better understand these mechanisms behind the schemes' performance, I use a publicly-available dataset from a teacher incentive experiment in Andhra Pradesh, India, containing both individual- and group-based piece-rate bonuses. I exploit the experimental nature of the data to test for the presence of peer pressure in the group-based scheme. I first document the existence of peer pressure in the group-based scheme, which mitigates free-rider incentives. Based on this result, I estimate the structural parameters of my model to recover the optimal incentive schemes that maximize the expected value of student achievement, minus the expected payment to teachers. I find that an optimally designed group-based scheme could increase student academic achievement by about twice as much as the results obtained from an optimally designed individual-based incentive scheme.

In my third paper, I study the impact that teachers may have on the academic performance of Black and Hispanic students, which I refer to as minority students. To do so, I estimate the distribution of matching effects between teachers and minority students. These matching effects capture the teachers' ability to reduce the achievement gap between their assigned students. Then, I study the relationship between the estimated matching effects and a set of teachers' characteristics and skills. This allows me to explore what type of teachers are better suited to teach minority students. I find that teachers can have a meaningful impact on their minority students' performance: a one-standard-deviation increase in the teacher matching effect generates achievement gains of 0.05 standard deviations. I do not find a relationship between the estimated matching effects and the teachers' race. However, I find that the matching effects are higher for teachers with better control of students' behavior. This evidence suggests that how teachers teach matters in improving minority students' performance.

Keywords: Teacher Effectiveness, Educational Policy Design, Achievement Gaps, Optimal Contracts.

Summary for Lay Audience

My thesis consists of three chapters on the Economics of Education. In the first chapter, I take a structural approach to studying the extent to which teacher behavior and teacher interactions with students determine teaching contribution to test score growth in the classroom. Teachers' contribution may differ across classrooms, as it may depend on the types of students taught, and because teachers may adjust their effort to new contexts. The estimated model suggests that teacher and student efforts play a significant role in determining student knowledge. My findings indicate that teachers who are effective in teaching low-performing students may not be as effective teaching high-performing students. In a counterfactual, I assigned teachers assessed as highly effective according to value-added estimates, to classrooms with a high proportion of low-achieving students. The results suggest that the value-added measures overstate the expected performance of certain teachers in the reassigned classrooms. Additionally, I quantify the benefits of reassigning teachers based on their comparative advantages, effort choices, and endowments. Compared to the assignment based on the value-added specification, the new reassignment produces higher gains for the low-performing students.

In my second chapter, I estimate the potential gains of paying teachers according to varying, optimally-designed linear schemes. The empirical literature considers a variety of schemes that have been evaluated under randomized controlled trials. However, there is little evidence about their relative performance. To better understand these mechanisms behind the schemes' performance, I use a publicly-available dataset from a teacher incentive experiment in Andhra Pradesh, India, containing both individual- and group-based piece-rate bonuses. I exploit the experimental nature of the data to test for the presence of peer pressure in the group-based scheme. I first document the existence of peer pressure in the group-based scheme, which mitigates free-rider incentives. Based on this result, I estimate the structural parameters of my model to recover the optimal incentive schemes that maximize the expected value of student achievement, minus the expected payment to teachers. I find that an optimally designed group-based scheme could increase student academic achievement by about twice as much as the results obtained from an optimally designed individual-based incentive scheme.

In my third paper, I study the impact that teachers may have on the academic performance of Black and Hispanic students, which I refer to as minority students. To do so, I estimate the distribution of matching effects between teachers and minority students. These matching effects capture the teachers' ability to reduce the achievement gap between their assigned students. Then, I study the relationship between the estimated matching effects and a set of teachers' characteristics and skills. This allows me to explore what type of teachers are better suited to teach minority students. I find that teachers can have a meaningful impact on their minority students' performance: a one-standard-deviation increase in the teacher matching effect generates achievement gains of 0.05 standard deviations. I do not find a relationship between the estimated matching effects and the teachers' race. However, I find that the matching effects are higher for teachers with better control of students' behavior. This evidence suggests that how teachers teach matters in improving minority students' performance.

Acknowlegements

I am greatly indebted to my supervisor Nirav Mehta and my dissertation committee members, Salvador Navarro and Todd Stinebrickner, for their support, encouragement, and guidance in developing my research. Nirav taught me about the philosophy of economic research and how to think as an applied economist. I will always be grateful for his support and sense of humor, which encouraged me to be a better thinker during hard times.

I would like to thank Audra Bowlus, Joseph Mullins, and Peter Streufert for their valuable comments and recommendations. My research greatly benefited from my interaction with the members of the Applied Micro and Macro Reading Group, Sergio Ocampo Diaz, Rory Mcgee, and Baxter Robinson. I extend my gratitude to all seminar participants at the University of Western Ontario, for being a constant source of feedback. I sincerely appreciate the financial support from the Social Science Graduate Research Award Fund. Finally, I am deeply grateful to my family and friends for their patience and support. Especially my wife, Cecilia Diaz Campo.

Contents

List of Figures

List of Tables

List of Appendices

Chapter 1

Introduction

My thesis consists of three chapters on the Economics of Education. In the first chapter, I take a structural approach to studying the extent to which teacher behavior and teacher interactions with students determine teaching contribution to test score growth in the classroom. Standard value-added models of educational production usually treat teacher's contribution to test score growth as unvarying across contexts. However, teachers' contribution may be determined by their behavior and interactions with students in the classroom. Thus, I develop and estimate a model in which teachers and students interact by exerting effort to produce student knowledge. Teacher effectiveness may differ across classrooms, as it may depend on the types of students taught, and because teachers may adjust their effort to new contexts. The estimated model suggests that teacher and student efforts play a significant role in determining student knowledge. My findings indicate that teachers who are effective in teaching low-performing students may not be as effective teaching high-performing students. In a counterfactual, I assigned teachers assessed as highly effective according to value-added estimates, to classrooms with a high proportion of low-achieving students. The results suggest that the value-added measures overstate the expected performance of certain teachers in the reassigned classrooms. This result is consistent with recent findings suggesting that similar reassignment policy efforts did not achieve their main goal of promoting the academic performance of low-achieving students. Additionally, I quantify the benefits of reassigning teachers based on their comparative advantages, effort choices, and endowments. Compared to the assignment based on the value-added specification, the new reassignment produces higher gains for the low-performing students.

In my second chapter, I estimate the potential gains of paying teachers according to varying, optimally-designed linear schemes. Evidence that teachers play a key role in promoting student academic achievement has led to considerable policy interest in teacher performance pay schemes. The empirical literature considers a variety of schemes that have been evaluated under randomized controlled trials. However, there is little evidence about their relative performance. A possible deficiency in group-based schemes is the presence of the so-called *free-rider* effect: if a teacher's responsibility over the totality of student achievement decreases, the teacher may exert less effort because they have less impact on total production. However, there are also potential benefits of using group-based schemes. First, averaging teacher output reduces noise, which benefits risk-averse teachers. Second, peer pressure can increase effort. To better understand these mechanisms, I use a publicly-available dataset from a teacher incentive experiment in Andhra Pradesh, India, containing both individual- and group-based piece-rate bonuses. I exploit the experimental nature of the data to test for the presence of peer pressure in the group-based scheme. I first document the existence of peer pressure in the group-based scheme, which mitigates free-rider incentives. Based on this result, I estimate the structural parameters of my model to recover the optimal incentive schemes that maximize the expected value of student achievement, minus the expected payment to teachers. I find that an optimally designed group-based scheme could increase student academic achievement by about twice as much as the results obtained from an optimally designed individual-based incentive scheme.

In my third paper, I study the impact that teachers may have on the academic performance of Black and Hispanic students, which I refer to as minority students. To do so, I estimate the distribution of matching effects between teachers and minority students. These matching effects capture the teachers' ability to reduce the achievement gap between their assigned students. Then, I study the relationship between the estimated matching effects and a set of teachers' characteristics and skills. This allows me to explore what type of teachers are better suited to teach minority students. I find that teachers can have a meaningful impact on their minority students' performance: a one-standard-deviation increase in the teacher matching effect generates achievement gains of 0.05 standard deviations. I do not find a relationship between the estimated matching effects and the teachers' race. However, I find that the matching effects are higher for teachers with better control of students' behavior. This evidence suggests that how teachers teach matters in improving minority students' performance.

Chapter 2

The Determinants of Teaching Effectiveness: Evidence from a Model of Teachers' and Students' Interactions

2.1 Introduction

Teachers are widely believed to play a key role in promoting students' study efforts, motivation, and academic achievement (Hanushek and Rivkin, 2006; Blazar and Kraft, 2017; Jackson, 2018). Thus, it is not surprising that policymakers have turned their focus to interventions targeting teachers. For instance, vast resources have been allocated to measuring individual teachers' contributions to test score growth, a measure commonly referred to as teachers' *effectiveness* (Kane et al., 2013). These measures have then been used to support students' learning using multiple policy levers, such as assigning teachers measured as highly effective to classrooms with below-average test scores. However, contrary to what was expected, recent findings show that some of these policies' efforts did not achieve their main goal of promoting the academic achievement of low-performing students (Glazerman et al. 2013; Stecher et al. 2018). These findings suggest that commonly used measures of teachers' effectiveness may not be as useful for certain policies as previously thought.

The standard value-added models of teachers' effectiveness treat teachers as an educational input that can be considered invariant across contexts such as classrooms (Jackson et al., 2014; Koedel et al., 2015). However, teachers' contributions to students' learning may be determined by what teachers actually do in the classroom; thus, teachers' effectiveness may not be a fixed attribute. In fact, recent evidence suggests that how teachers teach (i.e., what they do in the classroom) plays an essential role in promoting students' academic achievement (Garrett and Steinberg, 2015; Taylor, 2018; Aucejo et al., 2021). Thus, not accounting for the decisions made in the classroom may render extrapolations— which are crucial for designing effective policies—problematic. In particular, teachers' effectiveness may not be invariant to commonly suggested interventions, such as transferring highly effective teachers to poor-performing schools or replacing the least effective teachers (Hanushek, 2011; Chetty et al., 2014b).

In this paper, I take a structural approach to study the extent to which teachers' behavior and interactions with students may determine teachers' effectiveness in a classroom. I develop and estimate a model in which teachers and students interact by exerting effort to produce knowledge. Teachers' effectiveness may differ across classrooms because it may depend on the types of students being taught and because teachers may adjust their effort. The model nests standard value-added specifications, which allows me to study whether commonly used measures of teachers' effectiveness are enough to inform policymakers and how much could be gained by providing a cohesive framework to study its determinants. Doing so would enable a better understanding of how teachers would respond to new education policies and, consequently, the extent to which students' academic achievement would be affected, which is crucial for designing effective interventions.

Prior research acknowledges the heterogeneity in teachers' effectiveness but cannot comprehensively evaluate how it varies across contexts for several reasons. First, students' achievement may be determined by educational inputs, such as teachers' and students' efforts, that are difficult to measure and, as such, are not commonly found in databases. However, it may be important to assess the relative importance of these inputs to separate teachers' contributions to students' achievement from other possible confounding factors, which are typically unobserved or mismeasured. Second, recent research using direct classroom observation has studied the effects of different teaching practices on students' achievement, yet very little research considers how teachers make their effort decisions to implement these practices. It may be important to consider these decisions because the amount of effort exerted by teachers may change across classrooms, likewise changing their contributions to students' learning. Finally, some teachers may be more effective in classrooms with low-achieving students, while others may be more effective in classrooms with high-achieving students. Not accounting for these comparative advantages may miss important dimensions of teaching that could better direct teachers' assignments.

To address these challenges, the model takes into account several interrelated components. First, it allows for flexible distributions of teachers' and students' endowments. In particular, teachers may differ in their general teaching skills and their per-unit (i.e., marginal) effort costs. Students may differ in their stocks of initial knowledge and their home environments. Furthermore, I model a flexible distribution of the classroom environment, which allows some classrooms to be more prone to disruptions than others, regardless of teachers' effort. I allow teachers', classrooms', and students' endowments to be flexibly correlated. Second, teachers contribute to the production of students' knowledge with their general teaching skills and by exerting effort to keep the class on task and well-behaved, which is a teaching *practice* (i.e., an intermediate input to knowledge production) that has been shown to be effective in promoting students' achievement (Kane et al., 2014). The amount of effort exerted by teachers may depend on their preferences for promoting classroom knowledge. The effects of the teaching inputs may differ across classrooms due to changes in students' and classrooms' endowments. Third, students contribute to their knowledge gain with their initial stocks of knowledge and learning efforts. Finally, the teaching practice may also promote achievement by increasing students' engagement in their learning, thereby increasing students' effort. Thus, teachers who are effective motivators may have another mechanism to promote students' achievement.

Estimating such a model requires access to a rich dataset containing student-, classroom-, and teacher-level data on multiple inputs. The unique information in the Measures of Effective Teaching Longitudinal Database (METD) allows me to overcome these data challenges. I use Ronald Ferguson's 7Cs framework from the Tripod student survey to measure what teachers do in each classroom (Ferguson and Danielson, 2015). I use the rich sets of measures from the 7Cs framework to assess the impact of teachers' behavior on students' achievement. I complement these measures with measures of students' home environments and effort from the Tripod survey. I estimate the model using data on 6th grade middle school students from the first year of the METD (2009-2010). I then exploit the random assignment of teachers to classrooms in the second year of the METD to check the validity of my model to predict students' and teachers' efforts, as well as students' academic achievement in the holdout data.

The database contains multiple measurements of the educational inputs. I treat these measurements as noisy measures of underlying latent inputs. Following the factor model literature, I define a measurement system combining continuous and ordinal measures of the latent variables. I exploit the closed-form solution of the model to estimate the measurement system in one step using simulated maximum likelihood with analytical gradients, which provides fast convergence.

The estimated model reveals that teachers' and students' efforts play a key role in determining students' academic achievement. The effect of the teaching practice of keeping students on task is higher for students with a lower initial stock of knowledge. In comparison, the effect of the general teaching skill is higher for students with a higher initial stock of knowledge. For teachers, the marginal utility of increasing the classroom knowledge is higher in classrooms with higher initial stocks of knowledge. I also find that being in a well-behaved classroom substantially reduces students' utility cost of exerting effort, implying a complementarity between students' effort and the teaching practice. Overall, these results suggest that teachers' contributions to student knowledge vary across classrooms due to changes in students' and teachers' efforts and the contribution of the teaching inputs.

The estimates also suggest that teachers with higher general teaching skills or higher per-unit effort costs tend to be sorted into classrooms with higher initial stocks of knowledge and more supportive home environments. Additionally, teachers with lower per-unit effort costs tend to be sorted into classrooms with worse classroom environments. As previous research suggests, not accounting for the non-random sorting of teachers into classrooms may generate biased estimates of teachers' contributions to students' knowledge (Kane et al., 2013; Rothstein, 2017).

I use my estimates to study the impact of changing teachers' and students' endowments on academic achievement considering their behavioral responses. I find that, compared to an average teacher, teachers with higher general teaching skills are more effective in teaching students with higher initial knowledge. On the other hand, teachers with lower per-unit effort costs are more effective in teaching students with lower initial knowledge. The teachers' endowments are correlated: I find that teachers with higher general teaching skills also tend to have higher per-unit effort costs. Overall, these results suggest that teachers who are effective in teaching low-performing students may not be as effective teaching high-performing students, and vice versa. Accordingly, I expand upon standard value-added models, which usually consider a single dimension of teaching effectiveness, by taking into account the comparative advantages of teachers.

I use the model to conduct a series of policy-relevant counterfactuals. The counterfactuals are motivated by recent policy interventions that did not achieve their main goal of promoting students' academic achievement. For instance, the Talent Transfer Initiative (Glazerman et al., 2013) provided incentives to highly effective teachers, based on value-added measures, to transfer to schools with high proportions of low-achieving students. The program did not significantly increase middle school students' academic achievement in poor-performing schools.

In the first two counterfactuals, I quantify the achievement gains of reassigning teachers, who are measured as highly effective using a standard value-added specification, into classrooms with a high proportion of low-achieving students.[1] The first counterfactual reassigns the selected teachers into classrooms with below-average test scores from the previous year, while the second counterfactual reassigns these teachers into classrooms with high proportions of minority students. In the third counterfactual, I replace the least effective teachers with the most effective teachers based on their value-added estimates. Overall, the results suggest that the value-added estimates overstate the expected performance of the selected teachers in the reassigned classrooms. Moreover, I quantify the effects of reassigning teachers considering their effort choices, and the full set of endowments. Compared to the previous reassignment based on the value-added specification, the new reassignment produces higher gains for the low-performing students. For instance, the first counterfactual suggests that the reassignment using the value-added specification would reduce one quarter of the achievement gap between the low-achieving classrooms and the rest, whereas if behavioral responses were taken into account, the achievement gap would be reduced by a third.

This paper contributes to the teachers' effectiveness literature in several ways. First, a growing body of evidence suggests that teachers who use good practices can effectively increase students' achievement (Kane et al. 2011; Garrett and Steinberg 2015). The natural next step I take is studying how teachers exert effort to implement these practices. By taking this step, I can study how the teaching practices would vary across classrooms due to changes in teachers' effort and how these variations would ultimately affect students' achievement.[2]

Second, this paper relates to work studying heterogeneity in teachers' effectiveness due to the matching of teachers and students (see, for example, Loeb et al. 2014, Fox 2016, Gershenson et al. 2018, Aucejo et al. 2020, and Ahn et al. 2021). I focus on two teaching inputs that may have heterogeneous effects on students' achievement: the general teaching skill and the desirable teaching practice of keeping students on task and well-behaved. There is a new strand of the literature on this topic studying heterogeneity in the effect of teaching practices on

[1]As opposed to tracking interventions where students are sorted into new classrooms, the counterfactual reassignment keep the classroom composition fixed and match teachers to already established classrooms.

[2]Steinberg and Garrett (2016) find that the average prior achievement of the classroom positively influence measured teaching practices' scores. My model rationalizes this evidence by allowing teachers to adjust their effort to implement the teaching practice in response to changes in the classroom composition.

students' achievement and the implications for reassignment policy (Lavy, 2015; Aucejo et al., 2021; Graham et al., 2020). I expand on this literature by distinguishing the teaching practice from the endowments that may determine it. Accordingly, teachers may adjust their teaching practice to new classrooms. Thus, my framework can suggest the impact that the teaching practice may have on different classrooms and the most effective teachers to implement it.

Third, there is an ongoing debate about the validity of value-added estimates predicting teachers' effects on students' achievement due to bias generated by unmeasured or mismeasured factors that may be correlated with the estimated teachers' effectiveness (see, for example, Chetty et al. 2014a, Rothstein 2017, and Horvath 2015). For example, if some teachers are systematically assigned to disruptive classrooms (i.e., with a worse classroom environment), their estimated value-added may understate their true effectiveness in those classrooms. I address this concern by explicitly taking into account multiple inputs that are commonly unobserved by researchers, such as the students' home and classroom environments. Given that measurement error in the observed inputs may produce biased estimates, I exploit the multiple measures of the educational inputs available in the database and follow a factor model approach to separate the distribution of latent variables from measurement error (see, for example, Cunha et al. (2010) and Agostinelli and Wiswall (2016a)). Furthermore, the decisions of teachers and students are allowed to interact, allowing for flexible complementarity patterns that are not present in the traditional (linear-in-means) value-added model.

Finally, in recent years, the majority of US school districts have implemented teacher evaluation systems that contain measures based on students' test scores (Steinberg and Donaldson 2016; Steinberg and Kraft 2017). Given the extensive use of these measures in ongoing policies, it is crucial to assess the extent to which teachers' preferences for promoting students' achievement may be an important determinant of their effectiveness. Not accounting for teaching preferences may render extrapolations problematic because teachers may not be willing to exert the same amount of effort in all classrooms. To address this, my model allows teachers' preferences for academic achievement to depend on the level of initial knowledge in the classroom.

My model is closest in spirit to that of Todd and Wolpin (2018), which also models the effort choices of teachers and students in a classroom. The authors develop and estimate the model to better understand the reasons for the poor performance of high school students in Mexico. My model is designed to study how teachers' effectiveness may vary across classrooms, and whether commonly used measures are capable of capturing it. As such, my model nests standard specifications used in the teaching effectiveness literature, such as the standard value-added model, and allows teachers' contribution to students' knowledge to differ across classrooms. Additionally, I allow for flexible distributions of commonly unobserved endowments to capture potential non-random sorting of teachers to classrooms.

The remainder of the paper is organized as follows. Section 2.2 presents the model. Section 2.4 discuss identification. Section 2.3 discuss potential limitations of the standard value-added model. Section 2.5 discuss the estimation method. Section 2.6 describes the data and estimation sample. Section 2.7 discusses estimation results. Section 2.8 presents a comparative statics analysis. Section 2.9 present the results from the counterfactuals. Finally, section 2.10 concludes.

2.2 The Model

In this section, I present a model of the interaction between teachers and students in a classroom to produce knowledge. The model nests commonly used value-added specifications of educational production. Additionally, it allows for flexible distributions of teachers' and students' endowments to capture the potential non-random sorting of teachers to classrooms. Teachers' and students' decisions are the outcomes of a sequential Nash game with complete information where teachers internalize the students' actions. Teachers choose effort to maximize utility from their students' achievement, net the cost of effort of keeping the classroom well-behaved and on task. After observing the teacher's effort, students exert effort to maximize utility from achievement, net the cost of learning effort. Each classroom is a separate game.

2.2.1 Basic Setup

Consider teacher j in classroom c, which is endowed with a continuum of students of measure N_{cj}. Teacher j has two types of endowment: A general teaching skill, θ_{Gj}, and a per-unit cost of effort, θ_{Ej}. I allow for a flexible distribution of the teacher endowments, by treating them as a set of teacher-specific parameters, which may be correlated with the students' and classrooms' characteristics.

Student i is endowed with an initial stock of knowledge, K_{icj}, and a home environment, H_{icj}.[3] All endowments are observed by the teacher and the students. Students' endowments are formed by a structure that depends on a vector of exogenous student variables, X_{icj}, and unobservable shocks to the econometrician, ξ_{Kicj} and ξ_{Hicj}:[4]

$$\begin{aligned} \ln K_{icj} &= \mu_{Kcj} + X_{icj}\beta_K + \ln \xi_{Kicj} \\ \ln H_{icj} &= \mu_{Hcj} + X_{icj}\beta_H + \ln \xi_{Hicj}. \end{aligned} \tag{2.1}$$

The unobserved shocks, ξ_{Kicj} and ξ_{Hicj}, may be correlated with each other but are uncorrelated with the observed variables and teachers' skills. In particular, $\ln \xi_{Kicj}$ and $\ln \xi_{Hicj}$ follow a joint normal distribution with mean zero and a variance-covariance matrix, Λ. To capture the potential nonrandom sorting of students to teachers, the classroom-specific intercepts, μ_{Kcj} and μ_{Hcj}, are allowed to be arbitrarily jointly distributed with the observed student variables and the teachers' endowments.

Teacher j chooses managerial effort, E_{Mcj}, to keep the class on task and well-behaved. However, some classrooms may be harder to manage than others, regardless of the teacher's effort, due to inherently bad classroom environments. Therefore, I model the classroom management practice, M_{cj}, as the *effective input* provided by the teacher j in classroom c, which may depend on the teacher's managerial effort and the classroom environment, μ_{Mcj}, and is

[3] The home environment captures characteristics associated with the cost of students exerting effort. Similar to Agostinelli et al. (2019), the home environment includes measures such as the number of books and computers in the household and parents' education.

[4] The exogenous student variables are student age, race, gender, special education status, gifted status, English learner status, and twice-lagged test score.

given by

$$M_{cj} = \mu_{Mcj} E_{Mcj}^{\delta_E}. \tag{2.2}$$

I use the number of students in the classroom as a measure of the classroom environment: following the work of Lazear (2001), all else equal, a larger classroom may be prone to more disruptions, which in turn may adversely affect student learning.[5] The classroom environment is allowed to be arbitrarily jointly distributed with the students' and teachers' endowments.

Student i's achievement is a function of the student's initial stock of knowledge, K_{icj}, the student's learning effort, L_{icj}, the classroom management practice, M_{cj}, the teacher j's general skill, θ_{Gj}, and the average of the initial stock of knowledge in classroom c, $\bar{K}_{cj}$. The production function is given by

$$Y_{icj} = K_{icj}^{\alpha_K} \bar{K}_{cj}^{\alpha_{\bar{K}}} L_{icj}^{\alpha_L} M_{cj}^{\alpha_{Micj}} \theta_{Gj}^{\alpha_{Gicj}}. \tag{2.3}$$

I allow for the parameters governing the marginal product of the classroom management practice and general teaching skill to be affected by the initial stock of knowledge, by setting

$$\begin{aligned} \alpha_{Micj} =& \alpha_M + \alpha_{MK} \ln K_{icj}, \\ \alpha_{Gicj} =& 1 + \alpha_{GK} \ln K_{icj}. \end{aligned} \tag{2.4}$$

All else equal, a negative coefficient value of α_{MK} and α_{GK} implies a lower marginal product of M_{cj} and θ_{Gcj} for students with a higher stock of initial knowledge, while the opposite would happen if the interactions terms were positive. In this regard, it could be that two different students have the same level of classroom management practice, but if $\alpha_{MK} < 0$, the student with the lower initial stock of knowledge may benefit more from being more focused due to a better-behaved classroom (i.e., due to an increase in M_{cj}).

The utility cost of studying depends on learning effort and may be determined by the student's home environment, H_{icj}, and classroom management practice, M_{cj}:

$$C_{icj} = c_L H_{icj}^{\beta_H} M_{cj}^{\beta_M} L_{icj}. \tag{2.5}$$

A better home environment and classroom management practice may increase students' engagement in their learning, which in turn may reduce the utility cost of learning.[6] Therefore, reducing the utility-cost of learning represents a natural channel through which teachers could improve students' achievement.

After observing the teacher's effort, student i chooses learning effort, L_{icj} to maximize the utility from achievement minus the utility cost of effort, C_{icj}

$$U^s(L_{icj}) = \frac{Y_{icj}^{\gamma_L} - 1}{\gamma_L} - C_{icj}, \tag{2.6}$$

[5]As I further discuss in Section 2.6.1, I measure teachers' effort using survey questions that directly reference the teacher behavior in the classroom, such as "My teacher doesn't let people give up when the work gets hard.", while for the management practice, I choose survey questions that relate to the effective impact of teachers' effort in the classroom, such as "student behavior in this class is under control".

[6]One would expect home and classroom environment to reduce the cost of learning, i.e., $\beta_H < 0$ and $\beta_M < 0$, but I do not impose this restriction in my estimation.

where the students' preference for achievement, $\gamma_L \in (-\infty, 1]$, allows for a flexible substitution pattern between inputs: as I further discuss in Section 2.2.2, all else equal, this specification allows the utility-maximizing learning effort to decrease in the student's initial stock of knowledge and the teacher's general teaching skill when $\gamma_L < 0$. Furthermore, if $\gamma_L < 0$, the utility-maximizing learning effort may decrease in the classroom management practice if γ_L is sufficiently low to outweigh the effect of the practice on students' utility cost of effort.

Teachers maximize their utility from average classroom achievement minus their cost of effort:

$$U^t(E_{Mcj}) = \frac{\bar{Y}_{cj}^{\gamma_{E_{cj}}} - 1}{\gamma_{E_{cj}}} - \theta_{Ej} E_{Mcj}, \tag{2.7}$$

where $\bar{Y}_{cj}$ is the geometric mean of students' achievement considering students' best responses to teacher j's effort.[7] I allow for additional flexibility by letting the teachers' preference for student achievement, $\gamma_{E_{cj}}$, to depend on the average initial stock of knowledge in the classroom:

$$\gamma_{E_{cj}} = \gamma_E + \gamma_{EK} \ln \overline{K_{cj}}. \tag{2.8}$$

For instance, teachers may receive higher utility from improving the performance of high-achieving classrooms: all else equal, a positive coefficient, $\gamma_{EK} > 0$, implies that teacher j's marginal utility from increasing the classroom average achievement, $\bar{Y}_{cj}$, would be higher in classrooms with higher average initial stocks of knowledge. Conversely, teachers may be interested in reducing the achievement gap across classrooms, and so, have higher marginal benefits of increasing the knowledge of classrooms with lower initial stocks of knowledge, which would be captured by a negative coefficient, $\gamma_{EK} < 0$. Additionally, as shown in Appendix A.1, this specification allows teachers' efforts to decrease in response to changes in the home and classroom environments, as well as their general teaching skills when $\gamma_{E_{cj}} < 0$.

2.2.2 Solution

The equilibrium concept is sub-game perfect Nash equilibrium. The timing of the model is as follows: In stage 1, teacher j chooses the level of managerial effort for classroom c, E_{Mcj}. In stage 2, students choose the level of learning effort, L_{icj}. Finally, in stage 3, students' achievement is produced.

In stage 2, students' best responses can be obtained from their first-order condition. At the interior solution for learning effort, the marginal benefit must equal the marginal cost of effort. As shown in Appendix A.1, the sufficient condition for an interior solution for student effort requires that $\alpha_L \gamma_L < 1$.[8]

[7] The geometric mean, $\bar{Y}_{cj} = \exp\big(E(\ln Y_{icj})\big)$, generates a simple closed-form solution, which is useful in the estimation of the model.

[8] The sequential game assumption rules out zero-effort equilibria. This assumption is consistent with the data as more than 90% of the students in the sample report studying more than 0 hours per week and complete part of their weekly homework.

Student i's best response to the classroom management practice is given by the function

$$\ln L_{icj}^{br} = \frac{\ln\left(\frac{\alpha_L}{c_L}\right) + \gamma_L\alpha_K \ln K_{icj} + \gamma_L\alpha_{\bar{K}} \ln \bar{K}_{cj} + \gamma_L\alpha_{G_{icj}} \ln \theta_{Gj}}{1-\alpha_L\gamma_L} + \frac{(\gamma_L\alpha_{M_{icj}} - \beta_M)\ln M_{cj} - \beta_H \ln H_{icj}}{1-\alpha_L\gamma_L} \tag{2.9}$$

Students may respond to the classroom management practice through production (i.e., $\alpha_{M_{icj}} \ln M_{cj}$) and the utility cost of effort (i.e., $-\beta_M \ln M_{cj}$). The overall response to the classroom management practice also depends on the curvature parameter γ_L. If $\gamma_L < 0$ and the utility-cost parameter β_M is not too negative, students may reduce their effort in response to a better classroom management practice (i.e., $\frac{\partial \ln L_{icj}^{br}}{\partial \ln M_{cj}} = \frac{\gamma_L\alpha_{M_{icj}} - \beta_M}{1-\alpha_L\gamma_L} < 0$). Similarly, if $\gamma_L < 0$ and $\alpha_K > 0$, all else equal, knowledgeable students would study less than students with lower initial knowledge (i.e., $\frac{\partial \ln L_{icj}^{br}}{\partial \ln K_{icj}} = \frac{\gamma_L\alpha_K}{1-\alpha_L\gamma_L} < 0$). Additionally, students may respond differently to increases in the teaching inputs due to the heterogeneity in the parameters governing the marginal products of the teaching inputs (e.g., $\frac{\partial^2 \ln L_{icj}^{br}}{\partial \ln \theta_{Gj} \partial \ln K_{icj}} = \frac{\gamma_L\alpha_{GK}}{1-\alpha_L\gamma_L} > 0$, if $\alpha_{GK} > 0$ and $\gamma_L > 0$).

In stage 1, the teacher maximizes their utility by exerting managerial effort. The first order condition is derived from Equation (2.7).[9] At the interior solution for managerial effort, the marginal benefit must equal the marginal cost of managerial effort, which is given by the per-unit cost of effort, θ_{Ej}. In equilibrium, the teacher's managerial effort is given by the known function

$$E_{Mcj}^* = \left[\left(\frac{1}{\theta_{Ej}} \times \delta_E \frac{\alpha_{Mcj} - \alpha_L\beta_M}{1-\alpha_L\gamma_L}\right)^{1-\alpha_L\gamma_L} \times \left(\frac{\alpha_L^{\alpha_L} \times \bar{K}_{cj}^{\alpha_K+\alpha_{\bar{K}}} \times \theta_{Gj}^{\alpha_{Gcj}} \times \mu_{Mcj}^{\alpha_{Mcj}-\alpha_L\beta_M}}{c_L^{\alpha_L} \times \bar{H}_{cj}^{\alpha_L\beta_H}}\right)^{\gamma_{Ecj}}\right]^{\frac{1}{1-\alpha_L\gamma_L-\delta_E\gamma_{Ecj}(\alpha_{Mcj}-\alpha_L\beta_M)}} \tag{2.10}$$

where $\bar{K}_{cj}$ and $\bar{H}_{cj}$ are, respectively, the geometric means of students' knowledge and home environment, averaged at the classroom level, and $\alpha_{Mcj} = \alpha_M + \alpha_{MK} \ln \bar{K}_{cj}$, and $\alpha_{Gcj} = 1 + \alpha_{GK} \ln \bar{K}_{cj}$.[10] As seen in Equation (2.10), teacher j's effort may vary across classrooms due to changes in the classroom endowments: $\bar{K}_{cj}$, $\bar{H}_{cj}$, and μ_{Mcj}. However, the model is flexible enough to allow the teacher effort to be constant across classrooms, by setting $\gamma_{EK} = 0$ and $\alpha_{MK} = 0$.

The closed-form classroom management practice is then

$$M_{cj}^* = \mu_{Mcj} E_{Mcj}^{*\delta_E}. \tag{2.11}$$

[9] As shown in Appendix A.1, the sufficient condition for an interior solution fo teacher effort is given by: $\alpha_L\gamma_L + \delta_E\gamma_{Ecj}(\alpha_{Mcj} - \alpha_L\beta_M) < 1$, where $\alpha_{Mcj} = \alpha_M + \alpha_{MK} \ln \bar{K}_{cj}$.

[10] The geometric mean of students' achievement, $\bar{Y}_{cj}$, in the teachers' utility function, Equation (2.7), makes teachers' effort choices depend on classroom-level and teacher-level endowments only. The continuum of students assumption implies that teachers' efforts do not depend on individual student-level shocks since $E(\ln \xi_{Kicj}) = 0$ and $E(\ln \xi_{Hicj}) = 0$.

In equilibrium, the student-level endogenous variables, L_{icj} and Y_{icj}, are given by the known functions:

$$\begin{aligned} L^* &= L(K_{icj}, H_{icj}, \bar{K}_{cj}, \bar{H}_{cj}, \mu_{Mcj}, \theta_{Ej}, \theta_{Gj}, \Gamma), \text{ and} \\ Y^* &= Y(K_{icj}, H_{icj}, \bar{K}_{cj}, \bar{H}_{cj}, \mu_{Mcj}, \theta_{Ej}, \theta_{Gj}, \Gamma), \end{aligned} \tag{2.12}$$

which are obtained by replacing Equation (2.11) in the students' best responses, Equation (2.9), and the production function, Equation (2.3). The vector Γ contains the production and behavioral parameters. The closed-form equations (2.10), (2.11), and (2.12) express teachers' and students' efforts inputs, and students' achievement as functions of exogenous variables and the structural parameters. These reduced-form equations are used for estimation in Section 2.5. The reduced-form equations would be log-linear if the parameters governing the marginal product of the classroom managements, $\alpha_{M_{icj}}$, the general teaching skills, $\alpha_{G_{icj}}$, and teachers' preferences for students' knowledge, γ_{Ecj}, were constant across students and classrooms. I study the implications of assuming log-linear specifications in Section 2.3.

2.2.3 Measurement System

I allow the achievement, efforts' choices, and endowments to be measured with error. For convenience, I group latent variables in a set $\Omega = \{K_{icj}, H_{icj}, Y_{icj}, L_{icj}, E_{Mcj}, M_{cj}, \mu_{Mcj}\}$. To simplify the notation, I omit the student, classroom and teacher sub index from Ω. The variables contained in the sets Ω are functions of observable and unobservable exogenous endowments. For example, K_{icj}, which is defined in Equation (2.1), is a function of the observed students' characteristics, X_{icj}, the latent shock, ξ_{Kicj}, and the latent effect, μ_{Kcj}, while teacher effort, E_{Mcj}, is determined by the classroom-level endowments (i.e., $\bar{K}_{cj}$, $\bar{H}_{cj}$, and μ_{Mcj}), and teacher-level endowments (i.e., θ_{Ej} and θ_{Gj}) as defined in Equation (2.10).

The data contain a rich variety of continuous and discrete measures. Let Ω_d represent the d^{th} variable in the set Ω. I denote a continuous measure for Ω_d as Z_{dm}, where m is a sub index for an specific measure of Ω_d. Each continuous measure relates to the corresponding latent variable through the following equation:

$$Z_{dm} = \kappa_{dm} + \lambda_{dm} \ln \Omega_d + \epsilon_{dm}, \tag{2.13}$$

where κ_{dm} is a location parameter, and λ_{dm} is a scale parameter. I follow the standard procedure in the literature and assume a log-linear system of equations and errors ϵ_{dm} that follow an independent normal distribution, $N(0, \sigma^2_{\epsilon,dm})$ (Cunha et al., 2010).

For the discrete measures, I consider an ordinal system of equations, where each measure Z^O_{dm}, has P categories, $p = P$ is the lowest category, and $P = 1$ is the highest:[11]

$$Z^O_{dm} = \begin{cases} P & \text{if } \lambda_{dm} \ln \Omega_d + \epsilon_{dm} < \kappa_{Pdm} \\ P-1 & \text{if } \kappa_{Pdm} < \lambda_{dm} \ln \Omega_d + \epsilon_{dm} < \kappa_{P-1dm} \\ \vdots & \\ 2 & \text{if } \kappa_{3dm} < \lambda_{dm} \ln \Omega_d + \epsilon_{dm} < \kappa_{2dm} \\ 1 & \text{if } \lambda_{dm} \ln \Omega_d + \epsilon_{dm} > \kappa_{2dm} \end{cases} \tag{2.14}$$

[11] Most of the measures have five categories ranging from "totally untrue" to "totally true".

where the κ's are unknown cutoff parameters to be estimated. As discussed in Agostinelli and Wiswall (2016a), the identification of the system of ordinal measures requires a parametric assumption for the distribution of the measurement error. For this reason, I assume $\epsilon_{d,m} \sim N(0,1)$.[12]

2.3 Implications for Teacher Reassignment

This section discusses the implications for teacher reassignment and the potential limitations of using the standard value-added specification to predict the gains of reassigning teachers to different classrooms. Consider that teacher $j = 1$, who may be replaced by teacher $j = 2$, is assigned to classroom c. For student i in classroom c, the potential achievement gain of being reassigned to teacher 2 is given by the difference in achievement due to the teachers' reassignment, $\Delta \ln Y_{ic(2,1)}$:

$$\Delta \ln Y_{ic(2,1)} = \ln F(K_i, \bar{K}_c, L_{ic2}, M_{c2}, \theta_{G2}) - \ln F(K_i, \bar{K}_c, L_{ic1}, \mu_{Mc}, M_{c1}, \theta_{G1}), \tag{2.15}$$

where the first term on the right-hand side of Equation (2.15) is the log production function of latent knowledge, Equation (2.3), if teacher 2 were assigned to classroom c, and the second term is the production function without reassignment. For student i, the gain of reassignment partially depends on the counterfactual amount of studying effort, L_{ic2}, that student i would exert under teacher 2, and the counterfactual classroom management practice, M_{c2}, that, as defined by Equation (2.2) depends on the counterfactual amount of effort that teacher 2 would exert in classroom c, and the classroom c's environment effect. Teacher 2 would be more effective than teacher 1 in teaching student i if $\Delta \ln Y_{ic(2,1)} > 0$.

Based on the production function defined in Equation (2.3), classroom c's averaged achievement gain due to the teachers' reassignment, $\Delta \ln \bar{Y}_{c(2,1)}$, can be decomposed into the gain due to changes in students' effort, $\Delta \ln \bar{L}_{c(2,1)}$, teachers' effort, $\Delta \ln E_{Mc(2,1)}$, and teachers' general teaching skills, $\Delta \ln \theta_{G(2,1)}$. The decomposition is given by

$$\Delta \ln \bar{Y}_{c(2,1)} = \underbrace{\alpha_L \Delta \ln \bar{L}_{c(2,1)}}_{\text{Change in student effort}} + \underbrace{\alpha_{Mcj}\delta_E \Delta \ln E_{Mc(2,1)}}_{\text{Change in teacher effort}} + \underbrace{\alpha_{Gcj}\Delta \ln \theta_{G(2,1)}}_{\text{Change in general teaching skill}},$$

where

$$\Delta \ln \bar{L}_{c(2,1)} = \underbrace{\frac{\gamma_L \alpha_{Gcj}\Delta \ln \theta_{G(2,1)}}{1-\alpha_L\gamma_L}}_{\text{Change in general teaching skill}} + \underbrace{\frac{(\gamma_L \alpha_{M_{cj}} - \beta_M)}{1-\alpha_L\gamma_L}\delta_E \Delta \ln E_{Mc(2,1)}}_{\text{Change in teacher effort}} \tag{2.16}$$

$$\Delta \ln E_{Mc(2,1)} = \underbrace{\frac{\gamma_{Ecj}\alpha_{Gcj}\Delta \ln \theta_{G(2,1)}}{1-\alpha_L\gamma_L - \delta_E\gamma_{Ecj}(\alpha_{Mcj}-\alpha_L\beta_M)}}_{\text{Change in general teaching skill}} - \underbrace{\frac{(1-\alpha_L\gamma_L)\Delta \ln \theta_{E(2,1)}}{1-\alpha_L\gamma_L - \delta_E\gamma_{Ecj}(\alpha_{Mcj}-\alpha_L\beta_M)}}_{\text{Change in marginal cost of teaching effort}}.$$

As seen in Equation (2.16), the gain for classroom c would ultimately depend on the impact of changing the teachers' general teaching skill, $\Delta \ln \theta_{G(2,1)}$, and the teachers' marginal cost of

[12]The only exception is "Time in a week spent doing homework for this class", a measure of student effort, for which I assume $N(0, \sigma^2_{\epsilon,dm})$ and take the value of the thresholds, κ_{2dm} to κ_{Jdm}, as known, which allows me to estimate $\sigma^2_{\epsilon,dm}$.

effort, $\Delta \ln \theta_{E(2,1)}$. Thus, Equation (2.16) can be simplified to the reduced-form equation given by

$$\Delta \ln \bar{Y}_{c(2,1)} = \frac{\alpha_{Gcj} \Delta \ln \theta_{G(2,1)} - \delta_E (\alpha_{M_{cj}} - \alpha_L \beta_M) \Delta \ln \theta_{E(2,1)}}{1 - \alpha_L \gamma_L - \delta_E \gamma_{Ecj} (\alpha_{Mcj} - \alpha_L \beta_M)}. \tag{2.17}$$

However, some of the classrooms originally assigned to teacher 1 may benefit more from being reassigned to teacher 2. In particular, the gain may differ across classrooms due to changes in the parameters governing the marginal product of the teaching inputs, α_{Mcj} and α_{Gcj}, as well as the teachers' preferences for promoting students' knowledge, which is captured by γ_{Ecj}. Similar to Ahn et al. (2021), I consider that teacher 2 has a comparative advantage in teaching low-performing classrooms if the gain of reassigning teacher 2 to a classroom with lower initial knowledge (i.e., lower $\ln \bar{K}_{cj}$) is higher than the gain of reassigning the teacher to a classroom with higher initial knowledge. In contrast, a teacher has a comparative advantage in teaching high-performing classrooms if the opposite happens.

Two special cases of Equation (2.15) are worth mentioning. First, if the teachers' preference parameters defined by γ_{Ecj} were equal to zero for all classrooms, the change in teachers' effort due to the reassignment would reduce to $\Delta \ln E_{Mc(2,1)} = -\Delta \ln \theta_{E(2,1)}$, which only depends on the change in teachers' marginal cost of effort, and thus, $\Delta \ln E_{Mc(2,1)}$ would be fixed across classrooms. Accordingly, all the gains differences across classrooms would be due to changes in the marginal product of the teaching inputs. Under this scenario, the achievement gain reduces to a model with linear interactions between the changes in teachers' endowments and the students' initial knowledge. As such, it would resemble specifications in the literature interacting teachers' latent effects with students' fixed inputs (Lockwood and McCaffrey, 2009; Ahn et al., 2021).

Second, if the parameters governing the marginal product of the classroom management practice, α_{Mcj}, the general teaching skill, α_{Gcj}, and teachers' preference for students' knowledge, γ_{Ecj} were fixed across classrooms, the gains would be homogeneous across classrooms, and as such, teachers would not have comparative advantages in teaching different types of students. In Section 2.8, I analyze the possible biases that would be produced by assuming homogeneous changes in the classrooms.

2.3.1 Standard Value-Added Model

A standard value-added specification of the achievement production function takes the form:

$$Z_{icj} = W_{icj} \beta + \mu_{Tj} + \epsilon_{icj}, \tag{2.18}$$

where Z_{icj} the state test score for student i (i.e., a measure of the latent log knowledge), W_{icj} is a vector of students and classroom variables, μ_{Tj} is a classroom-invariant teacher effect, and ϵ_{icj} is an idiosyncratic error term.

Equation (2.18) may potentially capture the reduced-form production of knowledge if the vector W_{icj} included all the students' and classrooms' endowments: $\ln K_{icj}$, $\ln H_{icj}$, $\ln \bar{K}_{cj}$,

$\ln \bar{H}_{cj}$, and $\ln \mu_{Mcj}$, and the closed-form knowledge function were log-linear.[13] Under these assumptions, the reduced form would perfectly capture the expected changes in efforts, and the teacher effect, μ_{Tj}, would be a weighted sum of the teachers' log general teaching skill, $\ln \theta_{Gj}$, and log per-unit cost of effort, $\ln \theta_{Ej}$. Within this frame, the potential gain of the reassignment for student i would be given just by the difference between the teachers' effects:

$$\Delta \ln Y_{ic(2,1)} = \mu_{T2} - \mu_{T1}, \tag{2.19}$$

which, as opposed to Equation (2.16), does not depend on the students' initial knowledge. Equation (2.19) does not allow for comparative advantages in teaching and may not be ideal for extrapolations for several reasons. First, the marginal product of the teaching inputs may change across classrooms. Accordingly, even if the teaching inputs are fixed across classrooms, their impact on students' achievement may change. Second, teacher 2 may adjust her effort to the new classroom; so, her effectiveness may be different from what was originally predicted by the value-added specification. Third, there could be omitted or mismeasured variables in the vector W_{icj}, and to the extent that the teachers' effects are correlated with the unobserved endowments, the omissions and/or mismeasures may generate biased estimates of the teachers' effects.

A more flexible value-added model could be specified, for example, by allowing for interactions between the teacher effects and students' observed variables. However, specifying such models may miss important dimensions of teachers' effectiveness. For example, a flexible value-added specification may be given by

$$Z_{icj} = G(W_{icj}, \mu_{Tj}) + \epsilon_{icj}, \tag{2.20}$$

where the function $G(W_{icj}, \mu_{Tj})$ allows for flexible interactions between the teacher effect, μ_{Tj}, and the observed variables in the vector W_{icj}. Specification (2.20) assumes a one-dimensional teacher effect. Conversely, the production function, Equation (2.3) allows for two types of teacher endowments, a general teaching skill, θ_{Gj}, and a per-unit cost of effort, θ_{Ej}, which are allowed to have different impacts on different classrooms. All else equal, classrooms where teachers with a high general teaching skill are most effective may be different from the classrooms where teachers with a low per-unit cost of effort are most effective. Therefore, there may exist comparative advantages of reassigning teachers to different classrooms that could not be studied with uni-dimensional specifications, such as Equation (2.20). In section 2.8, I discuss the implications of these comparative advantages for students' achievement using the estimates from the structural model.

A growing body of literature estimates the effect of different teaching practices, such as the classroom management practice, on students' achievements. However, it may be important to examine how teachers would adjust their effort to new classrooms to better understand the potential gains of different reassignment policies. To see this, I consider the following specification as an alternative to Equation (2.15):

$$\Delta \ln Y_{ic(2,1)} = \ln G(K_i, \bar{K}_c, \bar{M}_2, \mu_{T2}) - \ln G(K_i, \bar{K}_c, \bar{M}_1, \mu_{T1}), \tag{2.21}$$

[13]The reduced form achievement equation would be log-linear if the parameters governing the marginal product of the classroom management practice, α_{Micj}, the general teaching skill, α_{Gicj}, and teachers' preference for students' knowledge, γ_{Ecj} were constants across students and classrooms.

where the classroom management practices, $\bar{M}_1$, and $\bar{M}_2$, are assumed to be fixed endowments of teacher 1, and teacher 2 respectively. A specification such as Equation (2.21) would implicitly assume that the teachers' efforts are fixed across classrooms and that the classroom environments (i.e, μ_{Mcj}) are attributes of the teachers rather than the classrooms. Therefore, estimating a specification such as Equation (2.21) may generate biased predictions of the potential gains of reassignment.

2.4 Identification

My identification strategy combines multiple sources of variation. First, I use within-teacher variation coming from differences between the students and classrooms to which teachers are assigned, to identify the production and behavioral parameters, and the teachers' endowments. Second, I use within-classroom variation coming from differences between students assigned to a classroom to identify the students' initial endowments. Third, I use multiple measures of the latent inputs to identify the distribution of latent endowments and the measurement system parameters.

2.4.1 Production and Behavioral Parameters

Consider two classrooms with multiple students assigned to teacher j. For simplicity, assume that all inputs are observed without measurement error. The input parameters on the production function, Equation (2.3), are identified due to the input-specific exclusion restrictions. The home environment, H_{icj}, shifts students' learning effort, L_{icj}. Meanwhile, the classroom environment, μ_{cj}, and the averaged classroom home environment, $\bar{H}_{cj}$ are shifters of the classroom management practice, M_{cj}. A similar argument follows for the determinants of the classroom management practice, Equation (2.2), where the classroom averaged initial knowledge, $\bar{K}_{cj}$, and home environment, $\bar{H}_{cj}$ work as shifters for teachers' effort.

Given the production and classroom management parameters, the parameter for students' preference for achievement, γ_L, is identified from variation in students' initial stock of knowledge in the students' best responses, Equation (2.9). The remaining students' behavioral parameters, β_M, β_H, and c_L, are identified from variation in the classroom management practice and students' home environment. The teachers' behavioral parameter, $\gamma_{E,cj}$, is identified from variation in the classroom-level endowments, $\bar{K}_{cj}$, $\bar{H}_{cj}$, and μ_{Mcj}, using the reduced-form equation of teachers' effort, Equation (2.10).

2.4.2 Teachers' Endowments

Given the results from Section 2.4.1, teacher $j's$ general teaching skill, θ_{Gj}, which is a teacher-specific parameter, and the parameter governing the marginal product of it, α_{GK}, are identified by rearranging Equation (2.3) to get:

$$\ln Y_{icj} - \alpha_K \ln K_{icj} - \alpha_{\bar{K}} \ln \bar{K} - \alpha_L \ln L_{icj} - \alpha_{Micj} \ln M_{cj} = \ln \theta_{Gj} + \alpha_{GK} \ln K_{icj} \ln \theta_{Gj}, \quad (2.22)$$

where the left-hand side is the difference between students' achievement, $\ln Y_{icj}$, and the student- and classroom- level inputs in the production function, and the right-hand side is the general teaching skill and its corresponding interaction with the initial stock of knowledge (i.e., $\alpha_{GK} \ln K_{icj} \ln \theta_{Gj}$). A similar argument follows for the identification of the per-unit cost of effort, θ_{Ej}, using the reduced form for teachers' effort, Equation (2.10).

2.4.3 Students' Endowments and Measurement System

In reality, endowments, effort choices, and outcomes are measured with error. By having multiple measures of each educational input, I can separate the joint distribution of the latent unobserved shocks, $\ln \xi_{K_{icj}}$ and $\ln \xi_{H_{icj}}$, from the measurement errors, ϵ_{dm}. In particular, the multiple measures of students' endowments, the observed students' characteristic, X_{icj}, and the orthogonality between unobservables and observables allow identifying the joint distribution of latent endowments given by Equation (2.1). Consider two measures of K_{icj} in the measurement system defined by substituting Equation (2.1) into Equation (2.13):

$$\begin{aligned} Z_{K1icj} &= \kappa_{K1} + \lambda_{K1}(\mu_{Kcj} + X_{icj}\beta_K + \ln \xi_{Kicj}) + \epsilon_{K1icj} \\ Z_{K2icj} &= \kappa_{K2} + \lambda_{K2}(\mu_{Kcj} + X_{icj}\beta_K + \ln \xi_{Kicj}) + \epsilon_{K2icj} \end{aligned} \tag{2.23}$$

First, identification requires normalizing one location and scale parameter. Following the convention in the factor model literature, I normalize $\kappa_{K1} = 0$ and $\lambda_{K1} = 1$ (Agostinelli and Wiswall, 2016b). Second, given the orthogonality between observables and the latent shock and measurement errors, the vector β_K is identified by the regression parameters in the first measurement equation.[14] Third, the classroom-specific intercept, μ_{Kcj}, is identified as a fixed effect in the first measurement equation. Fourth, given β_K and μ_{Kj}, the measurement parameters, κ_{K2} and λ_{K2}, are identified by the regression parameters in the second measurement equation.

The identification of the variance-covariance matrix of latent shocks, Λ, works as follows. First, the variance of the unobserved shock, $\ln \xi_{Kicj}$, is identified by the covariance between the unobserved components of Z_{K1icj} and Z_{K2icj} since ϵ_{K1icj}, ϵ_{K2icj}, and $\ln \xi_{Kicj}$, are all orthogonal to each other. Second, given the variance of the latent shock, the variances of the measurement errors, $\sigma^2_{\epsilon,K1}$ and $\sigma^2_{\epsilon,K2}$, are identified using the variance of each measurement. The parameters for the home environment, $\ln H_{icj}$, are identified following the same approach. Finally, the covariance between the unobserved shocks, $\ln \xi_{Kicj}$ and $\ln \xi_{Hicj}$, is identified using the covariance between the first measure of $\ln K_{icj}$ and $\ln H_{icj}$.

Identification for students' achievement, students' effort, the classroom management practice, and teachers' effort follow from replacing the reduced form equations (2.10), (2.11), and (2.12) in the measurement system. For each latent variable, identification requires normalizing the location, κ_{dm}, and scale, λ_{dm} of one measure.[15] Once the variance-covariance matrix of

[14] For example, if the vector X_{icj} contained only one variable, then β_K would be given by $\frac{cov(\hat{Z}_{K1icj}, \hat{X}_{icj})}{var(\hat{X}_{icj})}$, where $\hat{X}_{icj}$ and $\hat{Z}_{icj}$ represents the variables demeaned at the classroom level.

[15] Ordinal variables require the normalization of one cutoff. For example, by setting $\kappa_{3dm} = 0$.

latent shocks, Λ, is identified, identification of the remaining parameters follows from the system of reduced-form equations, which includes the latent random shocks, with a known joint distribution, and independent measurement error terms that are all orthogonal to the students' characteristics and the fixed effects.

Identification of the distribution of latent shocks only requires two measures of each students' endowment, and all the noisy measures are dependent variables in the measurement system. Thus, I only require a single dedicated measure of each one of the other latent variables in the model: academic achievement, student effort, classroom management practice, teacher effort, and classroom environment.[16] However, the METD contains multiple measurements, which considerably increase the number of observations available. For example, a classroom with 15 students may have 45 measurements of teachers' effort, coming from three different students' survey items, which is beneficial for obtaining more precise point estimates of the parameters.

2.5 Estimation

Similar to Todd and Wolpin (2018), I estimate the model using simulated maximum likelihood.[17] The estimation procedure is as follows:

1. Choose values for the parameters $\{\beta_H, \beta_K, \Lambda, \vec{\mu}_K, \vec{\mu}_H, \Psi, \vec{\mu}_M, \vec{\theta}_E, \vec{\theta}_G, \Gamma\}$, where β_H, β_K, Λ and the vectors $\{\vec{\mu}_K, \vec{\mu}_H\}$ denote the students' initial endowment parameters, Ψ denotes the parameters of the measurement system, $\vec{\mu}_M$ denotes a vector of classroom environment effects, $\{\vec{\theta}_E, \vec{\theta}_G\}$ are vectors of the teacher-specific endowments, and Γ denotes the production and behavioral parameters.
2. For each classroom, calculate the average of the students' endowments, $\overline{\ln K_{cj}}$ and $\overline{\ln H_{cj}}$ using arithmetic means of the observable students' characteristics effects, $X_{icj}\beta_K$ and $X_{icj}\beta_H$, and fixed effects, μ_{Kcj} and μ_{Hcj}, in Equation (2.1).
3. Draw D student-level shocks for the unobserved latent variables, ξ_{Kicj} and ξ_{Hicj}.
4. For each draw d, calculate each student initial endowment, $\ln K^d_{icj}$ and $\ln H^d_{icj}$, and the reduced form equations given by (2.10), (2.11), and (2.12).
5. For each draw and given the measurement system parameters, calculate the likelihood of observing student i's measures, f^d_{icj}, which is a product of likelihood statements for continuous and discrete measures.
6. For each student, average all the draw-specific likelihoods, $f_{icj} = \frac{1}{D}\sum_1^D f^d_{icj}$.

[16] I require at least one dedicated measure of the classroom environment to separately identify μ_{Mcj} from δ_E in Equation (2.2) because the classroom environment effects and teachers' managerial efforts are defined at the classroom level. The identification of the classroom environment effect comes from variation in the classroom management practice that is not due to changes in teachers' effort.

[17] While Todd and Wolpin (2018) estimates distributions of classrooms' and teachers' endowments using random shocks, I estimate them using latent fixed effects (see, for example, Agostinelli et al. (2019)).

7. The total log-likelihood is given by

$$\text{Log Likelihood} = \sum_j \sum_c \sum_i \ln f_{icj} \tag{2.24}$$

To deal with the computational burden of estimating a large set of parameters, I exploit the closed-form solution of the model and maximize the log-likelihood using analytical gradients, which provides a fast convergence.[18] Additionally, I can estimate the teachers' and classrooms parameters, $\{\vec{\mu}_K, \vec{\mu}_H, \vec{\mu}_M, \vec{\theta}_E, \vec{\theta}_G\}$, as fixed effects in one step, as opposed to iterative procedures that may not converge to a global maximum (Greene, 2004).

The simulated maximum likelihood procedure allows me to estimate a non-linear model with multiple latent factors combining continuous and ordinal measures. Students contribute to the likelihood with their available measures. All the noisy measures are dependent variables in the likelihood function.

2.6 Data

The Measures of Effective Teaching Longitudinal Database (METLD) was compiled over two academic years (2009–2010 and 2010–2011) across six districts in the United States: Charlotte-Mecklenburg, Dallas, Denver, Hillsborough County, Memphis, and New York City. The METLD collected information about teachers' and students' behaviors, as well as district-wide administrative records on students' current and past achievements, and other students' characteristics. The METLD was initially designed to test the validity of different methods to identify effective teachers (Kane et al., 2013). To do so, teachers were randomly assigned to already established classrooms before the start of the 2010–2011 academic year. This unique feature of the database allows researchers to use the 2009-2010 data to estimate different models and then compare predicted and actual outcomes using the 2010–2011 data.

I estimate the model using data on 6^{th} grade students in the first year of the METD.[19] Sixth-grade schoolteachers are usually subject matter specialists who teach multiple classrooms per year, which allows me to identify the model using only one academic year (2009–2010). I focus on math teachers because they have been the focus of recent educational policy interventions in the US that did not achieve their main goal of increasing students' performance on standardized exams (Glazerman et al. 2013; Stecher et al. 2018).

The original sample of classrooms participating in the MET intervention contains 200 6^{th} grade math teachers, 366 6^{th} grade classrooms with at least one student answering the survey, and 8,078 6^{th} grade students. I restrict the original sample to include only students with full information on age, race, gender, special education status, gifted status, and English learner

[18] Greene (2009, 2018) provides a detailed analysis of how to perform simulated maximum likelihood with analytical gradients.

[19] In Denver, no middle schools signed up for the program, which reduces the number of districts in the sample to five (White et al., 2014).

status.[20] Furthermore, I only keep classrooms that have at least seven students who have at least one measure of initial knowledge, home environment, academic achievement, and teacher effort.[21] I restrict the sample to teachers assigned to more than one classroom. The final sample consists of 6,734 students, 298 classrooms, and 149 teachers.

Table 2.1 shows descriptive statistics of the final sample and the original sample of 6^{th} grade classrooms. The table also shows that no variable in Table 2.1 is significantly different between the original and estimation samples at any conventional significance level. As seen, students in the estimation sample are, on average, 11.5 years old, and approximately 62 percent of them are either African American or Hispanic students.

Table 2.1: Descriptive Statistics

	Estimation Sample		Original Sample	
	Mean	Std. Dev.	Mean	Std. Dev.
Student Variables				
Age	11.605	.556	11.577	.576
Gender:Male	.514	.500	.513	.500
Race: African American	.291	.454	.296	.457
Race: Hispanic	.328	.470	.331	.471
Race: Asian	.063	.243	.068	.251
Gifted Status	.101	.302	.102	.303
Special Education	.071	.257	.074	.262
English Language Learner	.156	.363	.151	.358
Twice Lagged Math State Test Score	.046	.935	.051	.950
Lagged Math State Test Score	.054	.928	.058	.945
Math State Test Score	.075	.955	.076	.962
Classroom Variables				
Classroom Size	25.217	6.811	25.156	7.066
Number of Teachers	149		200	
Number of Classrooms	298		366	
Number of Students	6,734		8,078	

2.6.1 Measures of Classroom Management and Teacher Effort

The measures of the classroom management practice and teachers' level of effort come from Ferguson's Tripod survey (Ferguson and Danielson, 2015). The survey was designed to capture students' perceptions of what teachers do in the classrooms and how classrooms work by asking students to rate their agreement with 35 statements using a five-point scale, from "to-

[20] I do not include the free lunch indicator as an exogenous student variable in my model because one district did not report it. Missing values for the twice-lagged math test score (13% of the observations) are imputed from a linear regression of the score of students' observable characteristics and a classroom fixed effect.

[21] I exclude classrooms that have fewer than seven observations to avoid estimating fixed effects based on very few observations.

tally untrue" to "totally true" (Raudenbush and Jean, 2015; Kuhfeld, 2017). These statements are grouped into seven dimensions: Care, Control, Clarify, Challenge, Captivate, Confer, and Consolidate.

Many different models have been proposed to test the dimensionality of the 7 Cs. For example, a model proposed by Ferguson and Danielson (2015) groups the Control and Challenge dimensions into a *Press* domain, which is defined as "keeping students busy and on task and pressing them to think rigorously", while the other five Cs are grouped into an *Academic Support* domain, which is defined as "caring teacher-student relationships, captivating lessons, and other practices that students experience as supportive".[22] The Control items, as opposed to the majority of the other items, references the effective control of the classroom rather than teachers' actions to keep the class well-behaved. In this regard, previous work using exploratory and theory-based factor analysis suggests two latent factors: one consisting of Control statements, and a separate factor containing statements from the other six Cs (Wallace et al., 2016; Kuhfeld, 2017).

My proposed model comprises a classroom management practice that may depend on teachers' effort and a classroom environment effect. Since Control items reference the effective implementation of teachers' actions to keep the class on task and well-behaved, I treat them as measures of the classroom management practice (e.g., "student behavior in this class is under control"). To measure teachers' effort, I focus on questions that directly reference teachers' actions in the classroom. In particular, I focus on items from the Challenge domain, such as "My teacher doesn't let people give up when the work gets hard," because they relate to actions taken in the classroom to keep students on task and well-behaved. Lastly, I treat the number of students in the classroom as a dedicated measure of the classroom environment.[23]

2.6.2 Selected Measurements

Table 2.2 shows all the measures listed for each variable of the structural model. The students' achievement measures include state test scores and scores in the Balanced Assessment in Mathematics (BAM) test, which is a supplemental test comprised of open-ended math questions. I treat these measures as continuous and standardize them to have a mean of zero and a variance equal to 1. The measures of students' initial stock of knowledge are the student's standardized scores on the math and science state tests from the prior year.[24]

For students' home environment, I select measures that relate to the cost of students exerting effort. In particular, the number of books and computers in the house are treated as continuous measures and standardized to have a mean of 0 and a variance equal to 1. Parents' level of education is a categorical variable with five categories, ranging from "did not finish high school" to "professional or graduate degree. " For studying effort, the logarithm of the time in

[22] I developed and estimated a version of the model with Academic Support as an additional teaching input. However, this input does not show significant variation conditional on the other inputs and the teaching skills.

[23] The classroom environment effect is also measured by the measures of the classroom management practice since the latent practice depend on the classroom environment and the teachers' effort.

[24] Twenty-seven percent of the sample has information on the prior year's science test score.

Table 2.2: Measures of Model Variables

Variable: Students' Initial Knowledge, $\ln K_{icj}$
Measures: 5-th grade math score in state test (administrative record, standardized). 5-th grade science score in state test(administrative record, standardized).

Variable: Students' Home Environment, $\ln H_{icj}$
Measures: number of books at home (student survey, in logs, standardized). Number of computers at home (student survey, in logs, standardized). Parent's level of education (student survey).

Variable: Classroom Environment, $\ln \mu_{Mcj}$
Measures: Classroom Size (in logs)

Variable: Students' Effort, $\ln L_{icj}$
Measures: Time in a week spent doing homework for this class (student survey, in logs). Percent of the homework completed (student survey). Stop trying when the work gets hard (student survey, reverse coded).

Variable: Classroom Management Practice, $\ln M_{cj}$
Measures: students in this class treat the teacher with respect (student survey). Student behavior in this class is under control (student survey). My classmates behave the way my teacher wants them to (student survey).

Variable: Teacher Managerial Effort, $\ln E_{Mcj}$
Measures: my teacher doesn't let people give up when the work gets hard (student survey). In this class, my teacher accepts nothing less than our full effort (student survey). My teacher pushes me to become a better thinker (student survey).

Variable: Student Knowledge, $\ln Y_{icj}$
Measures: 6-th grade math score in state test (administrative record, standardized). BAM score (supplemental test, standardized).

a week spent doing homework is an ordinal variable with seven categories.[25]

The measurements for teacher effort and classroom management practice come from the Tripod 7Cs and are treated as ordered categorical measures with a 5-point scale, ranging from "totally untrue" to "totally true." Finally, I use the logarithm of classroom size as a measure of the latent classroom environment effect.[26] The Appendix Table A.A.2 reports the mean and standard deviation of continuous measures and the proportion of students who answered each category of the ordinal measures, as well as the number of observations for each measure.

[25]The cutoffs of the logarithm of the time in a week spent doing homework are known, which allows me to recover the variance of the error term for this measure.

[26]The classroom size is the only measure that is not defined at the student level.

2.7 Estimation Results

The model contains 1,287 parameters, including 13 parameters of the utility and production functions, 57 parameters of the measurement system, and 298 parameters for the teachers' endowments; the rest of the parameters relate to the students' endowments. The sufficient conditions for an interior solution for teachers and students are not binding at the estimated parameters.

Table 2.3: Production and Utility Functions Parameter Estimates

	Structural Parameters	
	Value	SE
(1) Knowledge, Y_{icj}:		
Coefficient on students' initial knowledge: α_K	.949	.017
Coefficient on classroom initial knowledge: $\alpha_{\bar{K}}$	-.134	.017
Coefficient on students' effort: α_L	.318	.016
Coefficient on classroom management: α_M	.083	.007
Interaction on classroom management: α_{MK}	-.074	.002
Interaction on general teaching skill: α_{GK}	.111	.022
(2) Students' Utility:		
Coefficient on students' home environment: β_H	-.508	.024
Coefficient on classroom management: β_M	-.192	.014
Students' preference for achievement: γ_L	.059	.013
Intercept: $log(c_L)$	-1.675	.057
(3) Classroom Management Practice, M_{cj}:		
Coefficient on teachers' effort: δ_E	1.429	.018
(4) Teachers' Utility:		
Preference for students' achievement: γ_E	.799	.011
Interaction parameter: γ_{EK}	.13	.008

Table 2.4: Summary Statistics of the Educational Inputs and Achievement

	Mean	Standard Deviation
Student Knowledge, $\ln Y_{icj}$	-.009	.883
Student Initial Stock of Knowledge, $\ln K_{icj}$	-.023	.906
Student Home Environment, $\ln H_{icj}$	-.039	.565
Student Effort, $\ln L_{icj}$	.542	.362
Classroom Stock of Initial Knowledge, $\ln \bar{K}_{cj}$	-.061	.714
Classroom Home Environment, $\ln \bar{H}_{cj}$	-.056	.424
Classroom Environment, $\ln \mu_{Mcj}$	-.818	.531
Classroom Management Practice, $\ln M_{cj}$	.209	.675
Teacher Effort, $\ln E_{Mcj}$	.719	.393
Teacher General Teaching Skill, $\ln \theta_{Gj}$	-.175	.226
Teacher per-unit Effort Cost, $\ln \theta_{Ej}$	-2.311	.371

Table 3.8 presents estimates of the structural parameters of the production and utility functions. Estimates for the production function appear in the first panel of table 3.8. The point estimates indicate that the students' initial knowledge is a positive and significant determinant

of academic achievement ($\alpha_K > 0$). As seen in Table 2.4, the student's initial stock of knowledge is the educational input that varies the most (i.e., the standard deviation is .906). All else equal, increasing the (log) initial stock of knowledge by one standard deviation above the mean increases the mean log achievement by .859 points (.972 of a standard deviation of the latent achievement). However, this effect is ameliorated by the averaged classroom initial knowledge ($\alpha_{\bar{K}} < 0$): on average, increasing the classroom average initial knowledge, $\ln \bar{K}_{cj}$, by one standard deviation above the mean decreases the log achievement by .096 points (.109 of a standard deviation).

Students' achievement increases with their studying effort ($\alpha_L > 0$). As seen in Table 2.4, the mean student log effort is .542, which corresponds to 1.72 hours a week doing homework for the class. On average, increasing student log effort by one standard deviation above the mean increases the log knowledge by .115 points (.13 of a standard deviation).

The parameter governing the marginal product of the classroom management practice, α_{Micj}, is, on average, equal to .083, and decreases with the student stock of initial knowledge (i.e., $\alpha_{MK} = -.074$). To give a sense of the magnitude of this parameter, first I calculate the impact of a one standard deviation increase in the log classroom management practice if the log initial stocks of knowledge were at the mean. This change results in a .056 points (.063 of a standard deviation) increase in the average log knowledge. If the log initial stock of knowledge were one standard deviation below the mean, the average increase would be .150 points (.17 of a standard deviation) instead. In contrast, if the log initial stock of knowledge were one standard deviation above the mean, the mean log knowledge would increase by only .016 points (.018 of a standard deviation).

The parameter governing the marginal product of general teaching skill increases with the students' stock of initial knowledge (i.e., $\alpha_{GK} = .111$). Increasing the log general teaching skill by 1 standard deviation for a student whose log initial stock of knowledge is at the mean results in a .226 points increase in log knowledge (.256 of a standard deviation), while this increase would be .249 points (.282 of a standard deviation) if the student were one standard deviation above the mean instead. Accordingly, all else equal, a knowledgeable student benefits more from teachers' general teaching skill and less from the classroom management practice.

The estimated parameters for the students' preferences and cost function appear in the second panel of table 3.8. To ease their interpretation, I replace the point estimates in the students' best response function in Equation (2.9):

$$\begin{aligned}\ln L^{br}_{icj} =& .538 + .056 \ln K_{icj} - .01 \ln \bar{K}_{cj} + (.06 + .01 \ln K_{icj}) \ln \theta_{Gj} + \\ & (.20 - .00 \ln K_{icj}) \ln M_{cj} + .52 \ln H_{icj}.\end{aligned} \tag{2.25}$$

The students' preference for achievement parameter γ_L is positive but close to zero, which implies that students do not respond much to changes in their initial knowledge, the averaged classroom initial knowledge, and general teaching skill. However, the estimates indicate that students present strategic complementarity with teachers' effort. A better classroom management practice considerably increases learning effort through a reduction in study cost, as shown by the negative and significant coefficient β_M. All else equal, increasing the classroom management practice by one standard deviation above the mean results in .248 hours increase in the mean study hours. The parameter governing the marginal product of the home environment,

β_H, is negative, which indicates that students' effort increases with a more supportive home environment due to a reduction of the utility cost of studying. All else equal, increasing home environment by one standard deviation above the mean increases mean students' effort by .572 hours.

The third panel of table 3.8 presents the teacher effort parameter, δ_E, in the classroom management practice Equation (2.2). The positive δ_E implies that the classroom management practice increases with teacher effort. Finally, the fourth panel of table 3.8 shows that teachers' preference for students' knowledge, γ_{Ecj}, is, on average, equal to .79, and increases with the initial classroom knowledge (γ_{EK} = .13), meaning that the teachers' marginal utility from students' achievement is higher in classrooms with knowledgeable students.[27]

Table 2.5 shows the determinants of students' endowments. All else equal, students' initial knowledge is higher for younger students and those who are gifted, as well as those with higher twice-lagged math scores. In contrast, all else equal, African American and Hispanic students, as well as English language learners, and students with special education status have lower initial knowledge. There is no significant difference due to gender in the initial stock of knowledge. The results are similar for the determinants of students' home environments. Additionally, Table 2.5 shows that there exists a positive correlation between the unobserved shocks of the students' endowments, $\ln \xi_{Kicj}$ and $\ln \xi_{Hicj}$.

Table 2.5: Determinants of Student Endowments

	Structural Parameters			
	Initial Knowledge, K_{icj}		Home Environment, H_{icj}	
	Value	SE	Value	SE
Exogenous Student Variables				
Age	-.041	.006	-.064	.01
Gender:Male	-.001	.012	-0.001	.016
Race: African American	-.155	.013	-.198	.023
Race: Hispanic	-.032	.015	-.32	.022
Race: Asian	.132	.026	-.154	.035
Gifted Status	.138	.02	.052	.03
Special Education	-.083	.022	-.014	.035
English Language Learner	-.123	.019	-.042	.029
Twice Lagged Math Score	.529	.007	.058	.011
Unobserved Shocks				
$SD(\ln \xi_{Kicj})$	.342	.0142		
$SD(\ln \xi_{Hicj})$			.379	.007
$Corr(\ln \xi_{Kicj}, \ln \xi_{Hicj})$	.108	.007		

Table 2.6 presents a correlation matrix of teachers' and students' endowments averaged at the classroom level. For the teachers' endowments, the correlation shows that those who are more skilled at increasing students' academic achievement (i.e., a higher $\ln \theta_{Gj}$) tend to have

[27] I performed a likelihood ratio test forcing the interaction parameters α_{GK}, α_{MK}, and γ_{EK} to be equal to zero. I reject the hypothesis that these restrictions are valid at all standard confidence levels.

a higher per-unit cost of effort (i.e., a higher $\ln\theta_{Ej}$). For the classrooms' endowments, the correlations show that those with a higher initial knowledge (i.e., a higher $\ln\bar{K_{cj}}$) tend to have students with better home environments (i.e., a higher $\ln\bar{H_{cj}}$). However, these two classrooms' endowments are not strongly correlated with the classroom environment, $\ln\mu_{Mcj}$.

Table 2.6 also shows that teachers with higher general teaching skill tend to be sorted into classrooms with higher initial knowledge and better home environments. A similar pattern can be seen for teachers with a higher per-unit cost of effort. However, teachers with higher general teaching skill tend to be sorted to classrooms with a slightly worse classroom environment , while teachers with a lower effort cost tend to be assigned to classrooms with a worse classroom environment. Finally, Table A.A.1 in the appendix shows the parameters of the measurement system.

Table 2.6: Correlation Matrix of Teachers' and Averaged Students' Endowments

	$\ln\theta_G$	$\ln\theta_E$	$\overline{\ln K_{cj}}$	$\overline{\ln H_{cj}}$	$\ln\mu_M$
$\ln\theta_G$	(.051)				
$\ln\theta_E$	.342	(.138)			
$\overline{\ln K_{cj}}$	.077	.050	(.51)		
$\overline{\ln H_{cj}}$	.145	.132	.667	(.18)	
$\ln\mu_M$	-.052	.348	.043	.001	(.282)

Note: The elements in parenthesis on the main diagonal are the variance of the endowments

2.7.1 Model Fit

Table 2.7 presents statistics for each measurement and compares them to the predicted statistics from simulations of the model. I report the mean and standard deviation of continuous measures and the proportion of students who answered each category of the ordinal measures. Moreover, the last column of table 2.7 reports the fraction of variation of each measure that is due to the variation in its corresponding latent factor (i.e., the signal ratio).[28] Overall, Table 2.7 shows that the model closely fits the statistics.

The first panel shows the statistics for the measures of academic achievement. As seen, there is no meaningful difference between the actual and simulated statistics for both measures. However, the measures differ in the fraction of the variance that comes from variation in the latent factor. In particular, latent achievement accounts for 79 percent of the state test score variation, while it accounts for 62 percent of the variation in the BAM test.

The second and third panels of table 2.7 show statistics for the measures of students' endowments. Similar to the measures of academic achievement, the model fits the mean and the standard deviation of both measures of students' initial knowledge very well. The latent

[28]The fraction of variation explained by the model is defined as $\frac{\lambda^2_{dm}\sigma^2_{True}}{\lambda^2_{dm}\sigma^2_{True}+\sigma^2_\epsilon}$. Ordered categorical measures are treated as discrete measures of a continuous latent variable, with total variation given by the variation explained by the model (i.e., $\lambda^2_{dm}\sigma^2_{True}$) plus the variation in the measurement error (i.e., σ^2_ϵ).

factor for students' initial knowledge explains 82 percent of the lagged math test score variation. In comparison, it accounts for 60 percent of the variation of the lagged science test score. These results highlight the importance of treating educational inputs as latent factors; not accounting for measurement error can generate biased estimates of the effect of students' initial knowledge on other educational inputs and, ultimately, students' achievement. As seen in the third and fourth panels, the model also fits very well with the statistics for students' home environment and effort.

The last two panels of table 2.7 compare the statistics for classroom management practice and teachers' effort. The first measure of classroom management practice, "students treat the teacher with respect," has five categories ranging from "totally untrue" to "totally true," and the model fits the proportions very well. The fraction of variation explained by the model is 31 percent, which can partially be due to the measurement error term being defined at the student level while the latent factor is defined at the classroom level. The fits for the other measures of classroom management practice are similar, but they are noisier than the first measure. The fit for the measures of teacher effort is also good. However, the fraction of variation explained by the model is smaller. The least noisy measure is "my teacher pushes me to become a better thinker," for which the latent effort explains 15 percent of the total variation.

Table 2.7: Model Fit: Mean and Standard Deviation

	Actual		Model		
	Mean	SD	Mean	SD	$\frac{\lambda^2_{dm}\sigma^2_{True}}{\sigma^2_{Total}}$
(1) Measures of ln Y_{icj}					
State Test	0	1	-.001	.997	.786
BAM	0	1	-.002	1.001	.621
(2) Measures of ln K_{icj}					
Previous Year Math	0	1	-.009	1.001	.821
Previous Year Science	0	1	-.008	.998	.602
(3) Measures of ln H_{icj}					
Number of Books	0	1	-.012	.993	.320
Number of Computers	0	1	-.006	.991	.244
Parents' Education					.319
Did not finish HS	19.8		19.97		
High School	20.86		21.21		
Some College	13.66		13.89		
4-year College	18.41		18.06		
Professional or Graduate degree	27.27		26.87		
(4) Measures of ln L_{icj}					
log Hours/week of study	.845	.84	.855	.82	.15
Homework Completed:					.137
None of it/Never Assigned	3.6		3.48		
Some of it	11.38		11.38		
Most of it	23.23		23.70		
All	53.74		53.22		
All plus some extra	8.05		8.22		
Stop Trying:					.116
Totally Untrue	4.54		4.51		
Mostly Untrue	5.72		5.69		
Somewhat Untrue/True	14.77		14.99		

Table 2.7: Model Fit: Mean and Standard Deviation

	Actual		Model		
	Mean	SD	Mean	SD	$\frac{\lambda_{dm}^2 \sigma_{True}^2}{\sigma_{Total}^2}$
Mostly True	22.65		22.76		
Totally True	52.31		52.05		
(5) Measures of ln M_{cj}					
Students Treat with Respect:					.313
Totally Untrue	6.92		6.83		
Mostly Untrue	9.36		9.35		
Somewhat Untrue/True	26.41		26.71		
Mostly True	32.22		32.11		
Totally True	25.10		25.00		
Behavior under control:					.228
Totally Untrue	12.56		12.64		
Mostly Untrue	12.96		12.74		
Somewhat Untrue/True	26.89		27.06		
Mostly True	26.31		26.45		
Totally True	21.28		21.11		
Classmates behave:					.279
Totally Untrue	14.76		14.72		
Mostly Untrue	14.79		14.78		
Somewhat Untrue/True	31.02		31.36		
Mostly True	25.26		24.95		
Totally True	14.17		14.19		
(6) Measures of ln E_{Mcj}					
Doesn't let give up:					.134
Totally Untrue	3.92		3.92		
Mostly Untrue	4.21		4.11		
Somewhat Untrue/True	16.44		16.38		
Mostly True	25.41		25.53		
Totally True	50.02		50.05		
Better Thinker:					.152
Totally Untrue	6.22		6.19		
Mostly Untrue	6.12		6.10		
Somewhat Untrue/True	20.61		20.72		
Mostly True	26.33		26.40		
Totally True	40.72		40.59		
Accepts Nothing Less:					.126
Totally Untrue	3.78		3.81		
Mostly Untrue	4.49		4.44		
Somewhat Untrue/True	19.64		19.50		
Mostly True	26.79		26.98		
Totally True	45.31		45.28		

The model aims to predict the impact of commonly suggested reassignment policies, such as transferring teachers with high value-added estimates to low-performing classrooms. Thus, it is important to check whether the model captures moments that relate to these teachers and classrooms. Table 2.8 compares the actual and predicted means of selected measures, conditional on being below or above mean characteristics of the classrooms and teachers. The selected measures are the math score of the state test (a measure of students' achievement), the lagged score in the math state test (a measure of initial knowledge), the number of books in the

Table 2.8: Model Fit: Conditional Means

	Actual		Model	
	Below Mean	Above Mean	Below Mean	Above Mean
(1) Classroom Averaged Lagged Score:				
State Test Score (Y_{icj})	-.451	.648	-.474	.636
Lagged Test Score (K_{icj})	-.495	.709	-.522	.689
Number of Books (H_{icj})	-.247	.325	-.241	.290
log Hours/week of Study (L_{icj})	.690	1.052	.705	1.018
Students Treat with Respect (M_{cj})	3.479	3.742	3.511	3.692
Doesn't let give up ($E_{M,cj}$)	4.078	4.208	4.086	4.202
(2) Classroom Averaged Number of Books:				
State Test Score (Y_{icj})	-.398	.439	-.387	.416
Lagged Test Score (K_{icj})	-.406	.451	-.368	.396
Number of Books (H_{icj})	-.373	.380	-.397	.392
log Hours/week of Study (L_{icj})	.646	1.051	.669	1.023
Students Treat with Respect (M_{cj})	3.469	3.719	3.517	3.669
Doesn't let give up ($E_{M,cj}$)	4.054	4.216	4.089	4.188
(3) Estimated Value-Added :				
State Test Score (Y_{icj})	-.222	.256	-.251	.269
Lagged Test Score (K_{icj})	-.107	.086	-.06	.054
Number of Books (H_{icj})	-.017	.011	-.069	.068
log Hours/week of Study (L_{icj})	.807	.893	.772	.923
Students Treat with Respect (M_{cj})	3.463	3.763	3.401	3.793
Doesn't let give up ($E_{M,cj}$)	4.025	4.259	4.023	4.26
(4) Proportion of Minority :				
State Test Score (Y_{icj})	.408	-.359	.429	-.383
Lagged Test Score (K_{icj})	.383	-.335	.386	-.348
Number of Books (H_{icj})	.311	-.307	.309	-.309
log Hours/week of Study (L_{icj})	1.062	.633	1.014	.680
Students Treat with Respect (M_{cj})	3.700	3.488	3.675	3.512
Doesn't let give up ($E_{M,cj}$)	4.205	4.065	4.149	4.127

household (a measure of the home environment), the time in a week spent doing homework (a measure of students' effort), the categorical variable "students treat the teacher with respect" (a measure of the classroom management practice), and "the teacher doesn't let students give up when the work gets hard" (a measure of teacher effort).[29] Given the model's parametric restrictions, this is a challenging test.

Classroom-Averaged Lagged Score: The first panel of Table 2.8 shows the means of the selected measures, conditional on being assigned to classrooms that are below or above the mean classroom-averaged lagged math state test score. Overall, Panel (1) shows that the model fits well with the relationship between the selected measures and the averaged classrooms' lagged score. For example, the first row of the first column shows that the actual mean state test score for classrooms below the mean is $-.451$, while the model predicts $-.474$. Meanwhile, the first row of the second column shows that the actual mean state test score for classrooms above the average is .648, while the model predicts .636.

[29] In Table 2.8, the measures of classroom management and teacher effort are treated as continuous with values ranging from 1 (Totally Untrue) to 5 (Totally True).

Classroom-Averaged Number of Books: The second panel of Table 2.8 shows the means of the selected measures, conditional on being below or above the mean of the classroom-averaged number of books at home, which is a measure of the home environment. This is a demanding test because the home environment enters the model through the students' cost of effort only. Similar to Panel (1), the model fits very well with the conditional moments.

Estimated Value-Added: In the third panel of Table 2.8, I estimate a standard value-added specification of teachers' effectiveness using teacher fixed effects.[30] I then estimate the mean of the selected measures, conditional on being assigned to teachers who are below/above the mean value-added score. The panel tests the ability of the model to replicate correlation patterns of the commonly used value-added estimator. As in the previous panels, the model predicts the patterns in the data. In particular, the actual mean state test score for students assigned to teachers below the value-added mean is −.222, while the model predicts −.251. Similarly, the actual mean for teachers above the value-added mean is .256, while the model predicts .269. Table A.A.3 in the Appendix shows the fit of the value-added regression. As shown, the model fits the coefficients of the regression, the proportion of variation being explained, and the coefficient of variation of the value-added estimates.

Proportion of Minority Students: Lastly, the fourth panel of Table 2.8 shows the means of the selected measures, conditional on being below or above the mean classroom proportion of minority students, or students who are either African American, Hispanic, or Native American. Since students' ethnicity only appears in their initial endowments and not directly in the achievement function, this is also a demanding test. As in the previous panels, Panel (4) shows that the model fits well with the patterns observed in the data.

2.7.2 Out-of-Sample Fit

The joint estimation of the production and behavioral parameters with the flexibly-correlated latent endowments ameliorates the sources of bias due to the non-random sorting of teachers to classrooms. However, a prevalent concern in the teachers' effectiveness literature is that teachers may be consistently assigned to classrooms based on unobserved traits of their students, which in turn may generate biased estimates of teachers' effectiveness (see, for example, Rothstein (2017)). The main goal of the METD study was to test the validity of commonly used measures of teachers' effectiveness by eliminating the systematic sorting of teachers to classrooms (Kane et al., 2013). This was operationalized by randomly assigning teachers to classrooms in the second year of the program (2010-2011) and testing the validity of different measures that were estimated using the data from the first year of the program (2009-2010). This feature of the METD allows me to explore whether my model can predict different statistics using from the randomized data.

[30] The value-added specification controls for the students' observable characteristics (excluding the twice-lagged math score), the lagged math state test score, the previous variables averaged at the classroom level, and the classroom size, as well as teacher fixed effects to measure teachers' effectiveness.

Table 2.9: Out-of-Sample Fit: Mean and Standard Deviation

	Actual		Model	
	Mean	SD	Mean	SD
(1) Students' achievement:				
State Test Score	.044	.965	.073	.975
	[.004,.084]			
BAM socre	.044	1.007	.015	.982
	[-.002,.089]			
(2) Students' Effort				
log Hours/week of study	.859	.771	.865	.825
	[.824,.893]			
Homework Completed	3.599	.917	3.528	.919
	[3.557,3.639]			
Stop Trying	4.215	1.089	4.145	1.123
	[4.166,4.263]			
(3) Classroom Management Practice:				
Behavior under control	3.447	1.272	3.402	1.284
	[3.390,3.504]			
Classmate behave	3.172	1.204	3.194	1.249
	[3.118,3.225]			
(4) Teachers' Effort :				
Doesn't let give up	4.229	1.070	4.178	1.052
	[4.181,4.276]			
Accept Nothing Less	4.247	.960	4.089	1.059
	[4.204,4.290]			

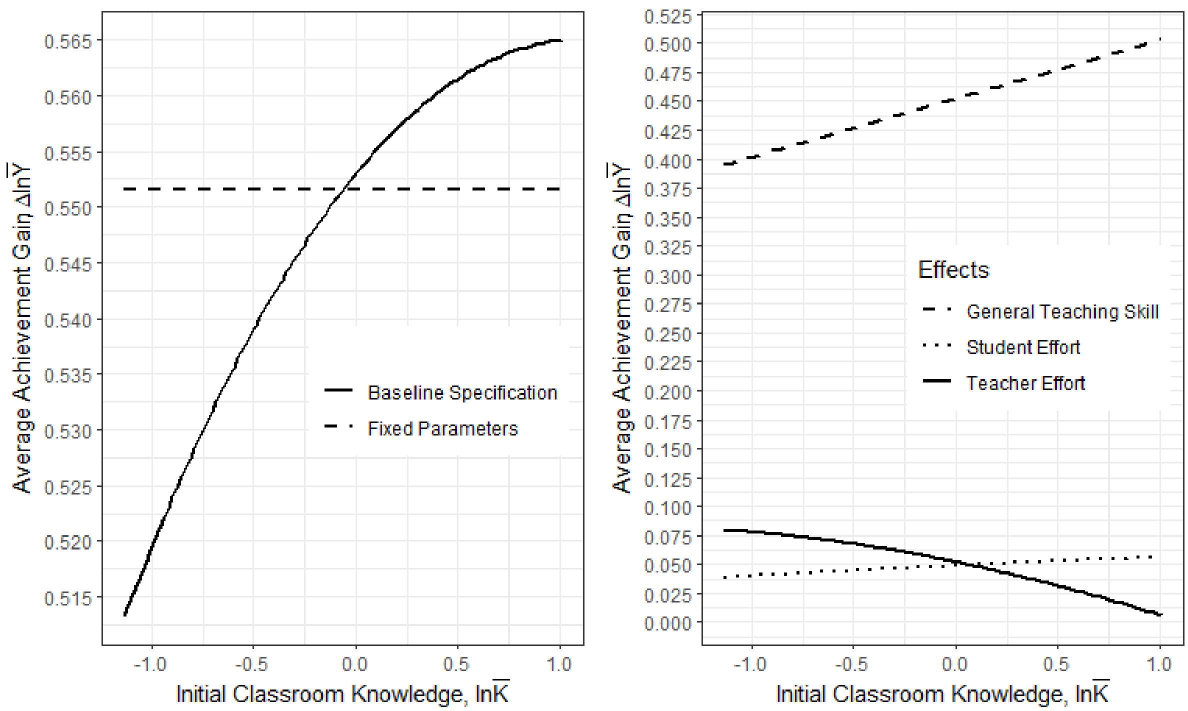
Note: 95% confidence intervals reported under each estimated mean coefficient. Sample consist of 100 teachers assigned to 2,274 students in 100 classrooms.

The new sample consists of 2,274 students, 100 classrooms, and 100 teachers. Thus, this is a challenging test because the sorting mechanisms between teachers and classrooms change, and the randomized data consists of a subset of the teachers included in the estimation sample.

Before conducting the validation, I recover the fixed effects parameters of the second-year classrooms. In particular, I recover μ_{Kcj} and μ_{Hcj}, defined in Equation (2.1), using the lagged math score and the number of books at home. I use the first measure of the classroom management practice to recover the classroom environment effects, μ_{Mcj} in Equation (2.2). To perform the out-of-sample validation, I use the observed characteristics of the new students and the full set of estimated parameters to predict the mean and standard deviation of the unused measures of students' knowledge, students' effort, classroom management practice, and teachers' effort.[31]

As seen in Table 2.9, the model predicts the mean and standard deviation of the measures very well. The model tends to slightly underpredict the mean of the measures of teachers'

[31]The measure of teachers' effort, "My teacher pushes me to become a better thinker", is not available in the second year of the program.



(a) Overall Effect (b) Effect Decomposition

Figure 2.1: Achievement Gains due to Changes in the General Teaching Skill

effort, as well as the second and third measures of students' effort. However, the predicted mean of the measures of academic achievement, the first measure of students' effort, and the measures of managerial practice all lie within the 95% confidence interval of the sample means.

2.8 Comparative Statics

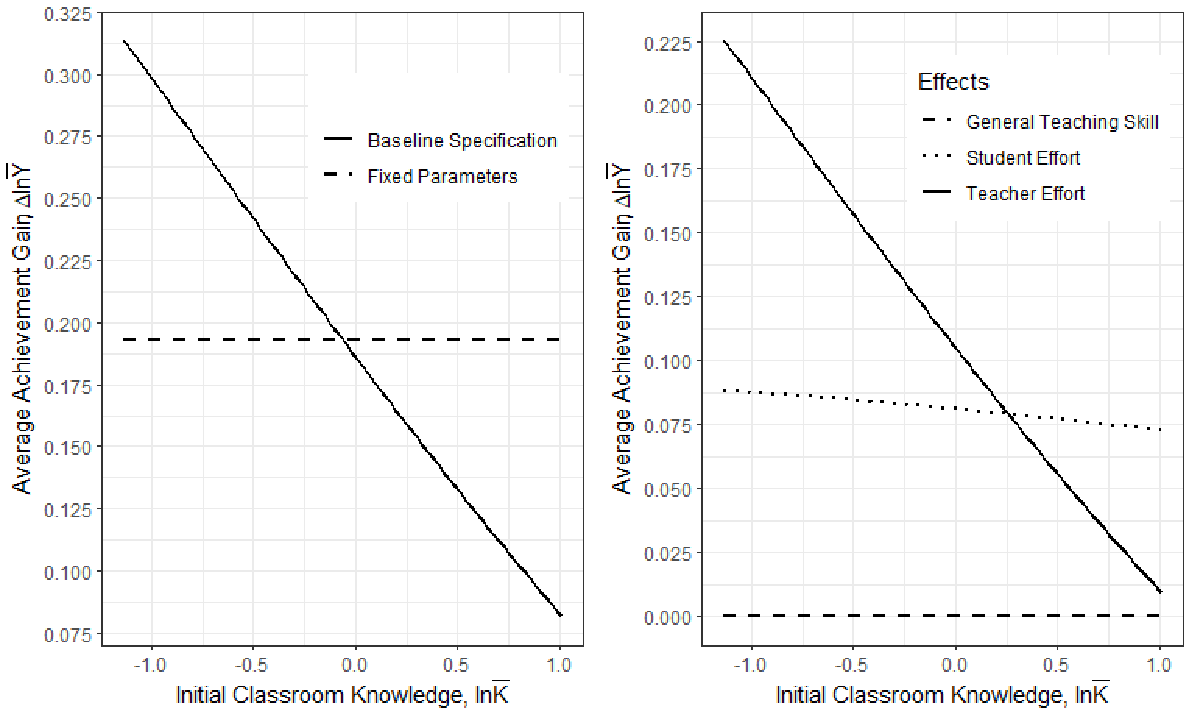
A crucial aspect of my theoretical model is that it allows teachers to have comparative advantages in teaching certain types of students. In what follows, I use the estimated model to study whether some classrooms may benefit more from a particular type of teacher. I use the classroom achievement gain, Equation (2.17), to summarize the effects of changing teachers' general teaching skills and marginal costs of effort on academic achievement considering the behavioral responses of teachers and students. To explore the heterogeneous effects of these changes, I evaluate the gains at different points of the classroom initial knowledge distribution. In particular, I calculate the gains of reassignment for values of $\ln \bar{K}_{cj}$ ranging from 1.5 standard deviations below the mean (i.e., at $\ln \bar{K}_{cj} = -1.132$), to 1.5 standard deviations above the mean (i.e., at $\ln \bar{K}_{cj} = 1.01$).

To better understand the importance of increasing students' access to teachers with higher general teaching skills, I study the gains of replacing an average teacher (with average endowments) with a teacher who has a general teaching skill that is two standard deviations above the mean. The first panel from Figure 2.1 graphs the gains from the proposed reassignment against the initial classroom knowledge (X-axis). The solid line shows the model-predicted gains. The dashed line is the gain, ignoring the heterogeneity in the parameters governing the marginal product of the teaching inputs and teachers' preferences and equalizing them to their corresponding means. As seen, the proposed reassignment produces higher gains in classrooms with a higher initial knowledge; ignoring the heterogeneity in the parameters overstate the gains in classrooms with a below-average initial knowledge, while the opposite is true for classrooms with an above-average initial knowledge. On an average classroom, the proposed reassignment increase achievement by .55 standard deviations. If the initial classroom knowledge was set to be 1.5 standard deviations below the mean, the reassignment would increase the classroom's achievement by .51 standard deviations. In contrast, classrooms with an initial knowledge that is 1.5 standard deviations above the mean would increase their knowledge by .565 standard deviations due to the reassignment. Overall, Figure 2.1 suggests that increasing the general teaching skill is a practical mechanism to increase students' achievement, and classrooms with a higher initial knowledge benefit more from the proposed reassignment.

To better understand the mechanisms behind the achievement gain, Panel (b) from Figure 2.1 decompose the overall gain into the effect of changing students' and teachers' efforts, as well as the direct effect of increasing the general teaching skill. The dashed line in Panel (b) shows the direct effect of increasing the general teaching skill by two standard deviations above the mean. As seen, this effect is higher in classrooms with higher initial knowledge. This is due to the higher marginal product of the general teaching skill, captured by α_{Gcj}, in classrooms with higher initial knowledge. Similarly, the dotted line shows that the gains from the increased students' effort are higher in classrooms with higher initial knowledge. In contrast, the solid line shows that the gains due to changes in teachers' efforts decrease with the initial classroom knowledge. This decrease is due to the lower marginal product of the teaching effort in classrooms with higher initial knowledge, which ameliorates the overall gains.

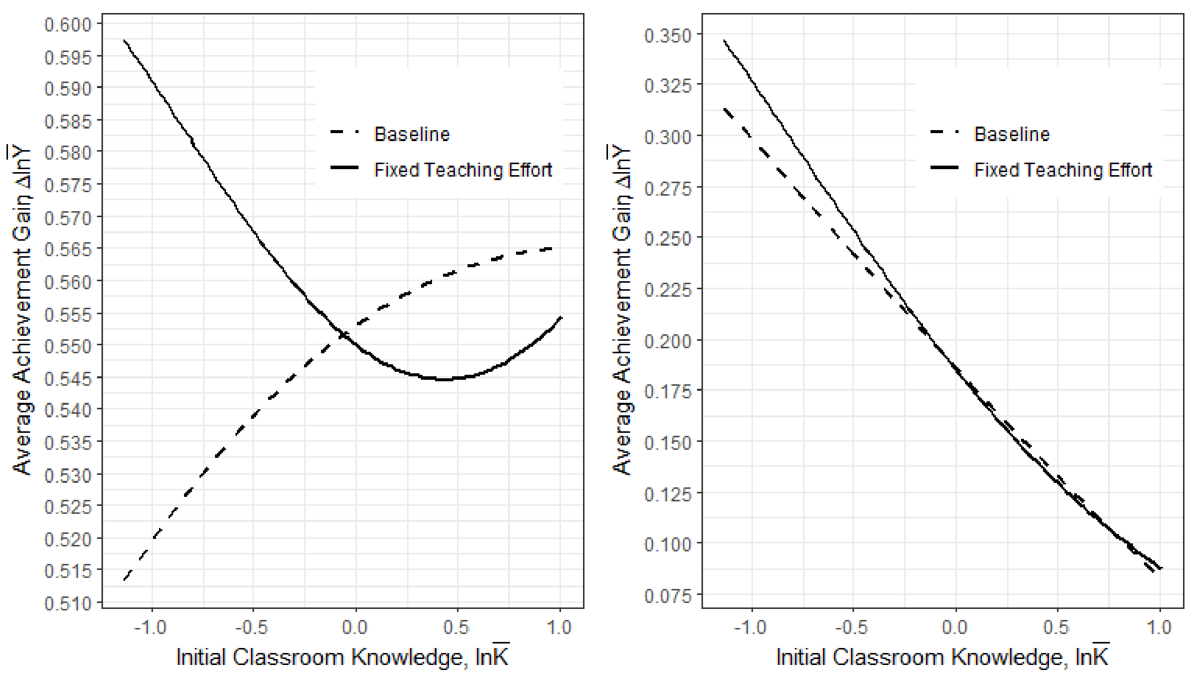
The first panel of Figure 2.2 shows the gains from replacing an average teacher with a teacher who has a marginal cost of effort that is two standard deviations below the mean. As seen by the intersection between the solid and dashed line, an average classroom can increase its knowledge by .195 standardized points due to the reassignment. However, this average masks important heterogeneity in the gains. In particular, a classroom that is 1.5 standard deviations below the mean would gain .313 standardized points due to the reassignment, while a classroom that is 1.5 standard deviations above the mean would only gain .087 points. The second panel of Figure 2.2 shows that both efforts' effects are decreasing on the initial classroom knowledge. Overall, Figure 2.2 shows that decreasing teachers' cost of effort leads to a considerable increase in students' achievement. As opposed to the general teaching skill, a reduction in teachers' per-unit cost of effort has a bigger impact in classrooms with lower initial achievement.

An important feature of the theoretical model is that it predicts the counterfactual amount of effort teachers would exert in the reassigned classroom. Not accounting for these endogenous choices may generate biased predictions of the achievement gains. To give a better sense of the impact of the efforts' adjustments, I contrast the equilibrium gains of reassignment from Figure 2.1 and Figure 2.2, with the gains ignoring the endogenous effort adjustments of the reassigned teacher. In particular, I set the reassigned teacher's effort equal to the effort she would exert in



(a) Overall Effect (b) Effect Decomposition

Figure 2.2: Achievement Gains due to Changes in the Marginal Cost of Effort



(a) Change in General Teaching Skill (b) Change in Marginal Cost of Effort

Figure 2.3: Gains of reassignment: Baseline Versus Fixed Teaching Effort

an average classroom. The first panel from Figure 2.3 graphs the achievement gains from increasing the general teaching skill by two standard deviations above the mean. The dashed line represents the baseline case, taking into account teacher effort responses, while the solid line shows the gains, ignoring teacher effort adjustments. As seen by the solid line, not accounting for teaching effort adjustments overstate the achievement gains by .10 standard deviations in classrooms that are 1.5 standard deviations below the mean and understate the gain by 0.01 standard deviations in classrooms that are 1.5 standard deviations above the mean.[32] These results suggest that not accounting for the teaching efforts' adjustments may generate large biases, especially for classrooms at the lower tail of the initial knowledge distribution. The second panel from Figure 2.3 shows a similar pattern from bringing a teacher with a marginal cost of effort that is two standard deviations below the mean, but the biases are smaller.

2.9 Counterfactual Simulations

I now use the estimated model to simulate policy-relevant reassignments of teachers to low-performing classrooms. The purpose of this exercise is twofold: to study the extent to which standard value-added models overpredict teachers' effectiveness in low-performing classrooms and to show how much could be gained by considering the endogenous interactions between teachers and students. The counterfactuals are based on reassignment interventions aimed at reducing the achievement gap between disadvantaged students and the rest. These interventions targeted classrooms that had low average test scores in the past or a high proportion of minority students. Additionally, I consider the commonly suggested intervention of replacing the least-effective teachers, who usually teach in low-performing classrooms, with the most-effective teachers, as measured by their value-added scores.

In the first counterfactual, I reassign teachers who are measured as highly effective using value-added estimates to classrooms ranked at the bottom of the (simulated) distribution of lagged math test scores. This counterfactual is motivated by a recent policy intervention, the Talent Transfer Initiative (TTI) (Glazerman et al., 2013), which provided incentives to highly effective teachers, based on value-added measures, to transfer to low-achieving schools, as measured by the previous year's achievement. The TTI was implemented in 114 schools in ten districts across seven states over two academic years, 2009-2010 and 2010-2011. The program targeted teachers in the top 20 percent of the value-added distribution and schools with the lowest achievement in the district. The TTI used an experimental design, where the student achievement in treated schools was compared to achievement in comparable schools in a control group. However, the TTI did not show a significant impact of the transfers on the math achievement of middle school students (6^{th} to 8^{th} grade).

In the second counterfactual, I reassign teachers in the top 20 percent of the value-added distribution to classrooms with high proportions of minority students. This counterfactual is motivated by recent evidence from the Intensive Partnerships for Effective Teaching (IP) initiative (Stecher et al., 2018), which was designed to increase low-income minority students'

[32]The solid line becomes increasing when the effect of a higher general teaching skill starts to dominate the other effects.

access to effective teachers by, among other things, revising recruitment and hiring practices and adjusting placement and transfer procedures.[33] As a whole, the IP initiative did not have a significant impact on low-income minority students' achievement. Though multiple factors could explain this result, I examine one potential explanation: the possibility that teachers who are measured as highly effective using a value-added specification may not perform as well in classrooms with higher proportions of minority students, who tend to have lower initial stocks of knowledge.[34]

The final counterfactual is driven by a commonly suggested policy of replacing teachers based on their value-added estimates (Hanushek, 2011; Chetty et al., 2014b). Teachers with lower value-added estimates tend to be sorted into classrooms with high proportions of low-achieving students. To keep the analysis comparable to the previous counterfactuals, I select the bottom 20 percent teachers of the simulated value-added distribution and replace them with teachers in the top 20 percent of the same distribution following a similar descending order. Accordingly, the top teachers selected for this counterfactual are the same as the top teachers from the previous counterfactuals.

For the first two counterfactuals, I first simulate the model and estimate teachers' value-added using fixed effects.[35] This closely matches the commonly used specifications in the literature. Second, I rank teachers based on their estimated value-added and select the top 20 percent in each district, 30 teachers in total. Third, I rank classrooms based on the selection criteria and keep the first thirty (top/bottom 10% of the classroom distribution). Fourth, I sort the selected teachers into the selected classrooms in decreasing order (i.e., the best teacher to the worst classroom, the second-best teachers to the second-worst classroom, etc.). Fifth, for the selected classrooms, I predict the change in achievement from the standard value-added model, which equals the difference between the value-added estimates of the selected teachers and the value-added estimates of the original teachers (see Equation (2.19)). Finally, I reassign the selected teachers to the selected classrooms and simulate the model again to obtain the performance of the selected classrooms after the transfers, considering the behavioral response of teachers and students. For the final counterfactual, I select the bottom 20 percent teachers of the simulated value-added distribution and replace them with teachers in the top 20 percent of the same distribution following a similar descending order. I repeat this process 100 times and report the averaged results.

[33]The IP initiative defines low-income as eligible for free or reduced-price lunch, and minority students as those classified as Black, Hispanic, or Native American.

[34]I focus only on minority students because reduced-price lunch status is missing for one of the districts.

[35]The value-added specification controls for the students' observable characteristics (excluding the twice-lagged math score), the lagged math state test score, the previous variables averaged at the classroom level, and the classroom size.

Table 2.10: Achievement Gap

	Coefficient	SE
(1) Lowest Lagged Math Score		
Achievement Gap	-1.052	.040
(2) Highest Share of Minority Students		
Achievement Gap	-0.586	0.028
(3) Bottom 20 percent Teachers		
Achievement Gap	-0.667	0.040

Note: Each row is a separate regression. The dependent variable is math test score and sample size is 6,603 students. Each regression controls for an indicator of whether the student belongs to a classroom at the bottom 20 percent of the lagged score (Panel 1), the top 20 percent of the minority distribution (Panel 2), and bottom 20 percent of the value-added distribution (Panel 3).

Table 2.10 shows the achievement gap between students in low-performing classrooms and other students in the sample. As seen in the first panel of Table 2.10, the average math test score of students in classrooms at the bottom 10 percent of the lagged score distribution is approximately 1.05 of a standard deviation below the rest of the students in the sample. The second panel of Table 2.10 shows the average state test score for classrooms with a high proportion of minority students is approximately .59 standard deviations below the rest. The third panel of Table 2.10 shows that students assigned to teachers at the bottom 20 percent of the value-added distribution perform approximately .66 standard deviations below other students'. Overall, these statistics highlight considerable gaps between groups.

The first row of Table 2.11 summarizes the results of the counterfactual reassignment. The first column shows that the predicted increase in students' math scores due to the reassignment is .385 standard deviations. However, as shown in the second column, the increase after allowing for teachers' and students' behavioral responses is .247 of a standard deviation, which is .138 standard deviations points below expected. The increase is considerably smaller than was initially predicted by the value-added estimates. Thus, the simulated intervention reduces approximately one-quarter of the achievement gap (.235 percent of the achievement gap).

Table 2.11: Counterfactual Reassignment of Teachers to Classrooms

	Simulations	
	Value-Added Change	Unrestricted Change
(1) Reassignment Using Value-Added		
Bottom 20 percent Lagged Math Score	.385	.247
Top 20 percent Minority	.345	.235
Bottom 20 percent Teachers	.657	.507
(2) Reassignment Using Model		
Bottom 20 percent Initial Knowledge	-	.354
Top 20 percent Minority	-	.321
Bottom 20 percent Teachers	-	.570

To examine the potential gains of transferring teachers to the selected classrooms, I simulate the increase in students' achievement if teachers were reassigned using the structural model instead of their estimated value-added. This simulation may provide better outcomes because the model considers changes in the marginal product of the inputs, effort adjustments, and the full set of students' and classrooms' endowments. To find the matches, I first simulate the model for each classroom– teacher combination within each district and calculate the expected classroom performance. I then choose the most effective teacher for each selected classroom. I simulate the model under the new assignments and estimate the potential gains for the selected classrooms. The first row of the second panel of Table 2.11 shows the result of this intervention. The reassignment using the model increases students' math scores by .354 standard deviations, or equivalently, reduces approximately one third of the achievement gap, which is similar to what was predicted by the wrongly specified value-added specification, but considerably higher than (.107 standard deviations above) the actual change when using the value-added specification instead.

For the second counterfactual, I follow a similar procedure to the previous one, but this time I rank classrooms based on the proportion of minority students. Similar to the first counterfactual, the second row of table 2.11 shows that the expected achievement gain using the value-added specification is .345 standard deviations, which is considerably higher than the actual gain shown in the second column (.235 standard deviations, or 40 percent of the achievement gap). The second panel shows that when the structural model is used to perform the reassignments, the potential gain for the selected classrooms would be .321 standard deviations, which represents a reduction of approximately 55 percent of the achievement gap.

The third row of the first panel of Table 2.11 summarizes the results of replacing teachers based on value added. The first column shows an expected increase of .66 standard deviations. In contrast, the second column shows that the actual increase is .513 standardized points (77 percent of the achievement gap) which is .147 standardized points below what is predicted by the value-added estimates. However, the optimal assignment increases the selected students' achievement by .570 standardized points,

As seen in Table 2.11, all counterfactuals show a gap between the predicted and actual change in students' achievement when using the value-added specification. This gap may be due to several factors. First, the value-added specification does not control for students' home environments and classroom environments, which, as shown in Table 2.6, are correlated with the teachers' endowments. Not accounting for these endowments may bias the estimates of teachers' effectiveness due to omitted variable bias. Second, the impact of the students' lagged math scores is downward biased due to measurement error, which may also affect the estimates of teachers' effectiveness. Finally, the linear value-added model is misspecified because it does not allow the contribution of the teachers' inputs to vary across classrooms.

Table 2.12: Addressing Omitted-Variable Bias and Measurement Error

	Simulations	
	Value-Added Change	Unrestricted Change
Reassignment Using Value-Added		
Bottom 20 percent Lagged Math Score	.353	.287
Top 20 percent Minority	.317	.268
Bottom 20 percent Teachers	.617	.601

To better understand the importance of these factors, Table 2.12 shows the results of estimating teachers' value added considering the full set of students' and classrooms' latent endowments. Accordingly, any discrepancy between the predicted and actual changes in students' performance would be due to changes in the marginal product of the teaching inputs and effort adjustments. Similar to Table 2.9, the first row of Table 2.12 shows the predicted and actual changes in students' achievement for classrooms at the bottom 20 percent of the lagged score distribution. As shown, the actual change in students' achievement is .287 standard deviations, which is .04 standard deviations higher than the results from Table 2.9, but still, .067 standard deviations lower than the gains when the structural model is used instead. The results for classrooms at the top 20 percent of the minority distribution follow a similar pattern. The third row of Table 2.12 shows that after controlling for all students' and classroom endowments, the predicted changes closely match the actual changes in achievement for students who were assigned to teachers at the bottom 20 percent of the value-added distribution.[36]

Overall, the counterfactuals suggest that not accounting for students' and teachers' behavioral interactions in the classroom and the full set of endowments would overstate the predicted effectiveness of teachers in classrooms with a high proportion of low-achieving students. As opposed to the value-added specification, the reassignment from the structural model produces higher gains for the low-achieving students.

2.10 Conclusions

In this paper, I develop and estimate a model in which teachers and students exert effort to produce knowledge. The model expands upon standard value-added specifications by allowing teachers' effectiveness to differ across classrooms. The model fits multiple patterns observed in the data very well.

The estimated model shows that the contribution of the teaching inputs to students' achievement is heterogeneous across classrooms. In particular, teachers with high general teaching skills tend to be more effective in teaching students with high initial knowledge. On the contrary, teachers with low per-unit effort costs tend to be more effective in teaching students with lower initial knowledge.

I use the estimated model to perform multiple policy-relevant counterfactuals. The counterfactual results are consistent with recent findings in the literature. I find limitations in using standard value-added specifications to study the potential impact of commonly suggested reassignment interventions targeting low-performing students. In particular, transferring teachers who are thought to be highly effective to classrooms with below-average test scores reduces one-quarter of the achievement gap, which is considerably lower than initially predicted. Additionally, I find that much could be gained by considering the endogenous effort choices of teachers and students. In particular, the same reassignment using the model can reduce one-third of the achievement gap.

[36]However, this reassignment policy does not consider the potential gains of taking into account changes in the marginal products and effort adjustments.

These results have important implications for educational policy design. Student surveys recently started to be used by schools, contain rich information that can be used to evaluate the role of teachers in promoting students' knowledge. My findings suggest that accounting for teachers' behavior and interactions with students in the classroom may considerably increase the impact of commonly suggested reassignment policy interventions. Furthermore, by distinguishing teaching practices from teachers' skills and preferences, policymakers may evaluate the role of different teaching inputs in a classroom

Bibliography

Agostinelli, Francesco, Morteza Saharkhiz, and Matthew J Wiswall. 2019. Home and school in the development of children. Tech. rep., National Bureau of Economic Research.

Agostinelli, Francesco and Matthew Wiswall. 2016a. Estimating the technology of children's skill formation. Tech. rep., National Bureau of Economic Research.

———. 2016b. Identification of dynamic latent factor models: The implications of re-normalization in a model of child development. Tech. rep., National Bureau of Economic Research.

Ahn, Tom, Esteban Aucejo, Jonathan James et al. 2021. The importance of matching effects for labor productivity: Evidence from teacher-student interactions. Tech. rep.

Aucejo, Esteban, Patrick Coate, Jane Cooley Fruehwirth, Sean Kelly, and Zachary Mozenter. 2021. Teacher effectiveness and classroom composition: Understanding match effects in the classroom. *Working Paper* .

Aucejo, Esteban, Jane Cooley Fruehwirth, and Sean Kelly. 2020. Teachers and the gender gap in reading achievement. *Working Paper* .

Blazar, David and Matthew A Kraft. 2017. Teacher and teaching effects on students' attitudes and behaviors. *Educational evaluation and policy analysis* 39, no. 1:146–170.

Chetty, Raj, John N Friedman, and Jonah E Rockoff. 2014a. Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review* 104, no. 9:2593–2632.

———. 2014b. Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American economic review* 104, no. 9:2633–79.

Cunha, Flavio, James J Heckman, and Susanne M Schennach. 2010. Estimating the technology of cognitive and noncognitive skill formation. *Econometrica* 78, no. 3:883–931.

Ferguson, Ronald F and Charlotte Danielson. 2015. How framework for teaching and tripod 7cs evidence distinguish key components of effective teaching. *Designing teacher evaluation systems: New guidance from the measures of effective teaching project* :98–143.

Fox, Lindsay. 2016. Playing to teachers' strengths: Using multiple measures of teacher effectiveness to improve teacher assignments. *Education Finance and Policy* 11, no. 1:70–96.

Garrett, Rachel and Matthew P Steinberg. 2015. Examining teacher effectiveness using classroom observation scores: Evidence from the randomization of teachers to students. *Educational Evaluation and Policy Analysis* 37, no. 2:224–242.

Gershenson, Seth, Cassandra Hart, Joshua Hyman, Constance Lindsay, and Nicholas W Papageorge. 2018. The long-run impacts of same-race teachers. Tech. rep., National Bureau of Economic Research.

Glazerman, Steven, Ali Protik, Bing-ru Teh, Julie Bruch, and Jeffrey Max. 2013. Transfer incentives for high-performing teachers: Final results from a multisite randomized experiment. *National Center for Education Evaluation and Regional Assistance* .

Graham, Bryan S, Geert Ridder, Petra M Thiemann, and Gema Zamarro. 2020. Teacher-to-classroom assignment and student achievement. Tech. rep., National Bureau of Economic Research.

Greene, William. 2004. The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *The Econometrics Journal* 7, no. 1:98–119.

———. 2009. Discrete choice modeling. In *Palgrave handbook of econometrics*. Springer, 473–556.

Greene, William H. 2018. Econometric analysis.

Hanushek, Eric A. 2011. The economic value of higher teacher quality. *Economics of Education review* 30, no. 3:466–479.

Hanushek, Eric A and Steven G Rivkin. 2006. Teacher quality. *Handbook of the Economics of Education* 2:1051–1078.

Horvath, Hedvig. 2015. Classroom assignment policies and implications for teacher value-added estimation. *Unpublished manuscript* .

Jackson, C Kirabo. 2018. What do test scores miss? the importance of teacher effects on non–test score outcomes. *Journal of Political Economy* 126, no. 5:2072–2107.

Jackson, C Kirabo, Jonah E Rockoff, and Douglas O Staiger. 2014. Teacher effects and teacher-related policies. *Annu. Rev. Econ.* 6, no. 1:801–825.

Kane, Thomas, Kerri Kerr, and Robert Pianta. 2014. *Designing teacher evaluation systems: New guidance from the measures of effective teaching project*. John Wiley & Sons.

Kane, Thomas J, Daniel F McCaffrey, Trey Miller, and Douglas O Staiger. 2013. Have we identified effective teachers? validating measures of effective teaching using random assignment. In *Research paper. met project. bill & melinda gates foundation*. Citeseer.

Kane, Thomas J, Eric S Taylor, John H Tyler, and Amy L Wooten. 2011. Identifying effective classroom practices using student achievement data. *Journal of human Resources* 46, no. 3:587–613.

Koedel, Cory, Kata Mihaly, and Jonah E Rockoff. 2015. Value-added modeling: A review. *Economics of Education Review* 47:180–195.

Kuhfeld, Megan. 2017. When students grade their teachers: A validity analysis of the tripod student survey. *Educational Assessment* 22, no. 4:253–274.

Lavy, Victor. 2015. What makes an effective teacher? quasi-experimental evidence. *CESifo Economic Studies* 62, no. 1:88–125.

Lazear, Edward P. 2001. Educational production. *The Quarterly Journal of Economics* 116, no. 3:777–803.

Lockwood, JR and Daniel F McCaffrey. 2009. Exploring student-teacher interactions in longitudinal achievement data. *Education finance and policy* 4, no. 4:439–467.

Loeb, Susanna, James Soland, and Lindsay Fox. 2014. Is a good teacher a good teacher for all? comparing value-added of teachers with their english learners and non-english learners. *Educational Evaluation and Policy Analysis* 36, no. 4:457–475.

Raudenbush, Stephen W and Marshall Jean. 2015. To what extent do student perceptions of classroom quality predict teacher value added. *Designing teacher evaluation systems: New guidance from the measures of effective teaching project* :170–202.

Rothstein, Jesse. 2017. Measuring the impacts of teachers: Comment. *American Economic Review* 107, no. 6:1656–84.

Stecher, Brian M., Deborah J. Holtzman, Michael S. Garet, and et. al. 2018. *Improving teaching effectiveness: Final report: The intensive partnerships for effective teaching through 2015 - 2016.* Santa Monica, CA: RAND Corporation.

Steinberg, Matthew P and Morgaen L Donaldson. 2016. The new educational accountability: Understanding the landscape of teacher evaluation in the post-nclb era. *Education Finance and Policy* 11, no. 3:340–359.

Steinberg, Matthew P and Rachel Garrett. 2016. Classroom composition and measured teacher performance: What do teacher observation scores really measure? *Educational Evaluation and Policy Analysis* 38, no. 2:293–317.

Steinberg, Matthew P and Matthew A Kraft. 2017. The sensitivity of teacher performance ratings to the design of teacher evaluation systems. *Educational Researcher* 46, no. 7:378–396.

Taylor, Eric S. 2018. Skills, job tasks, and productivity in teaching: Evidence from a randomized trial of instruction practices. *Journal of Labor Economics* 36, no. 3:711–742.

Todd, Petra and Kenneth I Wolpin. 2018. Accounting for mathematics performance of high school students in mexico: Estimating a coordination game in the classroom. *Journal of Political Economy* 126, no. 6:2608–2650.

Wallace, Tanner LeBaron, Benjamin Kelcey, and Erik Ruzek. 2016. What can student perception surveys tell us about teaching? empirically testing the underlying structure of the tripod student perception survey. *American Educational Research Journal* 53, no. 6:1834–1868.

White, M, B Rowan, G Alter, and C Greene. 2014. User guide to the measures of effective teaching longitudinal database (met ldb). *Ann Arbor, MI: Inter-University Consortium for Political and Social Research, The University of Michigan* .

Chapter 3

Group or Individual Teacher Bonuses? An Estimation of the Potential Gains

3.1 Introduction

Student performance in standardized exams has been shown to be an important determinant of their future wealth. For example, in India, increasing test performance by one standard deviation has been found to increase earnings from 16 to 20 percent (Aslam et al., 2011), while in the United States, the average increase in earnings is around 12 percent (Hanushek, 2011). Further, vast literature shows that teachers play a crucial role in determining students' performance, with just one year of exposure to a high-quality teacher having a lifelong positive impact on a student's wealth (Chetty et al., 2014).[1]

The proven importance of teachers has motivated the creation of teachers' incentive programs worldwide. These programs reward teachers for the extra effort required to improve test scores.[2] On this regard, the empirical literature shows a variety of incentive schemes that have been evaluated under randomized controlled trials.[3] But still, there is no conclusive evidence about their relative performance. A major challenge is that these incentive programs were not necessarily designed using the same incentive schemes, budgets, and objective functions. Thus, it is not possible to make a meaningful comparison between them. However, given the importance of teachers, providing optimal incentive schemes should be a priority for policymakers. This paper is an attempt to fill this gap.

In this paper, I estimate a model of teacher effort under two different linear bonus schemes: an individual-based (II) and a group-based (GI) piece rate bonus. I use the model to bet-

[1]For a review see Koedel et al. (2015), Jackson et al. (2014) and Murnane et al. (2014)

[2]As an example, consider two popular incentive schemes in education: Tournaments and piece rate bonuses. In the first case, teachers are ranked and paid according to their relative performance. In the second case, their payment is a linear function of their absolute performance. Piece rate bonuses can also be divided into individual- or group-based categories. In the first case, a teacher's payment depends on their own performance, while in the second case their payment is a weighted average of the team (school) performance.

[3]Muralidharan and Sundararaman (2011), Glewwe et al. (2010), Barrera-Osorio and Raju (2017), Duflo et al. (2012), Loyalka et al. (2016) and Behrman et al. (2015) are examples for developing countries.

ter understand the performance of these bonuses if they were designed to maximize the expected value of student academic performance, minus the expected payment to teachers. I use data from Muralidharan and Sundararaman (2011), a teacher incentive experiment in Andhra Pradesh, India, which has the unique feature of containing both types of incentives, and a control group.

The experiment randomly assigned schools to a control group and two incentive schemes that paid teachers a bonus based on test-score improvements. These bonuses were designed to be jointly equivalent to providing schools with Rp.10,000 in inputs. Both schemes followed the same piece rate bonus formula but averaged test scores at different levels. The II scheme is based on average teacher-level improvements, while the GI scheme is based on average school-level improvements.

Even though the experiment shows that the II scheme produces a higher average treatment effect (ATE) than the GI scheme, it is not clear which one would be preferred if they were optimally designed. This is because GI schemes are subject to potential drawbacks and advantages that could be considered in their design. Accordingly, an optimally designed GI scheme may have a different bonus formula and produce better results than an optimally designed II scheme.

A possible deficiency in GI is the presence of the so-called free-riding effect, as discussed in Imberman and Lovenheim (2015). If teachers' responsibility over the total "achievement production" decreases, they may exert less effort because their impact on the total production would be smaller. Thus, the incentive strength is reduced. However, there are also potential benefits of using group incentives. First, averaging the outputs of multiple teachers reduce noise, which make these scheme more attractive for risk-adverse teachers. Second, peer pressure can increase effort because teachers will suffer if they deviate from the team norms (Kandel and Lazear, 1992).

My empirical strategy allows me to disentangle the importance of these mechanisms in the GI scheme. First, I model the piece rate bonuses taking into account the possibility of free-riding and peer-pressure in the GI scheme. I do so by modifying the well-known moral hazard model with an exponential utility function and constant absolute risk aversion (see, for example, Bolton and Dewatripont 2005 and Mehta 2017). Second, I exploit the experimental nature of the data to test for the presence of free riding and peer pressure in the GI scheme. Third, I estimate the structural parameters and use them to counterfactually recover the II and GI scheme that maximizes the expected value of students' performance, minus the expected payment to teachers.

My results suggest that teachers with a small share of students are subject to higher peer-pressure which ameliorates the free-riding incentives. Therefore, I take into account the observed *heterogeneous* effect of peer-pressure to model the optimal GI bonus scheme. I find that, contrary to the original design, the GI scheme is preferred to the II scheme. This highlights the importance of evaluating optimally designed contracts.

There exists a rich literature on randomized controlled trials for teacher incentives in schools. These papers' analyze the effect of a variety of schemes that were not necessarily designed to maximize student achievement. In addition, the results have been mixed on the success of such schemes. Glewwe et al. (2010) analyze an unique tournament scheme between schools in Kenya. They find mixed evidence on student performance and no variation on the reward formula. Using experimental data from Pakistan, Barrera-Osorio and Raju (2017) randomly assign teachers to different types of piece-rate cash bonus schemes and find no signs of exam score improvements. Muralidharan and Sundararaman (2011) and Muralidharan (2012) find a positive impact on student grades for both types of piece rate contracts with a larger effect of II compared to GI after the first year of the program. I show that this is not necessarily optimal if the policymaker wants to maximize the present value of student performance, net the cost of paying teachers.

Mehta (2017) is the closest paper to mine. He uses Muralidharan and Sundararaman (2011)'s results to calibrate the parameters of a utility function for the II scheme and then recovers the optimal II linear scheme. I focus on the GI scheme and consider different mechanisms that may influence its performance. Then, I study the performance of an optimally designed GI scheme and compare it with an optimally designed II scheme.

My paper is also closely related to Imberman and Lovenheim (2015), who analyze a tournament between schools in Houston, Texas and is therefore subject to teacher free-riding. Their primary goal is to test whether the share of students a teacher teaches affects the incentive strength. My paper uses a similar approach, and test for the presence of different models of peer-pressure to rationalize the results. Another related paper is Macartney et al. (2015) who proposes a semi-parametric method to retrieve the relationship between the incentive strength and teachers' effort and then estimates the optimal contract. I complement their analysis by focusing on the relative performance the GI scheme taking into account free-riding and peer pressure.

There is an ongoing debate about the benefits of using an atheoretical approach with experimental data versus using structural estimation.[4] I benefit from both approaches. First, the experimental data allow me to identify causal effects and test different GI models. Second, the contracts are exogenously assigned. Therefore, I do not need to model that selection problem, what reduces the number of assumption I need to identify the parameters[5].

The paper is organized as follows. In Section 3.2 I give a brief review of the Andra Pradesh dataset. Section 3.4 presents a model of optimal teachers' effort. In section 3.5, I present the empirical methodology that I am going to use to test the model's implications. Finally, in section 3.8, I estimate the structural parameters to recover the optimal contracts.

3.2 The Andra Pradesh Experiment

The Andra Pradesh experiment is a randomized controlled trial that took part in Andra Pradesh, India over three years (2005-2007) (see Muralidharan and Sundararaman 2011 for additional details about the experiment). The experiment consisted of four different treatments, which

[4] See, for example, Angrist and Pischke (2010) and Keane (2010)

[5] Bellemare et al. (2016) provides a summary of the benefits of combining structural estimation with experiments in personnel economics.

were effective after the first year, and a control group. From a total of 500 randomly selected schools, 100 were assigned into a teacher's individual-based piece rate incentive scheme (II); 100 were assigned into a group-based piece rate scheme (GI), and 100 schools were assigned to a control group. The other schools were assigned either to an extra contract teacher or a block grant.

Students from first to fifth grade comprised the population under study and took a set of standardized tests to measure their knowledge. Once a year, students had to take two tests in two subjects - math and Telugu (language) - that covered material up to the previous year (first test) and the current year (second test). This was done with the objective of covering a broad range of topics and reducing measurement error. Teachers were told that the program would last for at least two years and after that it will be subject to budget availability[6]. For now on, year 0 (June-July 2005) refers to the baseline test period, while year 1 (March-April 2006) refers to the first period with treated schools; and year 2 (March-April 2006) refers to the second period of treatment.

In years 1 and 2, treated teachers were exposed to a bonus payment on top of their baseline wage that was given by the following formula:

$$Bonus_t = \begin{cases} 500 \times (\bar{q}_t^k - \bar{q}_{t-1}^k) & if \quad \bar{q}_t^k - \bar{q}_{t-1}^k > T \\ 0 \quad Otherwise, \end{cases} \tag{3.1}$$

where $k \in \{II, GI\}$ is an indicator for the type of scheme, $t \in \{1, 2\}$ is the program year, and $\bar{q}_t^k$ represents students' average score in year t. For the II scheme, $\bar{q}_t^{II}$ represents the average performance for all students assigned to a teacher. Meanwhile, for the GI scheme, $\bar{q}_t^{GI}$ represents the average school performance. The difference between scores, $\bar{q}_t^k - \bar{q}_{t-1}^k$, is the percentage gain (or loss) in the average test scores from year $t-1$ to year t. During the second year, the schemes had a $T = 0$ threshold. However, in year 1, a $T = .05$ punishment was introduced because the year 0 tests were taken in June-July instead of March-April. Thus, students might forget material during the summer vacation making $\bar{q}_0^k$ smaller than what it should be to fairly compare it with $\bar{q}_1^k$.

There exist potential "threshold effects" that could undermine my analysis. First, the first year 5% threshold could disincentivize teachers who expect to produce less than 5%. However, Muralidharan and Sundararaman (2011) report that there is no evidence of this effect. Second, the bonus does not punish teachers, so they could strategically produce less in year 1 and more in year 2 to maximize the present value gains. To reduce the dynamic incentives, the year 2 lagged score was set to be the maximum between year 0 and year 1 classroom (school) scores. Third, teachers could have incentives to only focus on good students. To avoid this issue, low grades were assigned to students who did not take test.

[6]The experiment was extended for three more years with small changes in the incentives design (Muralidharan, 2012), but the database is not publicly available

3.3 Sample Selection

In this paper, I focus on the first year of the experiment and leave the second year for a future analysis considering dynamic aspects of the incentive schemes. I use the normalized scores provided by Muralidharan and Sundararaman (2011): year 0 scores are normalized relative to the distribution of scores across all schools for the same test. Scores in years 1 and 2 are normalized with respect to the score distribution of the control group for the same test level and subject, and then averaged to obtain an unique normalized score for each student.

Muralidharan and Sundararaman (2011) provides a detailed description of the data. In this section, I focus on the teachers' share. To test for the presence of free riding in the GI scheme, it is necessary to construct the teachers' share of assigned students. The bonus GI formula implies that the share of total output that corresponds to each teacher is given by the number of students that they are in charge of divided by the total number of students that in the school.

Table 3.1: Descriptive Statistics of Teachers' Share

	Year 1			
	Control	II	GI	p-Value
	(1)	(2)	(3)	(4)
Teacher Share:				
Average	.312	.307	.315	0.87
Standard Deviation	.155	.158	.159	
5% percentile	.129	.139	.138	
50% percentile	.261	.246	.253	
95% percentile	.630	.622	.632	
N of teachers per school:				
Average	3.2	3.26	3.17	0.87
Min	2	1	2	
Max	5	5	5	
N of Schools:	100	100	100	
N of Teachers:	320	326	317	

Table 3.1 provides descriptive statistics for teacher shares and the number of teachers in the school for each treatment group. As seen, there are approximately 3 teachers per school, which implies that each teacher is responsible of 31% of the students. The 95% percentile corresponds to a share of 63% of students, and the standard deviation is approximately .16 standard deviations, implying that very few teachers are responsible of a large share of students.

3.4 Theoretical Model

In this section, I develop a static hidden action model with an exponential utility function and normally distributed shocks. The model embeds the II and the GI schemes and is based on the workhorse CARA-Normal model from Bolton and Dewatripont (2005). The objective of this section is to predict the level of teacher effort under different schemes. To do so, I assume that teachers are willing to participate in a given bonus scheme, and I leave the details of the optimal design to section 3.8.1.

I define the education output as the test score q for student i assigned to teacher j. The educational production function is given by

$$q_{ij} = s_{ij} + e_j + \epsilon_{ij}, \tag{3.2}$$

where s_{ij} represents student i's expected performance without the presence of teacher effort, e_j is teacher effort, and ϵ_{ij} is an idiosyncratic independent error with variance σ^2. The variable s_{ij} captures characteristics of the students and teachers that directly affect the production of knowledge, such as teachers' innate instructional quality. As such, teacher j can contribute to student i's outcome even when she does not exert effort.

I define $\bar{q}_j$ as the average performance of students assigned to teacher j. The variance for the average performance is given by $\sigma_J^2 = \frac{\sigma_S^2}{N_j}$, where N_j is the number of students assigned to teacher j. Accordingly, the relevancy of the unobserved shock decreases as the number of students assigned to teacher j becomes larger.

Teachers are paid based on the average students' performance. The linear wage function for the GI scheme is given by

$$W^{GI}(e_j, e_{-j}) = \beta_0 + \beta_1 \sum_{j=1}^{J} \omega_j \bar{q}_j \tag{3.3}$$

where teacher j's share of students is given by $\omega_j := \frac{N_j}{\text{school size}}$. Naturally, the share is defined in the interval $\omega_j \in [0, 1]$. The parameter β_0 is the fixed compensation and β_1 represents the variable compensation (the contract slope). The average school performance is defined by $\sum_{j=1}^{J} \omega_j \bar{q}_j$, where J is the total number of teachers in a particular school, and e_{-j} is the effort of other teachers in the school. The wage equation for the II scheme is given by

$$W^{II}(e_j) = \beta_0 + \beta_1 \bar{q}_j, \tag{3.4}$$

which is equal to equation (3.3) after setting $\omega_j = 1$ and J=1. As seen, the II and GI schemes have the same contract slope, given by β_1. However, under the GI scheme, teachers' compensations depend on the share of students, ω_j, and the efforts of other teachers in the school.

The cost of effort, $C(e_j)$, is defined as a quadratic function $C(e_j) = \frac{1}{2}\frac{e_j^2}{\gamma}$, where γ is a parameter that captures the marginal cost of effort. Teachers under the GI scheme could be subject to peer-pressure. Based on Kandel and Lazear (1992), I model peer-pressure as the disutility that is generated by producing less than an established social norm that states how much teachers should produce. I assume that teachers can perfectly observe each other's effort. I propose a flexible peer-pressure function that is given by

$$P(e_j) = (\alpha - \rho\omega_j) \times (\bar{e} - e_j), \tag{3.5}$$

where $\bar{e}$ is the social norm, $\alpha \geq 0$ is the homogeneous punishment for deviating from it, and $\rho \geq 0$ captures the degree to which teachers with higher share are exposed to lower peer pressure. As seen by equation 3.5, teachers who exert less effort than the social norm would have a disutility due to peer pressure.

Teachers have a constant absolute risk aversion and an exponential utility that is given by

$$U^k(e_j) = -e^{-\xi[W^k - C(e_j) - \mathbb{I}^{GI} P(e_j,\omega_j)]} \tag{3.6}$$

where $k \in \{II, GI\}$, ξ is a teacher's coefficient of absolute risk aversion, and $\mathbb{I}^{GI}$ is an indicator for the GI scheme. As Bolton and Dewatripont (2005) show, maximizing the expected utility with respect to e_j is equivalent to maximizing the certainty equivalent compensation. Therefore, the optimal effort under the II scheme is given by

$$e_j^{II} = \beta_1 \gamma. \tag{3.7}$$

Accordingly, teachers' would exert more effort when their marginal cost is lower (i.e., higher γ) and when the bonus scheme is steeper (i.e., higher β_1). Under the GI scheme, the optimal level of effort is given by

$$e_j^{GI} = \omega_j \beta_1 \gamma + (\alpha - \rho\omega_j)\gamma. \tag{3.8}$$

The first term in equation (3.8) is the so-called free-riding effect, which is equal to the optimal effort under the II scheme multiplied by the share ω_j. Thus, teachers with a higher share will exert more effort. The second term in equation (3.8) captures the effect of peer pressure.

I consider three scenarios for the optimal teacher effort under the GI scheme. First, if there is no peer pressure, teachers would be only subject to free riding, and their effort would be

$$e_j^{fr} = \omega_j \beta_1 \gamma. \tag{3.9}$$

Second, if all teachers are subject to the same level of peer pressure, the optimal effort would be

$$e_j^{ho} = \omega_j \beta_1 \gamma + \alpha\gamma. \tag{3.10}$$

Under this scenario, teachers would exert a constant amount of effort on top of the aforementioned free-riding effect. Finally, if $\rho = \beta_1$, the optimal effort would be given by

$$e_j^{he} = \alpha\gamma, \tag{3.11}$$

implying that the optimal effort would not depend on the share of students being taught.

3.5 Testing the Group Incentive Models

The experimental nature of the Andra Pradesh dataset allows testing for the presence of free-riding and peer pressure using the students' outcomes, and the teachers' shares of students, ω_j. The empirical methodology exploits the predicted linearity of the teacher effort and the random assignment to treatments.

Thus, the ATE for the II scheme is given by

$$E[q_{ij} \mid I^{II}] - E[q_{ij} \mid I^{Control}] = E[e_j^{II}] = \beta 1 \, E[\gamma \mid I^{II}], \tag{3.12}$$

with a similar result holding for each GI model. Table 8 in Muralidharan and Sundararaman (2011) shows that the combined math and language ATE for the II scheme is 15.6 SD in year 1, while for the GI is 14.1 SD. These values are not statistically different from each other. Thus, this results cast doubt on the presence of pure free riding (without peer-pressure).

Still, it is not clear whether the free riding effect plays a role in the ATE estimate. For example, the models given by equations (3.10) and (3.11) could produce the same expected effort, but only the first one is subject to free-riding. However, the underlying mechanism that generates these results are crucial for the design of the optimal contract. Thus, it is necessary to disentangle the combined effects of free-riding and peer-pressure.

The main specification is based on the following regression that includes dummies for each treatment, and the interaction with ω_j:

$$\begin{aligned} q_{ij} = \pi_C + X_{ij}\beta + \pi_{II} I^{II} + \pi_{GI} I^{GI} + \\ \pi_W \omega_j + \pi_{IW} \times \omega_j \times I^{II} + \pi_{GW} \times \omega_j \times I^{GI} + \lambda_m + \epsilon_{ij} \end{aligned} \tag{3.13}$$

where q_{ij} is the normalized test score of student i taught by teacher j. The variables I^{II} and I^{GI} are indicators for each treatment group, and the variable ω_j represent teacher j's share. The vector X_{ij} includes student, teacher and school characteristics. The variable λ_m represents district fixed effects. The model is estimated pooling Math and Language outcomes, and the standard errors are clustered at the school level.

Based on equation (3.2), s_{ij} is approximated using a linear function of the previous year's standardized score and other characteristics included in X_{ij}. The parameter π_W recovers the expected performance of students assigned to teachers with a higher share of students. The parameters $\hat{\pi}_{II}$ and $\hat{\pi}_{GI}$ recover teachers' effort under each scheme, in the extreme case when $\omega_j = 0$.

The parameters of interest are π_{IW} and π_{GW}. The parameter π_{IW} captures any change in teacher effort due to changes in the share ω_j. A negative π_{IW} coefficient would imply that teachers with a higher share have a higher cost of effort, which may be due to teachers being assigned to larger, harder-to-teach classrooms or the selection of less able teachers to larger classrooms. The parameter π_{GW} captures the effects mentioned above, plus any additional incentives to exert effort due to free riding.

The empirical methodology is based on the derivative of equation (3.8) with respect to the share ω_j, which is given by

$$\frac{\partial e_j^{GI}}{\partial \omega_j} = \underbrace{\beta_1 \omega_j \frac{\partial \gamma}{\partial \omega_j}}_{\text{Selection}} + \underbrace{\beta_1 \gamma}_{\text{Free Riding}} + \underbrace{\frac{\partial \tilde{P}(\omega_j, \gamma_j)}{\partial \omega}}_{\text{Peer Pressure}} \tag{3.14}$$

The first term captures the change in effort due to selection in effort (which is also present in the

II scheme). The second term captures the pure free-riding effect, i.e., the change in the incentive strength due to an exogenous increase in the share of production a teacher is responsible for. The last term captures the change in effort due to changes in peer-pressure, where $\tilde{P}(\omega_j, \gamma_j) = (\alpha - \rho\omega_j)\gamma$. Since the peer pressure effect ameliorates the free-riding incentives, any significant difference between schemes must be due to the free-riding effect.

The key identifying assumption is that, conditional on the share ω_j, the marginal cost of effort, γ, is the same under the GI and II scheme. Intuitively, any difference between schemes is due to the teachers' responses to peer pressure and free-riding, and it is not due to sorting. Formally, the assumption is given by

Assumption 1 $\mathrm{E}[\gamma \mid II, \omega_j] = \mathrm{E}[\gamma \mid GI, \omega_j] \quad \forall \omega_j$

3.6 Balance Test

Even though the dataset is a randomized controlled trial, it was not designed to randomly assign teachers to students. Therefore, I provide a balancing test to check whether there is evidence of non-random sorting on observables. The following regression captures the relationship between the share ω_j for each incentive scheme and a set of explanatory variables.

$$\begin{aligned} z_{ij} = \pi_C + X_{ij}\beta + \pi_{II} I^{II} + \pi_{GI} I^{GI} + \\ \pi_W \omega_j + \pi_{IW} \times \omega_j \times I^{II} + \pi_{GW} \times \omega_j \times I^{GI} + \lambda_m + \epsilon_{ij}, \end{aligned} \tag{3.15}$$

where z_{ij} is a specific explanatory variable. I test the equality of π_2 and π_3. I include district dummy variables and clustered errors at the school level. The estimates of these two parameters are provided in Table 3.2. Columns (3) and (8) present the p-Value of the t-test for the equality of the II and GI interaction coefficient on each year respectively.

The results show limited evidence of differential sorting between treatments. Only two variables are significant. The household affluence index is significantly different between schemes at the 1%, and the teacher's year of service is correlated with GI share at the 10% level. As a robustness check, I run my main specification (given by equation (3.13)) using the interaction of this variables with the treatment indicators, and I do not find significant impact on the results. Thus, I conclude that the outcome cannot by attributed to non-random sorting based on these variables.

3.7 Results

Table 3.3 summarizes the results for all the regression specifications for year 1. Column 1 shows the results after controlling for student characteristics. Column 2 and 3 add school and teacher characteristics respectively[7]. Notice that half of the sample is lost after controlling for

[7]Student characteristics include previous year score, parent literacy index, household affluence index, gender, caste, grade, and subject. School characteristics include proximity index, infrastructure index and the number of

Table 3.2: Balance Test

	Control (1)	II (2)	GI (3)	p-Value (4)
Students:				
Baseline Score	-0.294*	0.138	0.0409	0.5838
Male	-0.0696	0.0373	-0.00994	0.3877
SC	0.197*	-0.138	0.00628	0.3205
ST	0.0967*	0.00512	-0.0900	0.3267
OBC	-0.220	0.0330	0.151	0.4762
FC	-0.0744	0.0996	-0.0673	0.1882
Hh Affluence	-0.310	0.536*	-0.364	0.0024
Parent Lit.	-0.144	-0.236	-0.00891	0.2230
Teachers:				
Male	0.315*	-0.166	-0.109	0.8273
Head Master	0.562***	-0.0275	0.188	0.3822
Regular Teacher	-0.401**	0.0159	-0.224	0.3638
Education Level	0.657**	-0.897**	-0.694*	0.5983
Training Qual.	0.354	-0.241	-0.401	0.6665
Salary	-1,351	3,689*	1,240	0.2287
Years of Service	-7.507**	12.37***	4.805	0.0893

Parental education ranges from 0 to 4; a point is added for each of the following: father's literacy, mother's literacy, father having completed tenth grade, and mother having completed tenth grade.
Household affluence index sums seven binary variables including ownership of land, ownership of current residence, residing in a pucca house (house with four walls and a cement and concrete roof), and having each of electricity, water, toilet, and a television at home.
Scheduled castes (SC) and *tribes* (ST) are considered the most socioeconomically backward groups in India. Other backward caste (OBC) are also included.
Teacher education is coded from 1 to 4 indicating tenth grade, twelfth grade, college degree, and master's or higher degree.
Teacher training qualifications is coded from 1 to 4 indicating no training, a diploma, a bachelor's degree in education, and a master's degree in education.

teacher characteristics.

Table 3.3 shows the effects of increasing ω_j for each treatment. Overall, there is no evidence of selection due to change in the cost of effort as $\hat{\pi}_{IW}$ is not significantly different from zero. Additionally, there is no evidence of free-riding effects as $\hat{\pi}_{GW}$ is negative and not significantly different from $\hat{\pi}_{IW}$. These results, combined with the positive and significant GI scheme intercept (i.e., $\pi_{GI} > 0$,), imply that teachers under the GI scheme would exert effort even when their share of students approaches zero. Thus, these results reject the pure free-riding hypothesis with no peer pressure.

Additionally, the results suggest that the peer-pressure effect must compensate for the free-riding incentives generated by a higher share, meaning that teachers with a higher share are subject to lower peer pressure. Accordingly, the evidence from Table 3.3 is consistent with the predictions from the heterogeneous peer-pressure model (i.e., $e_j^{he} = \alpha\gamma$). The results are consistent after controlling for observable characteristics, implying that the results are not driven by students', schools' and teachers' observable characteristics.[8]

teachers. Teacher characteristics include gender, designation (headmaster, regular teacher, contract or community teacher), education level and training qualifications.

[8] I do not find evidence of pure free-riding in the second year data. However, I do find an smaller π_{GI}, which

Table 3.3: Regression results

	District Fixed Effect			School Fixed Effect		
	(1)	(2)	(3)	(4)	(5)	(6)
II intercept ($\hat{\pi}_{II}$)	0.223*	0.278**	0.332***	-	-	-
	(0.118)	(0.122)	(0.125)	-	-	-
GI intercept ($\hat{\pi}_{GI}$)	0.283**	0.287**	0.321**	-	-	-
	(0.115)	(0.120)	(0.125)	-	-	-
II slope ($\hat{\pi}_{IW}$)	-0.203	-0.311	-0.506	-0.247	-0.302	-0.496
	(0.303)	(0.320)	(0.317)	(0.409)	(0.398)	(0.467)
GI slope ($\hat{\pi}_{GW}$)	-0.422	-0.399	-0.430	-0.206	-0.130	-0.309
	(0.269)	(0.276)	(0.283)	(0.328)	(0.319)	(0.338)
Control slope ($\hat{\pi}_{CW}$)	0.110	0.0891	0.130	-0.133	-0.0832	0.0222
	(0.178)	(0.214)	(0.212)	(0.248)	(0.238)	(0.278)
Student Characteristics	NO	YES	YES	NO	YES	YES
School Characteristics	NO	YES	YES	NO	YES	YES
Teacher Characteristics	NO	NO	YES	NO	NO	YES
Observations	42,145	37,617	33,768	42,145	37,617	33,768

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The sources of variation that I use to identify the parameters of interest are the following. First, I use the between-school variability. The school's average share decrease as the number of teachers increases. All else equal, bigger schools are going to have more free-riding incentives. Second, I use the within-school variability. Those teachers in the GI scheme with relatively lower shares are going to have more incentive to deviate from the II optimal effort. It is worth mentioning that each school has at most one teacher assigned to each grade and they teach both subjects. Therefore, all the within-school variability is coming from teachers that are being assigned to more classrooms (grades) than others. Thus, there is no marginal assignment of teachers to students. Thus, the non-random sorting opportunities are limited.

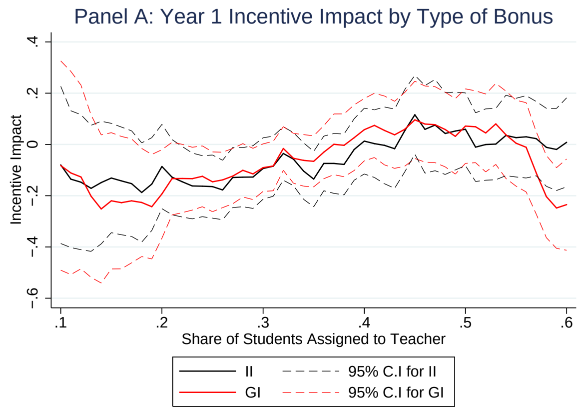
To give a better sense of the source of variation, the second panel of table 3.3 shows the effects of changing the shares after controlling for school fixed-effects. As seen, the results have similar magnitudes and are not qualitatively different from the results from the first panel. Accordingly, the lack of free-riding effects are not driven by unobserved school characteristics that may bias the results.

3.7.1 Heterogeneous Effects

The existence of non-linearities in peer-pressure could hide free-riding effects for low levels of teacher shares. Following Imberman and Lovenheim (2015), I estimate a local-linear regression to recover the heterogeneous impact of the incentives at different share levels. Panel A in Figure 3.1 presents the year 1 estimates for $\omega_j \times I^{II}$ and $\omega_j \times I^{GI}$ with 95% clustered confidence

is consistent with the results from Muralidharan and Sundararaman (2011) showing an smaller ATE for the GI scheme in the second year of the program.

Figure 3.1: Local linear regression



intervals centered at each share value, with a uniform kernel bandwidth of 0.15. I show the estimates for values of ω_j from 10% to 60%, because that range covers 90% of the sample. Below and above those values, the confidence intervals become very large. Consistent with the previous results, the estimates are not significantly different from each other. Therefore, I find no evidence of heterogeneous free-riding effects.

3.8 Structural estimation and counter-factual analysis

In the previous section, I used the original contract information to find evidence of heterogeneous peer pressure in the GI scheme. Still, nothing has been said about the optimality of its design. In this section, I estimate the model parameters to counterfactually recover the optimal bonus for the average school in Andra Pradesh. By doing so, I can compare my results with Muralidharan and Sundararaman (2011) and Mehta (2017) and show the relative performance of the optimally designed GI bonus scheme under the assumption of heterogeneous peer pressure.

The empirical strategy is based on estimating the average treatment effect (ATE) for each scheme expressing the outcome in monetary units. To do so, I recover the present value of the students' scores using a similar approach to Muralidharan and Sundararaman (2011). Since the contract slope is known, I can estimate the following econometric model:

$$q_{ij}^{\$} = \pi_C + \beta q_{ij(t=0)} + \pi_{II} I^{\text{slope}\times II} + \pi_{GI} I^{\text{slope}\times GI} + \lambda_m + \epsilon_{ijcs} \tag{3.16}$$

where the \$ supra index reflects variables expressed in monetary units. I now describe how to recover $q_{ij}^{\$}$. First, Aslam et al. (2011) estimate that a one standard deviation score increase on a standardized math (language) test produces a return of 16 (20) percent. So, multiplying these returns by the standardized score results ($q_{ij} \times 16\%$ and $q_{ij} \times 20\%$) provides the return for each student-teacher observation. Second, I follow the same methodology that Muralidharan and

Table 3.4: Estimated model parameters

Parameter	Estimate	Confidence I.
Teacher effort (SD):		
$\hat{e}_j^{II}$	.156	[.136, .175]
$\hat{e}_j^{GI}$	.141	[.121, .160]
Cost Parameter:		
$\hat{\gamma}$	4.83	[4.3, 5.52]
Peer Pressure (%):		
$\hat{\alpha}$	90.1	[77.8, 102]
Variance (U$s):		
$\hat{\sigma}^{\$^2}$	84676.7	[83541 , 85827]

Sundararaman (2011) use to recover the present value[9]. I then multiply the present value by the return of each student exam to get $q_{ij}^{\$}$.

To make the results comparable to Mehta (2017) calibration, I only include students' previous year scores, $q_{ij(y_0)}$ as an student characteristic. The indicators $I^{\text{slope}\times II}$ and $I^{\text{slope}\times GI}$ are indicator variables multiplied by the known contract slope when the observation is treated.

The identification strategy works as follows. First, teachers' efforts, $e_j^{II} = \beta_1\gamma$ and $e_j^{he} = \alpha\gamma$, are estimated using $\hat{\pi}_{II}$ and $\hat{\pi}_{GI}$ respectively. Therefore, they are the ATE for each bonus scheme. Second, I take advantage of knowing the slope to identify the marginal cost parameter, γ, from the ATE of the II scheme, $\hat{\pi}_{II}^{\$}$. Third, I identify the peer pressure parameter α as $\hat{\alpha} = \frac{\hat{\pi}_{GI}}{\hat{\pi}_{II}}$. Finally, I recover the student idiosyncratic error variance $\hat{\sigma}^{\$^2}$ from the residuals of the regression.

The results are summarized in Table 3.4. As seen, teacher effort under the II scheme increase students' performance by .156 SD, while teachers under the GI scheme increase it by .141 SD, which are equivalent to the ATEs from Muralidharan and Sundararaman (2011). The estimated peer pressure parameter is $\alpha = .901$, which implies that teachers under the GI scheme exert 90% of the effort under the II scheme.

3.8.1 Optimal Contract

In this section, I recover the optimal linear piece rate contract for an average school. The designer's problem is to maximize the performance of a representative classroom, minus the cost (the teachers' payment) subject to the teachers' willingness to participate (IRC) and their optimal effort choice (ICC).

$$
\begin{aligned}
&\max_{\beta_0,\beta_1} && E_\epsilon[\bar{q}_j - W_j] && \\
&\text{s.t} && E_\epsilon[-exp\{-\xi(W_j^k - C_j - \mathbb{I}^{GI}P_j)\}] \geq U(\bar{W}) && \text{(IRC)}\\
&\text{and} && \arg\max_e E_\epsilon[-exp\{-\xi(W_j^k - C_j - \mathbb{I}^{GI}P_j)\}], && \text{(ICC)}
\end{aligned}
$$

[9]The average annual wage for agricultural labour is calculated to be Rs. 28,000. The present value of 40 years of work, assuming a discount rate of 10%, constant wages and a 12 year delay between year 1 and the first year of work is Rs.87,662

Table 3.5: Counterfactual Results

Parameter	Estimate (1)	Confidence I. (2)
Optimal Contract Slope:		
$\hat{\beta}_1^{II}$	.242	[.218, .265]
$\hat{\beta}_1^{GI}$	.489	[.457, .521]
Optimal effort (SD):		
$\hat{a}_t^{II}$	1.168	[.906, 1.43]
$\hat{a}_t^{GI}$	2.362	[1.90, 2.81]

Following Mehta (2017), I evaluate this contract for a representative school. In particular, I focus on an average school, which has 37.5 students and three teachers. I also assume that students are evenly distributed between teachers. Hence, $\omega_t = \frac{1}{J} = \frac{1}{3}$. I set $\xi = 6.7 \times 10^3$ as Mehta (2017) using the estimates from Cohen and Einav (2007).

Solving the maximization problem for each case, the optimal effective slope β_1 is given by

$$\begin{aligned} \beta_1^{II} &= \frac{1}{1 + \xi\gamma^{-1}\sigma_J^2} \\ \beta_1^{GI} &= \frac{1}{\alpha + \xi\gamma^{-1}\frac{\sigma_J^2}{J\alpha}}. \end{aligned} \tag{3.17}$$

Equation (3.17) suggests that if $\alpha = 1$, the optimal slope under the GI scheme would be higher due to the lower risk of an extreme shock, which is capture by the lower variability given by $\frac{\sigma_J^2}{J\alpha}$ instead of σ_J^2. When $\alpha = \frac{1}{3}$, the GI slope would be equivalent to the optimal slope under pure free-riding (i.e., with pee-pressure parameters given by $\rho = 0$ and $\alpha = \frac{1}{3}$).

Table 3.5 presents the results of estimating the optimal slopes and the predicted effort under each contract. As column (1) shows, the optimal II slope induces teachers to increase the expected student's performance by 1.16 SD points, which is 7.5 times larger than the result from Muralidharan and Sundararaman (2011) and almost equal to Mehta (2017). Meanwhile, the optimal GI contract with heterogeneous peer pressure increases teacher effort by 2.36 SD points, which is 14 times larger than the original contract and almost two times larger than the optimal II scheme. Given that the average return from a standardized test in India is 18 percent (Aslam et al. (2011)), the optimal II would increase students' wealth by 21 percent. In comparison, the optimal GI contract with heterogeneous peer effect would increase it by 42 percent. The original bonus increased students' wealth by only 3 percent.

The main intuition behind these results is that a trade-off exists between higher incentives, through β_1, and smaller risk. Steeper contracts generate higher effort. However, if teachers are risk-averse, steeper contracts do not necessarily generate higher utility, and the IRC binds. Thus, there is a limit to how steep the slope can be to meet the IRC. The GI scheme allows teachers to diversify risk as the school-level variance of the shock is a fraction of the teacher-level variance. At $\alpha = .901$, for any given slope, teachers under the GI scheme can produce almost as in the II scheme but with lower risk. Thus, the designer can offer a steeper scheme to teachers under the GI scheme.

To check the contract optimality under different peer pressure scenarios, I provide in Figure 3.2 the optimal bonus for the full range of possible peer-pressure parameters, from $\alpha = \frac{1}{3}$ (which is equivalent to the pure free-riding case) to $\alpha = 1$. Panel A shows that the optimal slope for GI is higher than the II slope, and it increases as α increases. The intuition behind this is that GI ensures that teachers can withstand their idiosyncratic classroom shocks by diversifying the risk with other teachers. Thus, the designer can provide more incentives through a steeper contract slope. In that sense, the optimal contract assuming pure free riding (without peer pressure) shows that teachers effort would increase by 0.46 SD points (Panel B), which is three times larger than the original result. Student wealth would increase by 8 percent, which doubles the original incentive scheme result but is less than half of the II optimal incentive scheme. Panel B also shows that teacher effort under GI steadily increases and becomes equal to II when α is 57 percent.

Panel C illustrates the optimal contract intercept which is the fixed compensation level. The intercept is increasing at low levels of α but then decreases.[10] This is due to two opposing effects. First, risk diversification between teachers allows the designer to provide less insurance (smaller β_0). Second, the incentive strength is low for low levels of α, which could lead to teachers placing more value on being more insured than having a steeper contract.

My empirical methodology is tempered by the fact that ξ represents the mean estimated absolute risk aversion parameter from Cohen and Einav (2007) who use Israeli data on deductible choices in auto insurance contracts. Thus, I check the robustness of my results by increasing the ξ parameter by three times the calibrated value. If $\xi = 20 \times 10^3$, the optimal II effort increases student performance by 0.47 SD points, while the optimal GI with peer pressure will increase it by 1.17 SD points. Therefore, even if teachers are considerably risk-averse, the results are still significant and the relative performance of the GI scheme holds.

3.9 Conclusion

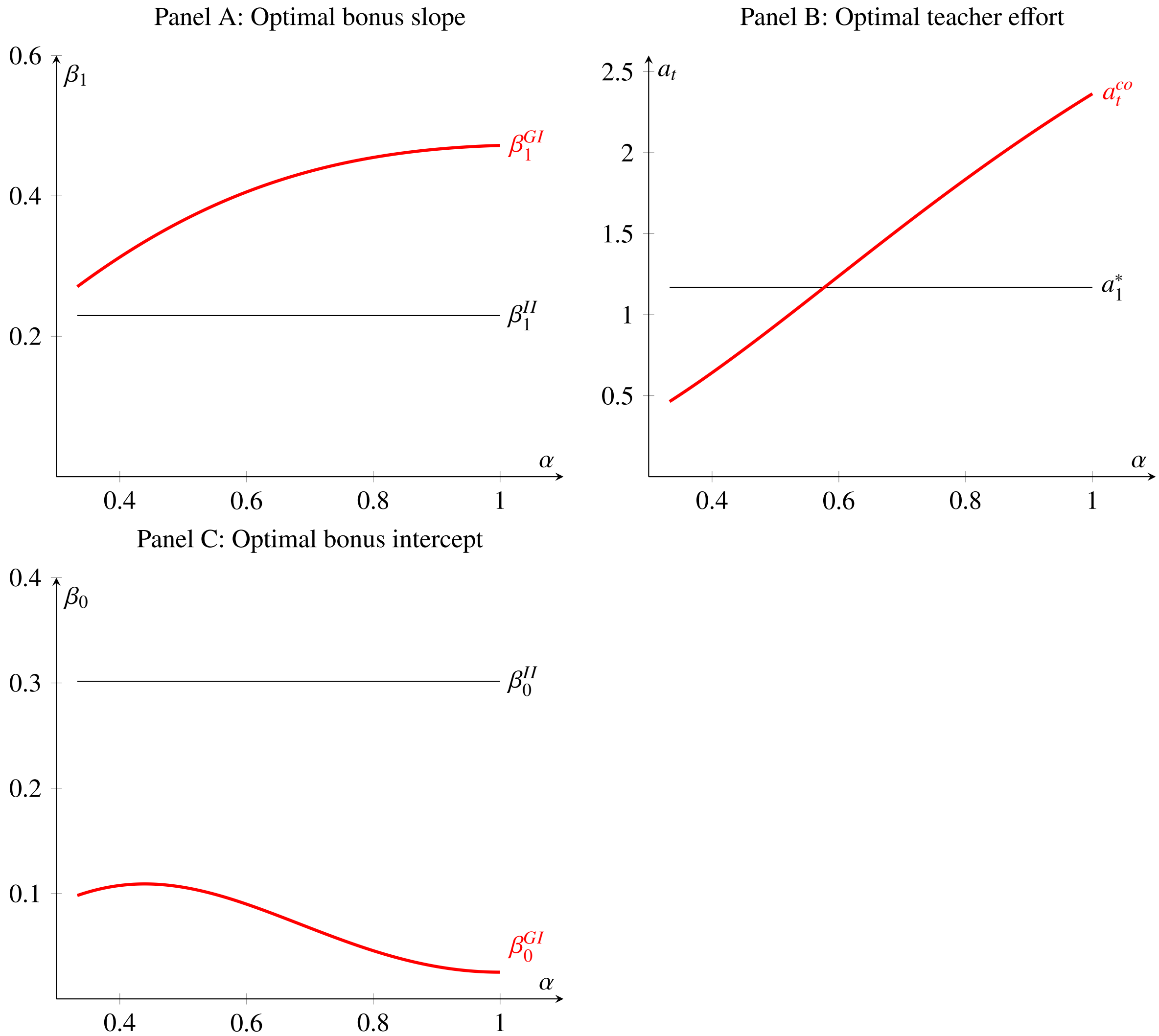
In this paper, I exploit the experimental nature of the data from Muralidharan and Sundararaman (2011) to test for the presence of free-riding and peer-pressure in the GI scheme. My results show that teachers with a small share of students are subject to higher peer pressure which mitigates the free-riding incentives. I consider the heterogeneous effects of peer pressure to model the optimal GI bonus scheme. I find that, contrary to the original design, the optimally-designed GI scheme would have different parameters and produce higher test scores than the II scheme.

An optimally designed piece rate bonus can increase student wealth by 8 to 42 percent. If teachers are subject to high levels of peer pressure, the optimally designed group-based contract can double the expected performance of students compared to the optimally designed individual contract.

The unique feature of this paper is that it exploits the experimental nature of the data to

[10] I recover β_0 using the incentive rationality constraint, normalizing the outside option to be the average baseline salary. The average lagged performance score is normalized to zero in the baseline year.

Figure 3.2: Simulation for different α



recover structural parameters and test for the presence of free-riding in the GI scheme. Still, the analysis is limited to a statistical model that does not formally treat the dynamic results. Having access to the larger panel from Muralidharan (2012) with more than three years of data could substantially enrich the analysis.

Bibliography

Angrist, Joshua D and Jörn-Steffen Pischke. 2010. The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *The Journal of economic perspectives* 24, no. 2:3–30.

Aslam, Monazza, Anuradha De, Geeta Kingdon, and Rajeev Kumar. 2011. Economic returns to schooling and skills–an analysis of india and pakistan. *Education Outcomes and Poverty in the South* .

Barrera-Osorio, Felipe and Dhushyanth Raju. 2017. Teacher performance pay: Experimental evidence from pakistan. *Journal of Public Economics* 148:75–91.

Behrman, Jere R, Susan W Parker, Petra E Todd, and Kenneth I Wolpin. 2015. Aligning learning incentives of students and teachers: results from a social experiment in mexican high schools. *Journal of Political Economy* 123, no. 2:325–364.

Bellemare, Charles, Steeve Marchand, and Bruce Shearer. 2016. Structural estimation and experiments: Applications to contracting models. *Journal of Institutional and Theoretical Economics JITE* 172, no. 2:342–363.

Bolton, Patrick and Mathias Dewatripont. 2005. *Contract theory*. MIT press.

Chetty, Raj, John N Friedman, and Jonah E Rockoff. 2014. Measuring the impacts of teachers ii: Teacher value-added and student outcomes in adulthood. *American economic review* 104, no. 9:2633–79.

Cohen, Alma and Liran Einav. 2007. Estimating risk preferences from deductible choice. *The American economic review* 97, no. 3:745–788.

Duflo, Esther, Rema Hanna, and Stephen P Rya. 2012. Incentives work: Getting teachers to come to school. *The American Economic Review* 102, no. 4:1241–1278.

Glewwe, Paul, Nauman Ilias, and Michael Kremer. 2010. Teacher incentives. *American Economic Journal: Applied Economics* 2, no. 3:205–227.

Hanushek, Eric A. 2011. The economic value of higher teacher quality. *Economics of Education review* 30, no. 3:466–479.

Imberman, Scott A and Michael F Lovenheim. 2015. Incentive strength and teacher productivity: Evidence from a group-based teacher incentive pay system. *Review of Economics and Statistics* 97, no. 2:364–386.

Jackson, C Kirabo, Jonah E Rockoff, and Douglas O Staiger. 2014. Teacher effects and teacher-related policies. *Annu. Rev. Econ.* 6, no. 1:801–825.

Kandel, Eugene and Edward P Lazear. 1992. Peer pressure and partnerships. *Journal of political Economy* 100, no. 4:801–817.

Keane, Michael P. 2010. Structural vs. atheoretic approaches to econometrics. *Journal of Econometrics* 156, no. 1:3–20.

Koedel, Cory, Kata Mihaly, and Jonah E Rockoff. 2015. Value-added modeling: A review. *Economics of Education Review* 47:180–195.

Loyalka, Prashant Kumar, Sean Sylvia, Chengfang Liu, James Chu, and Yaojiang Shi. 2016. Pay by design: Teacher performance pay design and the distribution of student achievement .

Macartney, Hugh, Robert McMillan, and Uros Petronijevic. 2015. Incentive design in education: An empirical analysis. Tech. rep., National Bureau of Economic Research.

Mehta, Nirav. 2017. The potential output gains from using optimal teacher incentives: An illustrative calibration of a hidden action model. *University of Western Ontario* 2017.

Muralidharan, Karthik. 2012. Long-term effects of teacher performance pay: Experimental evidence from india. *Society for Research on Educational Effectiveness* .

Muralidharan, Karthik and Venkatesh Sundararaman. 2011. Teacher performance pay: Experimental evidence from india. *Journal of political Economy* 119, no. 1:39–77.

Murnane, Richard J, Alejandro J Ganimian et al. 2014. *Improving educational outcomes in developing countries: Lessons from rigorous evaluations*. National Bureau of Economic Research Cambridge, MA.

Chapter 4

The Heterogeneous Effect of Teachers on Minority Students

Evidence shows that African American and Hispanic students in the United States consistently perform worse than White and Asian students on standardized tests. The achievement gaps start as early as kindergarten and persist through college (Dee 2005; Redding 2019; Fairlie et al. 2014). These gaps represent a major educational problem since Blacks and Hispanics constitute the largest share of students in K-12 public schools in many US districts (NCES 2021). Given the direct relationship between academic achievement and students' future outcomes, designing policy levers to increase low-achieving minorities' performance has become a priority for policymakers.

A natural way to reduce the achievement gaps is by increasing minorities' access to effective teachers. It is well documented that teachers play a major role in determining students' academic success (Hanushek and Rivkin 2006; Jackson et al. 2014; Koedel et al. 2015). Thus, identifying the best teachers for minority students is crucial for the design of an effective policy. However, the standard value-added models (VAMs) used to identify effective teachers do not consider the differential effect that some teachers may have on different types of students, such as minority students. In fact, it has been proven difficult to identify effective teachers for minority students using standard VAMs (Stecher et al. 2018).

In this paper, I study the degree to which teachers can affect the academic performance of Black and Hispanic students relative to White and Asian students. To do so, I estimate a flexible value-added model that allows each teacher to have a differential effect on minority students. These matching effects capture the teachers' ability to reduce the achievement gap between their assigned students. Then, I study the relationship between the estimated matching effects and a set of teachers' characteristics and skills. This allows me to explore what type of teachers are better suited to teach minority students. The characteristics include teachers' race and gender. Meanwhile, the measures of skills are the teachers' experience, preparedness to teach their subject, and a rich set of evaluations of what teachers do in a classroom.

I estimate the econometric model using The Measure of Effective Teaching database, which contains rich information on teachers and students from six large US districts: Charlotte-

Mecklenburg, Dallas, Denver, Hillsborough County, Memphis, and New York City. In particular, the database contains administrative records of students that expand for six years (2008-2009 to 2013-2014). These records match students to their assigned teachers using unique identifiers.

During the 2009-2010 and 2010-2011 academic periods, a subset of teachers was videotaped and evaluated by trained raters using multiple protocols of effective teaching. In particular, I use the Framework for Teaching (FFT) and the Classroom Assessment Scoring System (CLASS) protocol, both containing multiple measures of different dimensions of teaching. I group each protocol into two dimensions that relate to the teachers' quality of instruction and their management of the classroom behavior. Additionally, I use the Content Knowledge for Teaching Math Assessment (CKT) to measure teachers' understanding of the subject being taught.

I find that, on average, the minority gap is .08 standard deviations after controlling for students' observable characteristics and teacher-year fixed effects. However, this average masks important heterogeneity between teachers because teacher's effectiveness may vary from student to student. In particular, a one standard deviation increase in the matching effects translates into a relative achievement gain of .056 standard deviations, which implies a reduction of 70% of the conditional achievement gap. This result highlights the important role that teachers may play in reducing the prevalent disparities between minority students and high-performing students.

I do not find a significant relationship between the matching effect and the teachers' characteristics, experience, content knowledge, and instructional quality. However, I find a positive relationship with classroom management as measured by both protocols, implying that minority students benefit more from teachers who can keep students on task and well-behaved. This result suggests that how teachers teach matters in improving minority students' achievement.

There is a recent but growing body of literature studying the differential effect that teachers may have on students using flexible value-added specifications (Fox 2016; Loeb et al. 2014; Ahn et al. 2021). Using data from North Carolina, Ahn et al. (2021) find matching effects that are remarkably similar in magnitude to the ones I found. The combined evidence suggests that the magnitude of these effects may be similar across several US districts. Additionally, I relate the estimated matching effects to teachers' characteristics and multiple measures of teachers' skills.

4.1 Sample Restrictions

I restrict my analysis to elementary and middle school math teachers with complete information on the FFT and CLASS protocols, CKT scores, and at least five minority and non-minority students assigned. Additionally, I restrict the sample to Black and Hispanic students, traditionally categorized as minority students, and White and Asian students with complete information on lagged math and language scores and race. The final sample contains 110,731 students and 507 teachers and is summarized in Table 4.1.

As seen in the first panel of Table 4.1, minority students represent 59% of the sample. The sample is evenly distributed between male and female students. The average age of the students is 12 years, which implies that the average student is in 6^{th} grade. The second panel of Table 4.1 presents descriptive statistics of the teachers. As opposed to students, White teachers outnumber minority teachers, and a similar pattern occurs for female teachers.

Table 4.1: Descriptive Statistics

	Mean	Std. Dev.
Student-Year Variables:		
Race: Black	.279	.449
Race: Hispanic	.309	.462
Race: White	.345	.475
Race: Asian	.067	.249
Gender: Male	.501	.500
Age	12.00	1.42
Free-Reduced Price Lunch	.618	.486
Gifted Status	.097	.296
Special Education	.092	.290
English Language Learner	.116	.320
Elementary Level	.225	.417
Teacher Variables:		
Race: White	.633	.482
Race: Black	.288	.453
Race: Hispanic	.061	.236
Race: Other	.002	.139
Gender: Male	.191	.394
District Experience	7.43	6.81
Content Knowledge (CKT)	57.45	14.70
FFT: Management	2.65	.289
FFT: Instruction	2.41	.267
CLASS: Management	4.32	.247
CLASS: Instruction	3.87	.522
Elementary Level	.495	.50
Number of Teachers	507	
Number of Students	110,731	

Notes: Descriptive Statistics for 4^{th} to 8^{th} grade math teachers in the academic years 2008-2009 to 2013-2014. I only observe the district experience for 377 teachers.

Table 4.2 shows the minority achievement gap from an OLS regression with teacher-year fixed effects and clustered standard errors at the same level. I standardized students' achievement within a year, grade and district. The first column shows an unconditional minority gap of .50 standard deviation. As seen in the second column, the gap is reduced but persist after controlling for students' lagged score and other students' observable characteristics. The only other indicators with large gaps are the gifted and special education status. However, these students represent a relatively small share of the students in the sample.

Table 4.2: Student Racial Gaps in Math

	(1)	(2)
Minority	-.50***	-.08***
	(.012)	(.004)
Lagged Math Score	-	.630***
		(.003)
Lagged Language Score	-	.149***
		(.003)
Gender: Male	-	.013***
		(.004)
Free-Reduced Price Lunch	-	-.041***
		(.007)
Gifted Status	-	.154***
		(.009)
Special Education Status	-	-.0084***
		(.007)

Notes: All regressions include Teacher-Year fixed effect. The baseline category corresponds to White and Asian students. Standard errors are clustered at the Teacher-Year Level. The first column represents the unconditional minority gap. The second column controls for lagged math and language scores, gender, age, free-reduced price lunch status, gifted status, and special education status. * Significant at 10 percent. ** Significant at 5 percent. *** Significant at 1 percent.

4.2 The Model

My econometric specification is given by

$$y_{ijt} = X'_{ijt}\beta + \gamma_j \times Minority_i + \theta_{jt} + \epsilon_{ijt} \tag{4.1}$$

where y_{ijt} represents student i's math achievement under teacher j in year t. The vector X_{ijt} corresponds to observable student variables, while ϵ_{ijt} captures unobserved components.[1] The model also includes the level effect, θ_{jt}, that is common to all students assigned to teacher j in year t. The object of interest is the distribution of the matching effects given by γ_j, which captures the minority gap for students assigned to teacher j. A higher matching effect implies that minority students assigned to teacher j perform relatively better than the average minority student in the sample.

The teacher-year level effect controls for the possibility that minority students are system-

[1] The vector X_{ijt} includes students' lagged scores in math and language that are allowed to vary by education level and race, age, gender, gifted status, English language learner status, and missing indicators.

atically sorted into better (or worse) teachers in a particular year. In this regard, the interaction parameters γ_j could be reinterpreted as a difference-in-difference coefficient, where any general unobserved effect for teacher j in year t is removed. Accordingly, the key identification assumption is that, within teacher-year, the teacher-specific matching effect is uncorrelated with the unobserved shock, ϵ_{ijt}, conditional on the students' observable characteristics.

4.3 Estimation

4.3.1 Distribution of the Matching Effects

In principle, equation (4.1) could be estimated in one step using fixed effects to capture the level effects, θ_{jt}, and the matching effects, γ_j. However, such estimates would be contaminated with estimation error, which would upward bias the variance of the effects. Alternatively, one could shrink the estimated effects using standard post-estimation shrinking techniques. However, these estimates would underestimate the true variation of the effects. Accordingly, the true variability of the matching effect would lie between the shrunken and fixed effect (i.e., unshrunken) estimates (Raudenbush and Bryk 2002; Kraft 2017; Blazar 2018). Therefore, I use a two-step procedure to recover an unbiased estimate of the teachers' effects.

The two-step procedure works as follows. In the first step, I recover the vector of parameters β from the econometric model by estimating the equation (4.1) using fixed effects. Doing so allows students' characteristics to be flexibly correlated with the matching effects, γ_j, and teacher-year effects, θ_{jt}. Then, I calculate the residual between the students' scores and the part of the model explained by students' characteristics, $y_{ijt} - X'_{ijt}\beta$.

In the second step, I estimate the distribution of the teacher effects using restricted maximum likelihood (RML) using the estimated residual from the first stage as the dependent variable. The RML procedure recovers unbiased estimates of the true distribution of effects.[2] I estimate a two-level hierarchical structure for the level effect, which is defined at the teacher-year and teacher level: $\theta_{jt} = \bar{\theta}_j + \tilde{\theta}_{jt}$. I assume a joint-normal distribution that allows the matching effects, γ_j, to be correlated with the level effects defined at the teacher level, $\bar{\theta}_j$.

4.3.2 Relationship Between the Matching Effects and Teachers' Characteristics

tudying the relationship between teachers' characteristics and their corresponding matching effects requires having estimates of each teacher's effect. To do so, I recover the best linear unbiased estimators of the random effects, which are equivalent to the empirical Bayes estimates. Then, I regress the estimated effects on a set of teachers' characteristics: race, gender, teaching experience in the district, content knowledge, instructional quality, and classroom management. I standardize all variables to ease the interpretation of the results.

[2] See Blazar and Kraft (2017), Chetty et al. (2011) and Kraft (2017) for examples of value-added models using RML

Since the FFT and CLASS protocols are based on classroom observations, they may capture students' components that are not directly related to teachers' skills. Therefore, I estimate the regression

$$Protocol_{cjt} = \bar{X}'_{cjt}\beta + S_j + \xi_{cjt}, \tag{4.2}$$

where $Protocol_{jt}$ represents the classroom management or instructional quality of teacher j in classroom c in year t (i.e., year 2009-2010 and 2010-2011), as measured by the FFT and CLASS protocols. The vector $\bar{X}_{cjt}$ corresponds to observable characteristics of the students in the classroom, including lagged scores in math and language, average age, and the proportion of male, minority, gifted, and English language learner students. These variables capture changes in the protocols' scores that are due to changes in the composition of students in a classroom. The objects of interest are the teacher effects, S_j, which I estimate using fixed effects. I use the fixed effect estimates as measures of the teachers' managerial and instructional skills.

4.4 Results and Discussion

Table 4.3 presents the estimated variance-covariance matrix of teachers' effects. The first row shows the standard deviation of the matching effect. Increasing the matching effect by one standard deviation increases minorities' achievement by .056 standard deviations. The second row shows that the standard deviation of the level effect is .187 standard deviation, which is consistent with results from the standard value-added models (Hanushek and Rivkin 2006; Jackson et al. 2014; Koedel et al. 2015). The third row shows that the correlation between these effects is negative, although non-significant. Therefore, the overall performance of students assigned to teacher j is not directly related to the differential effect that some teachers may have on minority students. Finally, the last two rows show the standard deviation of the residual variation.

Table 4.3: Estimated Teacher Effects

	Math		
	Estimate	SE	95% Confidence Interval
Teacher-Level Effects (SD)			
Matching Effect	.056	.006	(.045,.070)
Level Effect	.187	.008	(.172,.204)
Corr(Matching, Level)	-.07	.1067	(-.277,.135)
Teacher-Year Effects (SD)			
Level Effect	.165	.004	(.159,.173)
Student-Year Effects (SD)			
Level Effect	.517	.001	(.515,.519)

Different mechanisms could explain the matching effects. The matching effects could be due to students being more motivated when matched to a demographically similar teacher, resulting in increased performance. Under this scenario, teachers would play a passive role and would not directly influence academic outcomes with their teaching skills. Accordingly, policymakers could increase the participation of minority teachers to promote the performance of minority students.

Alternatively, some teachers may have teaching skills that are more beneficial for minority students. Thus, policymakers could design training programs to increase the teaching skills that benefit minority students the most.

To explore the relevancy of these potential mechanisms, in Table 4.4, I present the relationship between the estimated matching effects and the set of teachers' characteristics and skills. Each row in the first column represents a separate regression. As seen in the first row of Column 1, the relationship between teachers' race and the matching effect is positive but not significant. The results show a similar pattern for teachers' gender, experience, content knowledge, and instructional quality. However, as measured by both protocols, I find a positive and significant relationship between teachers' managerial skill and the matching effects. This evidence suggests that teachers may play an active role in determining the minorities' performance through better classroom management.

In the second and third columns of Table 4.4, I combine the teachers' characteristics and skills in a single regression. These estimates remove any effect that the teaching skills may have on the estimated effect of teachers' race. The second column presents the results using the FFT measures of teachers' managerial and instructional skills, while the third column uses the CLASS protocol instead. I do not include the teachers' experience, as it is only observed for a subset of teachers. The relationship between the teachers' skills and characteristics is similar to the previous estimates. Overall, this evidence suggests that improving the managerial skills of teachers may be an effective mechanism to increase teachers' contribution to the academic performance of minority students.

Table 4.4: Relationship between Matching Effects and Teachers' Characteristics and Skills

	(1)	(2)	(3)
Teacher: Minority	.109	.122	.129
	(.093)	(.096)	(.097)
Teacher: Male	.024	.036	.057
	(.113)	(.114)	(.115)
District Experience	.051	-	-
	(.050)		
Content Knowledge	.023	.060	.050
	(.044)	(.047)	(.046)
FFT: Management	.085*	.167**	-
	(.044)	(.077)	
FFT: Instruction	.037	-.087	-
	(.044)	(.076)	
CLASS: Management	.097**	-	.126***
	(.044)		(.047)
CLASS: Instruction	-.028	-	-.056
	(.044)		(.046)

Each row in the first column represent a separate regression. The sample consists of 507 math teachers. * Significant at 10 percent. ** Significant at 5 percent. *** Significant at 1 percent.

Bibliography

Ahn, Tom, Esteban Aucejo, Jonathan James et al. 2021. The importance of matching effects for labor productivity: Evidence from teacher-student interactions. Tech. rep.

Blazar, David. 2018. Validating teacher effects on students' attitudes and behaviors: Evidence from random assignment of teachers to students. *Education Finance and Policy* 13, no. 3:281–309.

Blazar, David and Matthew A Kraft. 2017. Teacher and teaching effects on students' attitudes and behaviors. *Educational evaluation and policy analysis* 39, no. 1:146–170.

Chetty, Raj, John N Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2011. How does your kindergarten classroom affect your earnings? evidence from project star. *The Quarterly journal of economics* 126, no. 4:1593–1660.

Dee, Thomas S. 2005. A teacher like me: Does race, ethnicity, or gender matter? *American Economic Review* 95, no. 2:158–165.

Fairlie, Robert W, Florian Hoffmann, and Philip Oreopoulos. 2014. A community college instructor like me: Race and ethnicity interactions in the classroom. *American Economic Review* 104, no. 8:2567–91.

Fox, Lindsay. 2016. Playing to teachers' strengths: Using multiple measures of teacher effectiveness to improve teacher assignments. *Education Finance and Policy* 11, no. 1:70–96.

Hanushek, Eric A and Steven G Rivkin. 2006. Teacher quality. *Handbook of the Economics of Education* 2:1051–1078.

Jackson, C Kirabo, Jonah E Rockoff, and Douglas O Staiger. 2014. Teacher effects and teacher-related policies. *Annu. Rev. Econ.* 6, no. 1:801–825.

Koedel, Cory, Kata Mihaly, and Jonah E Rockoff. 2015. Value-added modeling: A review. *Economics of Education Review* 47:180–195.

Kraft, Matthew A. 2017. Teacher effects on complex cognitive skills and social-emotional competencies. *Journal of Human Resources* :0916–8265R3.

Loeb, Susanna, James Soland, and Lindsay Fox. 2014. Is a good teacher a good teacher for all? comparing value-added of teachers with their english learners and non-english learners. *Educational Evaluation and Policy Analysis* 36, no. 4:457–475.

NCES. 2021. Share of students enrolled in k-12 public schools in the united states in 2020, by ethnicity and state. *Accessed: July 22, 2022. https://www.statista.com/statistics/236244/enrollment-in-public-schools-by-ethnicity-and-us-state/* .

Raudenbush, Stephen W and Anthony S Bryk. 2002. *Hierarchical linear models: Applications and data analysis methods*, vol. 1. sage.

Redding, Christopher. 2019. A teacher like me: A review of the effect of student–teacher racial/ethnic matching on teacher perceptions of students and student academic and behavioral outcomes. *Review of educational research* 89, no. 4:499–535.

Stecher, Brian M., Deborah J. Holtzman, Michael S. Garet, and et. al. 2018. *Improving teaching effectiveness: Final report: The intensive partnerships for effective teaching through 2015 - 2016.* Santa Monica, CA: RAND Corporation.

Appendix A

Appendices to Chapter 2

Table A.1: Measurement System

	Value	SE		Value	SE
Measures of $\ln Y_{icj}$			Measures of K_{icj}		
State Test			Previous Year Math		
Location	0.00	-	Location	0.00	-
Scale	1.00	-	Scale	1.00	-
Standard Deviation	.469	.005	Standard Deviation	.42	.006
BAM			Previous Year Science		
Location	.048	.011	Location	-.13	.017
Scale	.89	.013	Scale	.857	.019
Standard Deviation	.623	.007	Standard Deviation	.586	.011
Measures of $\ln L_{icj}$			Measures of $\ln H_{icj}$		
log Hours/week of study			Number of Books		
Location	0.00	-	Location	0.00	-
Scale	1.00	-	Scale	1.00	-
Standard Deviation	.855	.012	Standard Deviation	.824	.014
Homework Completed:			Number of Computers		
Scale	1.10	.079	Location	0.00	.018
Threshold 1	-1.32	.062	Scale	.875	.037
Threshold 2	-.049	.051	Standard Deviation	.868	.013
Threshold 3	.314	.05	Parents Education		
Threshold 4	2.13	.056	Scale	1.21	.057
Stop Trying:			Threshold 1	-.975	.033
Scale	.979	.07	Threshold 2	-.221	.030
Threshold 1	-1.24	.05	Threshold 3	.210	.029
Threshold 2	-.792	.045	Threshold 4	.811	.030
Threshold 3	-.155	.044	Measures of $\ln E_{Mcj}$		
Threshold 4	.502	.046	Doesn't let give up:		
Measures of $\ln M_{cj}$			Scale	1.00	-
Students Treat with Respect:			Threshold 1	-1.154	.035
Scale	1.00	-	Threshold 2	-.764	.025
Threshold 1	-1.582	.035	Threshold 3	0.00	-
Threshold 2	-.975	.025	Threshold 4	.742	.019
Threshold 3	0.00	-	Better Thinker:		
Threshold 4	1.023	.023	Scale	1.08	.049

Table A.1: Measurement System

	Value	SE		Value	SE
Behavior under control:			Threshold 1	-.868	.050
Scale	.804	.027	Threshold 2	-.455	.047
Threshold 1	-1.128	.032	Threshold 3	.328	.046
Threshold 2	-.579	.029	Threshold 4	1.063	.047
Threshold 3	.241	.029	Accepts Nothing Less:		
Threshold 4	1.08	.031	Scale	.961	.048
Classmates behave:			Threshold 1	-1.179	.052
Scale	.922	.030	Threshold 2	-.766	.047
Threshold 1	-1.036	.035	Threshold 3	.091	.044
Threshold 2	-.430	.032	Threshold 4	.845	.044
Threshold 3	.525	.032	Measures of $\ln \mu_{Mcj}$		
Threshold 4	1.452	.034	log Classroom Size		
			Location	3.00	0.001
			Scale	-.160	.014
			Standard Deviation	.253	.007

Table A.2: Descriptive Statistics: Measures of Latent Variables

	Mean	Obs.
Measures of ln Y_{icj}		
State Test	0	6,603
BAM	0	4,764
Measures of ln K_{icj}		
Previous Year Math	0	6,214
Previous Year Science	0	1,811
Measures of ln H_{icj}		
Number of Books	0	5,250
Number of Computers	0	5,273
Parents' Education		3,960
Did not finish HS	19.8	
High School	20.86	
Some College	13.66	
4-year College	18.41	
Professional or Grad. degree	27.27	
Measures of ln L_{icj}		
log Hours/week of study	.85	5,267
Homework Completed:		5,218
None of it/Never Assigned	3.6	
Some of it	11.38	
Most of it	23.23	
All	53.74	
All plus some extra	8.05	
Stop Trying:		5,347
Totally Untrue	4.54	
Mostly Untrue	5.72	
Somewhat Untrue/True	14.77	
Mostly True	22.65	
Totally True	52.31	
Measures of ln E_{Mcj}		
Doesn't let give up:		5,388
Totally Untrue	3.92	
Mostly Untrue	4.21	
Somewhat Untrue/True	16.44	
Mostly True	25.41	
Totally True	50.02	
Better Thinker:		5,275
Totally Untrue	6.22	
Mostly Untrue	6.12	
Somewhat Untrue/True	20.61	
Mostly True	26.33	
Totally True	40.72	
Accepts Nothing Less:		5,215
Totally Untrue	3.78	
Mostly Untrue	4.49	
Somewhat Untrue/True	19.64	
Mostly True	26.79	
Totally True	45.31	

	Mean	Obs.
Measures of ln M_{cj}		
Students Treat with Respect:		5,419
Totally Untrue	6.92	
Mostly Untrue	9.36	
Somewhat Untrue/True	26.41	
Mostly True	32.22	
Totally True	25.10	
Behavior under control:		5,325
Totally Untrue	12.56	
Mostly Untrue	12.96	
Somewhat Untrue/True	26.89	
Mostly True	26.31	
Totally True	21.28	
Classmates behave:		5,313
Totally Untrue	14.76	
Mostly Untrue	14.79	
Somewhat Untrue/True	31.02	
Mostly True	25.26	
Totally True	14.17	

Table A.3: Model Fit: Estimated Value-Added Regression

	Actual		Model
	Coefficient	SE	Coefficient
Lagged Math Score	.602	.011	.583
Av. Lagged Math Score	.179	.028	.211
Age	-.039	.008	-.034
Gender:Male	-.016	.014	.015
Race: African American	-.146	.025	-.141
Race: Hispanic	-.039	.025	-.079
Race: Asian	.141	.037	.050
Gifted Status	.211	.034	.161
Special Education	-.246	-.032	-.071
English Language Learner	-.048	.025	-.105
log(Class Size)	-.247	.062	-.062
R^2	.705		.689
Standard Deviation of Teachers' Effects	.260		.271

Note: The dependent variable is math test score and sample size is 6,135 students assigned to 149 teachers. The control variables include students' characteristics, their corresponding classroom average, (log) classroom size, and teacher fixed effects. The standard deviation of the estimated teacher fixed effects gives the standard deviation of the value added.

A.1 Additional Model Material

Consider N_{cj} students in classroom c assigned to teacher j. Students choose studying effort $L_{icj} \in [0, \infty)$, and the teacher chooses managerial effort $E_{Mcj} \in [0, \infty)$. The teacher exerts effort to maximize utility from their students' achievement, net the cost of effort. Students exert effort to maximize utility from achievement, net the cost of learning effort. The teacher moves first and students move second. The teacher values the average response of the students, which I denote as $\bar{L}^{br}_{cj}(E_{Mcj})$. Assume that the utility functions are twice differentiable and strictly concave. Suppose further that the knowledge production function is increasing in the teacher's and students' effort. Similar to Gal-Or (1985), I define the subgame perfect Nash equilibrium with sequential move as:

Definition (Nash Equilibrium with Sequential Move). Students' and teachers' efforts $(L^*_{1cj}, L^*_{2cj}, ..., L^*_{N_{cj}cj}, E^*_{Mcj})$ corresponds to a Nash equilibrium with sequential move if:

$$L^*_{icj} \equiv L^{br}_{icj}(E^*_{Mcj}) = \underset{L}{\text{argmax}} \quad U^s(L_{icj}, E^*_{Mcj}) \quad \forall \text{i in classroom c,} \quad \text{and}$$

$$E^*_{Mcj} = \underset{E_M}{\text{argmax}} \quad U^t(E_{Mcj}, \bar{L}^{br}_{cj}(E_{Mcj})).$$

The function $L^{br}_{icj}(E^*_{Mcj})$ is the students' best response, Equation (2.9), when teachers exert the equilibrium amount of effort, E^*_{Mcj}. I omit the endowments from Definition A.1 to keep the notation simple.

If an interior solution satisfying Definition A.1 exists, then students' first order conditions are:

$$U^s_L(L^*_{icj}, E^*_{Mcj}) = 0 \quad \forall \text{i in classroom c,}$$

where the subscript L denotes the partial derivative of the students' utility function with respect to L_{icj}, and the second order conditions are

$$U^s_{LL}(L^*_{icj}, E^*_{Mcj}) < 0 \quad \forall \text{i in classroom c.}$$

Given the model parametrization, the students' second order condition holds if $\alpha_L \gamma_L < 1$. Additionally, the sign of the slope of students' best response $L^{br}_{1cj}(E^*_{Mcj})$ is determined by

$$\frac{\partial L^{br}}{\partial E_M} = -\frac{U^s_{LE}}{U^s_{LL}} \lesseqgtr 0,$$

which, as Equation (2.9) shows, is positive if $\frac{\gamma_L \alpha_{M_{icj}} - \beta_M}{1 - \alpha_L \gamma_L} > 0$.

For the teacher, the first order condition is

$$U^t_E(E^*_{Mcj}, \bar{L}^*_{cj}) + U^t_{\bar{L}}(E^*_{Mcj}, \bar{L}^*_{cj}) \frac{\partial \bar{L}}{\partial E_M} = 0,$$

and the second order condition is

$$U^t_{EE}(E^*_{Mcj}, \bar{L}^*_{cj}) + 2 \times U^t_{\bar{L}E}(E^*_{Mcj}, \bar{L}^*_{cj})\frac{\partial \bar{L}}{\partial E_M} + U^t_{\bar{L}}(E^*_{Mcj}, \bar{L}^*_{cj})\frac{\partial^2 \bar{L}}{\partial E^2_M} + U^t_{LL}(E^*_{Mcj}, \bar{L}^*_{cj})(\frac{\partial \bar{L}}{\partial E_M}) < 0.$$

Notice that the strict concavity of the teacher utility function does not guarantee by itself the existence and uniqueness of an equilibrium. To find the sufficient conditions, I first replace the students' best response, Equation (2.9), and the classroom management practice, Equation (2.2), in the knowledge production function, Equation (2.3), and then apply the geometric mean:

$$\bar{Y}_{cj} = \Big[\frac{\alpha_L^{\alpha_L} \times \bar{K}_{cj}^{(\alpha_K+\alpha_{\bar{K}})} \times \mu_{Mcj}^{(\alpha_{Mcj}-\alpha_L\beta_M)} \times \theta_{Gj}^{\alpha_{Gcj}}}{c_L^{\alpha_L} \times \bar{H}_{cj}^{\alpha_L\beta_H}}\Big]^{\frac{1}{1-\alpha_L\gamma_L}} \times E_{Mcj}^{\frac{\delta_E(\alpha_{Mcj}-\alpha_L\beta_M)}{1-\alpha_L\gamma_L}}$$

where $\alpha_{Mcj} = \alpha_M + \alpha_{MK} \ln \bar{K}_{cj}$, and $\alpha_{Gcj} = 1 + \alpha_{GK} \ln \bar{K}_{cj}$. The average classroom knowledge is increasing in teachers' effort if the effort's exponent is positive, $\frac{\delta_E(\alpha_{Mcj}-\alpha_L\beta_M)}{1-\alpha_L\gamma_L} > 0$. Thus, after replacing $\bar{Y}_{cj}$ in the teacher's utility function, Equation (2.7), the second order condition for teacher j holds if

$$\alpha_L\gamma_L + \delta_E\gamma_{Ecj}(\alpha_{Mcj} - \alpha_L\beta_M) < 1.$$

The closed-form solution for teachers' effort, is

$$E^*_{Mcj} = \Bigg[\Big(\frac{\alpha_L}{c_L}\Big)^{\alpha_L\gamma_{Ecj}} \times \Big(\frac{\delta_E}{\theta_{Ej}} \times \frac{\alpha_{Mcj} - \alpha_L\beta_M}{1 - \alpha_L\gamma_L}\Big)^{1-\alpha_L\gamma_L} \times \Big(\frac{\bar{K}_{cj}^{\alpha_K+\alpha_{\bar{K}}} \times \theta_{Gj}^{\alpha_{Gcj}} \times \mu_{Mcj}^{\alpha_{Mcj}-\alpha_L\beta_M}}{\bar{H}_{cj}^{\alpha_L\beta_H}}\Big)^{\gamma_{Ecj}}\Bigg]^B \tag{A.1.1}$$

where the exponent B is given by

$$B = \frac{1}{1 - \alpha_L\gamma_L - \delta_E\gamma_{Ecj}(\alpha_{Mcj} - \alpha_L\beta_M)}.$$

Teacher's response to changes in endowments depends on the sign of the curvature parameter γ_{Ecj}. For instance, the teacher effort would be increasing in θ_{Gj} if $\gamma_{Ecj} > 0$, while it would be decreasing if $\gamma_{Ecj} < 0$ (assuming that $\alpha_{Gcj} > 0$).

Bibliography

Gal-Or, Esther. 1985. First mover and second mover advantages. *International Economic Review* :649–653.