Ability Tracking, School and Parental Effort, and Student Achievement: 
A Structural Model and Estimation*

Chao Fu † Nirav Mehta ‡

March 12, 2015

Abstract

We develop and estimate an equilibrium model of ability tracking, in which schools decide how to allocate students into ability tracks and choose track-specific effort; parents choose effort in response. The model is estimated using data from the ECLS-K. A counterfactual ban on tracking would benefit low-ability students but hurt high-ability students. Ignoring effort adjustments would significantly overstate the impacts. We then illustrate the tradeoffs involved when considering policies that affect schools’ tracking decisions. Setting proficiency standards to maximize average achievement would lead schools to redistribute their inputs from low-ability students to high-ability students.

*We thank Yuseob Lee, George Orlov, and Atsuko Tanaka for excellent research assistance. We thank John Bound, Betsy Caucutt, Tim Conley, Steven Durlauf, John Kennan, Lance Lochner, Michael Lovenheim, Jesse Rothstein, Jeffrey Smith, Steven Stern, Todd Stinebrickner, Chris Taber, and Kenneth Wolpin for insightful discussions. We have benefited from the comments from participants at the AEA 2014 winter meeting, Labo(u)r Day, and the CESifo Area Conference on the Economics of Education, and seminars at Amherst, ASU, Berkeley, Brock, Queen’s, Rochester, St. Louis, and UWO.

†Department of Economics, University of Wisconsin. 1180 Observatory Dr., Madison, WI 53706. Email: cfu@ssc.wisc.edu.

‡Department of Economics, University of Western Ontario. Email: nirav.mehta@uwo.ca
1 Introduction

Ability tracking, the practice of allocating students into different classrooms based on prior performance, is pervasive.\(^1\) It is also controversial, because it may benefit students of certain ability levels while hurting others. There is considerable policy interest in learning how ability tracking affects different types of students, and how policy changes, such as changing proficiency standards, would affect schools’ tracking choices and the distribution of student achievement. However, these questions are also very hard to answer. Indeed, the determinants and effects of tracking remain largely open questions, as evidenced by the mixed findings in the literature.\(^2\)

Carefully-designed experiments can potentially answer these questions. However, to learn the effects of tracking, experiments would have to be conducted on a representative distribution of schools, because, as Gamoran and Hallinan (1995) point out, the effectiveness of tracking depends on how it is implemented and on the school climate in which it is implemented. Similarly, to the extent that different policies may have very different impacts, one would need to run experiments under an extensive range of potential policies. The cost of doing so can become formidable very fast.

As a feasible alternative, we adopt a structural approach. We develop and estimate a model that treats a school’s tracking regime and track-specific effort, parental effort, and student achievement as joint equilibrium outcomes. The model is designed to address three interrelated components which have yet to be considered in a cohesive framework. First, changing peer composition in one classroom necessarily involves student re-allocation, hence changes peers in some other classroom(s). It is important not to treat classrooms in isolation when studying the treatment effect of changing peers, especially for policy making purposes, which require a balance between trade-offs. Second, by explicitly modeling school and parental effort inputs, we can then infer what these input levels and student achievement would be if tracking regimes, hence peer compositions, were changed. This allows us to decompose the effect of tracking into the direct effect of changes in peers and the indirect effects caused by the behavioral responses of schools and parents. As argued by Becker and

\(^1\)Loveless (2013) reports that 75% of US 8th-graders in the National Assessment of Educational Progress data were in tracked classes.

\(^2\)We review these studies below.
Tomes (1976), ignoring behavioral responses may bias estimated impacts of policy changes, which means that taking them into account is critical for policymaking. Finally, our explicit modeling of tracking regime choices allows us to predict how tracking regimes, which determine classroom-level peer composition, and subsequent school and parent inputs would change in response to a policy change.

In the model, each school is attended by children from different types of households, the distribution of which may differ between schools. A household type is defined by the child’s ability and by how costly it is for the parent to help her child learn. A child’s achievement depends on her own ability, effort inputs invested by the school and by her parent, and the quality of her peers. A parent maximizes her child’s achievement by choosing costly parental effort given both her child’s track assignment (which determines peer quality) and the effort invested by the school. Taking into account responses by parents, the school chooses a tracking regime, which assigns students to classrooms based on their ability, and track-specific effort inputs to maximize its own objective, which increases in the average achievement of its students and the fraction of students satisfying a proficiency standard.

This framework naturally allows policies to produce winners and losers because changes in tracking regimes at a school may differentially affect students of different ability levels and parental backgrounds. Moreover, given that the model allows schools to base tracking decisions on the composition of attendant households, policies that affect schools’ tracking decisions will necessarily create heterogeneous effects at the school level. That is, there will be a distribution of treatment effects within each school, and this distribution may itself differ across schools.

We estimate the model using data from the nationwide Early Childhood Longitudinal Study (ECLS-K), which are rich enough to allow us to model the interactions between schools and parents. Students are linked to their parents and teachers. For students, we observe prior test scores, class membership, and end-of-the-year test scores. Parents report the frequencies with which they help their children with homework (parental effort). Teachers report the class-specific workload (school effort), and, as a unique feature of the survey, teachers also report the overall level of

---

3See Epple and Romano (2011) for a recent review of the literature on peer effects in education.
4The importance of these types of input adjustments has been supported by evidence found in the literature. For example, Pop-Eleches and Urquiola (2013) and Das et al. (2011) both document changes in home inputs in response to quasi-experimental variation in school inputs.
ability among students in each of their classes relative to other students in the same school. This rare last piece of information (available for multiple classes) allows us to essentially “observe” two of the key components of the model. The first is the tracking regime used by each school. The second is the composition of peer ability in each track. Each track comprises a section of the school-specific ability distribution; the section is specified by the tracking regime. Such information directly reveals peer quality, despite the fact that a student’s ability is not directly observable and that the grouping of students into different tracks is endogenous. Therefore, we are able to separate the roles of own ability and peer quality in the technology, a common identification concern in this literature.

We use the estimated model to conduct two policy evaluations. First, we quantify the effects of allowing ability tracking on the distribution of student achievement by comparing outcomes from the baseline model, where schools choose tracking regimes, with counterfactual outcomes where no schools are allowed to track students, i.e. class composition is equal across classrooms in a school. Over 95% of schools in our sample practice ability tracking under the baseline and therefore are affected by this policy. A tracking ban increases peer quality for low-ability students and decreases peer quality for high-ability students. In response, schools increase their effort overall, but parents react differently depending on their child’s ability. The increase in the peer quality for low-ability students is effectively an increase in these households’ endowments, causing their parents to reduce their provision of costly effort. Conversely, parents of high-ability students increase their effort because their children now have lower quality peers. Altogether, banning tracking increases the achievement of students with below-median prior scores by 4% of a standard deviation (sd) in outcome test score and reduces the achievement of students with above-median prior scores by 5% sd. We find the average treatment effect from banning tracking is very small, so the effects of tracking are mostly distributional in nature.

It is worth noting, however, that peer quality is estimated to be an important input in the production function: holding all other inputs at their average levels and increasing peer quality by one standard deviation (sd) above the average causes a 24% sd increase in the outcome test score. Underlying the small net effect are the behavioral responses of parents and schools. Indeed, if the effort adjustments by schools and parents in response to the ban on tracking were ignored, one would
overstate the loss for students with above-median prior scores by 57%, and overstate the gain for students with below-median prior scores by over 100%. That is, the effect of banning tracking would appear much larger were we to hold school and parental effort constant at their levels under the status quo. Our work helps to cast light on the open question of why estimated peer effects are typically small, as reviewed by Sacerdote (2011). It is exactly the fact that peer quality is an important input that induces substantial behavioral responses by schools and parents, which in turn mitigate the direct effect of peer quality.

Because policies that affect schools’ tracking decisions will necessarily lead to increases of peer quality for some students and decreases for some other students, our second counterfactual experiment highlights the trade-offs involved in achieving academic policy goals. In the model, schools value both average achievement and the fraction of students testing above a proficiency standard, which policymakers may have control over. In this counterfactual, we search for region-specific proficiency standards that would maximize average student test scores in each Census region. Achievement-maximizing standards would be higher than their baseline levels in every region, but do not increase without bound because eventually the density of student ability decreases, reducing the gain in average achievement to increasing standards. Under these higher proficiency standards, schools would adjust their effort inputs and tracking regimes such that resources are moved away from low-ability students toward high-ability students, leading to decreases in achievement for low-ability students and increases in achievement for high-ability students.

This paper contributes to and brings together two strands of studies on ability tracking: one measuring how tracking affects student achievement and one studying the determinants of tracking decisions. We take a first step in the literature towards creating a comprehensive framework to understand tracking, by building and estimating a model where tracking regimes and track-specific inputs, parental effort, and student achievement are viewed as joint equilibrium outcomes.

There is considerable heterogeneity in results from empirical work assessing the effect of ability tracking on both the level and distribution of achievement. Argys et al. (1996) find that tracking reduces the performance of low ability students, while

---

Betts and Shkolnik (2000b) and Figlio and Page (2002) find no significant differences in outcomes for US high school students of the same ability level at tracked and untracked schools. Duflo et al. (2011) run an experiment in Kenya and find that students of all abilities gain from a tracking regime where students were assigned to high- and low-ability classrooms, relative to a control group where students were randomly assigned to the two classrooms. Gamoran (1992) finds that the effects of ability tracking on high school students vary across schools. The heterogeneous findings from these studies indicate that there is no one “causal effect” of ability tracking which generalizes to all contexts, and therefore highlight the importance of explicitly taking into account the fact that households differ within a school and that the distributions of households differ across schools, both of which are fundamental to our model and empirical analyses.\footnote{There is also a literature studying between-school tracking, e.g. Hanushek and Woessmann (2006).}

Fruehwirth (2013) identifies peer achievement spillovers using the introduction of a student accountability policy as an instrument.\footnote{See Manski (1993), Moffitt (2001), and Brock and Durlauf (2001) for methodological contributions concerning the identification of peer effects and Betts and Shkolnik (2000a) for a review on empirical work in this respect.} She uses this instrumental variable to eliminate the problem of nonrandom assignment in identifying peer achievement spillovers, which is a valid approach as long as students are not reassigned in response to the policy.\footnote{The author provides support for this assumption by showing that observable compositions of peer groups do not appear to change in response to student accountability. Assuming random assignment to classrooms within schools, Fruehwirth (2014) finds positive effects of peer parental education on student achievement.} Our paper complements her work. Unlike Fruehwirth (2013), we assume that a student’s achievement does not depend directly on the effort choices by other students (households), hence abstracting from direct social interactions within a class.\footnote{See Blume et al. (2011) for a comprehensive review of the social interactions literature.} Instead, we model schools’ decisions about student assignment and track-specific inputs in order to study the nonrandom assignment of students across classrooms, and how parents respond.

While our work focuses on how peer groups are determined within a school, a different literature studies how households sort themselves into different schools. Epple et al. (2002) study how ability tracking by public schools may affect student sorting between private and public schools. They find that when public schools track
by ability, they may attract higher ability students who otherwise would have attended private schools. This is consistent with our finding that high-ability students benefit from tracking. Caucutt (2002), Epple and Romano (1998), Ferreyra (2007), Mehta (2013), Nechyba (2000) develop equilibrium models to study sorting between schools and its effects on peer composition.\textsuperscript{10} Walters (2014) studies the demand for charter schools. Our work complements this literature by taking a first step toward studying schools’ tracking decisions, which determine class-level peer groups faced by households within a school, and emphasizing the interactions between a school and attendant households in the determination of student outcomes.

The rest of the paper is organized as follows. The next section describes the model. Section 3 describes the data. Section 4 explains our estimation and identification strategy. Section 5 presents the estimation results. Section 6 conducts counterfactual experiments. The last section concludes. The appendix contains additional details about the data and model.

2 Model

Each school is treated as a closed economy. A school makes decisions about ability tracking and track-specific inputs, knowing that parents will subsequently respond by adjusting parental effort.

2.1 The Environment

A school $s$ is endowed with a continuum of households of measure one. Households are of different types in that students have different ability levels ($a$) and parents have different parental effort costs ($z \in \{\text{low, high}\}$).\textsuperscript{11} Student ability $a$ is known to the household and the school, but $z$ is a household’s private information, which implies that all students with the same ability level are treated identically by a school. Let $g_s(a, z), g_s(a)$ and $g_s(z|a)$ denote, respectively, the school-$s$ specific

\textsuperscript{10}Mehta (2013) also endogenizes school input choices.

\textsuperscript{11}We treat student ability as one-dimensional in the model, and our empirical analysis focuses on one subject: Reading (student test scores are highly correlated (0.7) between Reading and Math). A more comprehensive model should jointly consider tracking and input decisions across multiple subjects, an interesting extension we leave for future research.
joint distribution of household types, marginal distribution of ability, and conditional
distribution of \( z \) given \( a \). In the following, we suppress the school subscript \( s \).

### 2.1.1 Timing

The timing of the model is as follows:

Stage 1: The school chooses a tracking regime and track-specific effort inputs.

Stage 2: Observing the school’s choices, parents choose their own parental effort.

Stage 3: Student test scores are realized.

### 2.1.2 Production Function

The achievement of a student \( i \) in track \( j \) depends on the student’s ability \( (a_i) \), the average ability of students in the same track \( (q_j) \), track-specific school effort \( (e_s^j) \), and parental effort \( (e_p^i) \), according to \( Y(a_i, q_j, e_s^j, e_p^i) \).\(^{12}\) The test score \( y_{ji} \) measures student achievement with noise \( \epsilon_{ji} \sim F_\epsilon (\cdot) \), which has density \( f_\epsilon \), such that

\[
y_{ji} = Y(a_i, q_j, e_s^j, e_p^i) + \epsilon_{ji}. \tag{1}
\]

### 2.2 Parent’s Problem

A parent derives utility from her child’s achievement and disutility from exerting parental effort. Given the track-specific school input \( (e_s^j) \) and the peer quality \( (q_j) \) of her child’s track, a parent \( i \) chooses her own effort to maximize the utility from her child’s achievement, net of her effort cost \( C_p (e_p^i, z_i) \):

\[
u(e_p^i, q_j, a_i, z_i) = \max_{e_p^i \geq 0} \{ \ln \left( Y(a_i, q_j, e_s^j, e_p^i) \right) - C_p (e_p^i, z_i) \},
\]

where \( u(\cdot) \) denotes the parent’s value function. Denote the optimal parental choice \( e_p^*(e_s^j, q_j, a_i, z_i) \). Notice that a parent’s optimal choice depends not only on her child’s ability \( a_i \) and her cost type \( z_i \), but also on the other inputs in the test score production, including the ones chosen by the school \( (q_j \ and \ e_s^j) \).

\(^{12}\)Arguably, the entire distribution of peer ability may matter. Following the literature, for feasibility reasons we have allowed only moments of the ability distribution, in our case the average and the variance of peer quality, to enter the production function. The variance of peer quality is dropped because it is found insignificant.
2.3 School’s Problem

A school cares about the total test score of its students and may also care about the fraction of students who can pass a proficiency standard \( y^* \). It chooses a tracking regime, which specifies how students are allocated across classrooms based on student ability, and track-specific inputs. Formally, a tracking regime is defined as follows.\(^{13}\)

**Definition 1** Let \( \mu_j(a) \in [0,1] \) denote the fraction of ability-\( a \) students assigned to track \( j \), such that \( \sum_j \mu_j(a) = 1 \). A tracking regime is defined as \( \mu = \{\mu_j(\cdot)\}_j \).

If no student of ability \( a \) is allocated to track \( j \), then \( \mu_j(a) = 0 \). If track \( j \) does not exist, then \( \mu_j(\cdot) = 0 \). Because all students with the same ability level are treated identically, \( \mu_j(a) \) is also the probability that a student of ability \( a \) is allocated to track \( j \). The school’s problem can be viewed in two steps: 1) choose a tracking regime; 2) choose track-specific inputs given the chosen regime. The problem can be solved backwards.

### 2.3.1 Optimal Track-Specific School Effort

Let \( e_s^* \equiv \{e^*_j\}_j \) denote a school effort vector across the \( j \) tracks at a school. Given a tracking regime \( \mu \), the optimal choice of track-specific effort solves the following problem:

\[
V_s(\mu) = \max_{e^s \geq 0} \left\{ \int_i \left\{ \sum_j \left[ E(\epsilon_i) \left( (y_{ji} + \omega I(y_{ji} > y^*)) | a_i \right) - C^s(e^*_j) \right] \mu_j(a_i) \right\} di \right. \\
\left. \text{s.t. } \begin{align*}
y_{ji} &= Y(a_i, q_j, e^*_j, e^p_i) + \epsilon_{ji} \\
e^p_i &= e^{ps}(e^*_j, q_j, a_i, z_i) \\
n_j &= \sum_a \mu_j(a) g_s(a) \\
q_j &= \frac{1}{n_j} \sum_a \mu_j(a) g_s(a) a.
\end{align*} \right\},
\]

where \( V_s() \) denotes the school’s value function given regime \( \mu \). The parameter \( \omega \geq 0 \) measures the additional valuation a school may derive from a student passing

---

\(^{13}\)Another type of tracking, which we do not model, may happen within a class. We choose to focus on cross-class tracking because it is prevalent in the data (95% of schools track students across classes). In contrast, when asked how often they split students within a class by ability levels, the majority of teachers in the ECLS-K either report never or report doing so for less than once per week.
the proficiency standard \((y^*)\).\(^{14}\) \(C^s(e^*_j)\) is the per-student effort cost on track \(j\). The terms in the square brackets comprise student \(i\)'s expected net contribution conditional on her being on track \(j\) (denominated in units of test scores), where the expectation is taken over both the test score shock \(\epsilon_{ji}\) and the distribution of parent type \(z_i\) given student ability \(a_i\), which is \(g_s(z|a)\). In particular, a student contributes by her test score \(y_{ji}\) and an additional \(\omega\) if \(y_{ji}\) is above \(y^*\). Increases in achievement increase the probability a student’s realized test score \(y_{ji}\) exceeds the threshold, but there is still some probability the child’s score may be below \(y^*\). Student \(i\)’s total contribution to the school’s objective is a weighted sum of her track-specific contributions, where the weights are given by her probabilities of being assigned to each track, \(\{\mu_j(a_i)\}_j\). The overall objective of a school is thus the integral of individual students’ contributions. There are four constraints a school faces, listed in (2). The first two are the test score technology and the optimal response of the parent. The last two identity constraints show how the tracking regime \(\mu\) defines the size \((n_j)\) and the average student quality \((q_j)\) of a track. Let \(e^{**}(\mu)\) be the optimal solution to (2).

### 2.3.2 Optimal Tracking Regime

A school’s cost may vary with tracking regimes, captured by the function \(D(\mu)\). Balancing benefits and costs of tracking, a school solves the following problem:\(^{15}\)

\[
\max_{\mu \in M_s} \left\{ V_s(\mu) - D(\mu) + \eta_\mu \right\}.
\]

where \(\eta_\mu\) is the school-specific idiosyncratic shifter associated with regime \(\mu\), which is i.i.d. across schools. \(M_s\) is the support of tracking regime for school \(s\), which is specified in Section 2.5.2.

\(^{14}\)Notice that whether or not passing \(y^*\) affects the school’s objective is allowed for, not imposed by, our model structure. It is an empirical question, the answer to which depends on whether or not the estimated \(\omega\) is strictly positive.

\(^{15}\)We assume the tracking decision is made by the school. In reality, it is possible that some parents may request that their child be placed in a certain track. We abstract from this in the model and instead focus on parental effort responses.
2.4 Equilibrium

Definition 2 A subgame perfect Nash equilibrium in school \(s\) consists of \(\{e^p(\cdot), e^s(\cdot), \mu^*\}\), such that
1) For each \((e^s_j, q_j, a_i, z_i)\), \(e^p(\cdot)\) solves parent’s problem;
2) \((e^s(\mu^*), \mu^*)\) solves school’s problem.\(^{16}\)

We solve the model using backward induction. First, solve the parent’s problem for any given \((e^s_j, q_j, a_i, z_i)\). Second, for a given \(\mu\), solve for the track-specific school inputs \(e^s\). Finally, optimize over tracking regimes to obtain the optimal \(\mu^*\) and the associated effort inputs \((e^p(\cdot), e^s(\cdot))\).\(^{17}\)

2.5 Further Empirical Specifications

We present the final empirical specifications we have chosen after estimating a set of nested models, as will be further discussed at the end of this subsection.

2.5.1 Household Types

There are six types of households in a school: two types of parents (low and high effort cost) interacted with three school-specific student ability levels \((a^s_1, a^s_2, a^s_3)\).\(^{18}\)

Household types are unobservable to the researcher but may be correlated with observable household characteristics \(x\), which include a noisy measure of student ability \((x^a)\) and parental characteristics \((x^p)\) containing parental education and an indicator of single parenthood. Let \(\Pr((a, z) | x, s)\) be the distribution of \((a, z)\) conditional on \(x\) in school \(s\). The joint distribution take the following form, the details of which are in Appendix A.1

\[
\Pr((a^s_l, z) | x, s) = \Pr(a = a^s_l | x^a, s) \Pr(z | x^p, a^s_l).
\]

This specification has three features worth commenting on. First, ability distributions are school-specific and discrete. This assumption allows us to tractably model

\(^{16}\)The equilibrium at a particular school also depends on the distribution of households \(g_s(a, z)\) and \(\eta_s\). We suppress this dependence for notional convenience.

\(^{17}\)This involves evaluating a school’s objective function at all possible tracking regimes.

\(^{18}\)Specifically, we use a three-point discrete approximation of a normal distribution for each school, as described in Appendix A.1.
unobserved student heterogeneity in a manner that allows ability distributions to substantially vary between schools, which is important to our understanding why schools make different tracking decisions. Second, given that the pre-determined student ability is itself an outcome from previous parental inputs, the latter being affected by parental types, the two unobserved characteristics of the households are inherently correlated with each other. We capture this correlation by assuming that the distribution of parental types depends not only on parental characteristics but also on student ability. Third, the noisy measure of student ability (previous test score) is assumed to be a sufficient statistic for ability, conditional on which the ability distribution does not depend on parental characteristics. We are not, however assuming that parental characteristics do not matter for ability, rather that the previous test scores summarize all the information. Indeed, the pre-determined ability is itself affected by parental investment in the past, as can be seen easily from the joint distribution of \((x^a, x^p)\).

### 2.5.2 Tracking Regime

The support of tracking regimes \((M_s)\) is finite and school-specific, and is subject to two constraints. First, the choice of tracking regimes in each school is constrained by both the number of classrooms and the size of each classroom (i.e., the fraction of students that can be accommodated in a classroom). Let \(K_s\) be the number of classrooms in school \(s\), we assume that the size of a particular track can only take values from \(\left\{0, \frac{1}{K_s}, \frac{2}{K_s}, \ldots, 1\right\}\). Schools with more classrooms have finer grids of \(M_s\). Second, the ability distribution within a track cannot be “disjoint” in the sense that a track cannot mix low- and high-ability students while excluding middle-ability students. Subject to these two constraints, \(M_s\) contains all possible ways to allocate students across the \(K_s\) classrooms. If a track contains multiple classrooms \(\left(n_j > \frac{1}{K_s}\right)\), the composition of students is identical across classrooms in the same track.

The cost of a tracking regime depends only on the number of tracks, and is given by

\[
D(\mu) = \gamma_{|\mu|},
\]

where \(|\mu| \in \{1, 2, 3, 4\}\) is the number of tracks in regime \(\mu\). \(\gamma = [\gamma_1; \gamma_2; \gamma_3; \gamma_4]\) is the vector of tracking costs, with \(\gamma_1\) normalized to 0.
The permanent idiosyncratic shifter in the school objective function, $\eta_{\mu}$, follows an extreme-value distribution. From the researcher’s point of view, conditional on a set of parameter values $\Theta$, the probability of observing a particular tracking regime $\tilde{\mu}$ in school $s$ is given by

$$\frac{\exp \left( V_s(\tilde{\mu}|\Theta) \right)}{\sum_{\mu'} \exp \left( V_s(\mu'|\Theta) \right)}.$$  

(3)

2.5.3 Achievement Function

Student achievement is governed by

$$Y(a, q, e^s, e^p) = \alpha_0 + a + \alpha_1 e^s + \alpha_2 e^p + \alpha_3 q + \alpha_4 e^s a + \alpha_5 e^s e^p,$$

(4)

where the coefficient on one’s own ability $a$ is set to one. The last two interaction terms allow the marginal effect of school effort on student achievement to depend on student ability and parental effort.

2.5.4 Cost Functions

The cost functions for both parent and school effort are assumed to be quadratic. The cost of parental effort is type-specific, where types differ in the linear coefficient $z \in \{z_1, z_2\}$, but are assumed to share the same quadratic coefficient ($c^p$), such that

$$C^P(e^p, z) = ze^p + c^p(e^p)^2.$$  

The cost of school effort is given by

$$C^s(e^s) = c^s_1 e^s + c^s_2(e^s)^2.$$  

2.5.5 Measurement Errors

We assume that both the school effort $e^s$ and the parent effort $e^p$ are measured with idiosyncratic errors. The observed school effort in track $j$ ($\tilde{e}^s_j$) is given by

$$\tilde{e}^s_j = e^{s*}_j + \zeta^s_j,$$  

(5)

where $\zeta^s_j \sim N(0, \sigma^2_j)$. For parental effort, which is reported in discrete categories, we use an ordered probit
model to map model effort into a probability distribution over observed effort, as specified in Appendix A.2.1. Let \( \Pr(\tilde{e}^p|e^p) \) denote the probability of observing \( \tilde{e}^p \) when model effort is \( e^p \).

### 2.5.6 Further Discussion of Model Specification

We have estimated a set of nested models including versions that are more general than the one presented to include additional and potentially important features.\(^\text{19}\) We have chosen the relatively parsimonious specifications presented here because, in likelihood ratio tests, we cannot reject that the simpler model fits our data as well as the more complicated versions.\(^\text{20}\) To be specific, we have allowed more general specifications in the following three broad aspects. The first involves more flexible test score production functions. For example, we have allowed for additional interaction terms between \( q \) and \( e^s \), as well as \( q \) and \( e^p \) to allow for the possibility that the effect of peer quality may vary with effort inputs. Following the discussion of non-linear peer effects in Sacerdote (2011), we also allow for an interaction effect between \( q \) and \( a \). To entertain the possibility that teaching may be more or less effective in classes with widely-heterogeneous students, we have also included in the production function the dispersion of class ability, as measured by the coefficient of variation, and its interaction with \( e^s \). The second aspect is on the household side, where we have allowed for additional heterogeneity in the effectiveness of parental effort besides that in parental effort costs. Finally, on the school side, we have allowed for 1) a nested logit counterpart of equation (3) in school’s regime choices, and 2) an additional payoff in the school’s objective function that increases with the fraction of students achieving outstanding test scores.

**Remark 1** Our finding that the non-linear terms are not important in the technology does not contradict the literature documenting evidence of non-linear peer effects. In a reduced form of our model, student achievement is a highly non-linear function of inputs, in particular peer quality, as illustrated by Appendix C.

\(^{19}\)The final choice of the detailed specification of the model is, of course, data-dependent. The more general models we have considered do not add much more complication to the estimation or the simulation. We view this as good news for future work that uses a different dataset because one can feasibly extend our current specification.

\(^{20}\)The conclusions from our counterfactual experiments are robust to the inclusion of these additional features (results available upon request).
3 Data

We use data from the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K). The ECLS-K is a national cohort-based study of children from kindergarten entry through middle school. Information was collected from children, parents, teachers, and schools in the fall and spring of children’s kindergarten year (1998) and 1st grade, as well as the spring of 3rd, 5th, and 8th grade (2007). Schools were probabilistically sampled to be nationally representative. More than 20 students were targeted at each school for the first (kindergarten) survey round. This results in a student panel which also serves as a repeated cross section for each school. The ECLS-K assessed student skills that are typically taught and developmentally important, such as math and reading skills. We focus on 5th grade reading classes.\(^{21}\)

The data are rich enough to allow us to model the interactions between schools and parents. For students, we observe their prior (3rd grade) test scores (which are related to their ability), class membership (to identify their ability track), and end-of-the-year test scores, where test scores are results from the ECLS-K assessment. Students are linked to parents, for whom we have a measure of parental inputs to educational production (frequency with which parents help their child with homework), and parental characteristics such as education and single-parenthood (which may affect parenting costs).\(^{22}\) Assuming that homework loads on students increase teachers’ effort cost, we use homework loads reported by the teacher to measure the school’s effort invested in each class.\(^{23}\)

For the tracking regime, we exploit data unique to this survey. In particular, we use teachers’ reports on the ability level of their classes, which are available for different classes in the same school.\(^{24}\) The question for reading classes is: “What is the reading ability level of this child’s reading class, relative to the children in your

\(^{21}\) We focus on reading instead of math because the former has a much larger sample size.

\(^{22}\) Ferreyra and Liang (2012) use time spent doing homework as a measure of student effort.

\(^{23}\) The same measure has been used in the study of the relationship between child, parent, and school effort by De Fraja et al. (2010). Admittedly, the use of homework loads as school effort is largely due to data limitations, yet it is not an unreasonable measure. Homework loads increase teachers’ effort, who have to create and/or grade homework problem sets. Moreover, a teacher may face complaints from students for assigning too much homework. See Appendix A.2 for the survey questions.

\(^{24}\) The ECLS-K follows many students at the same school. As such, we have the above information on classes for several classes at each school.
school at this child’s grade?” A teacher chooses one of the following four answers: a) Primarily high ability, b) Primarily average ability, c) Primarily low ability and d) Widely mixed ability.\textsuperscript{25} We use the number of distinct answers given by teachers in different classes as the number of tracks in a school. Classes with identical teacher answers to this question are viewed as in the same track. The size of each track is calculated as the number of classes in that track divided by the total number of classes. Although the relative ability ranking is a priori obvious between answers a), c) and b) or d), the relative ranking between b) and d) is less so. We use the average student prior test scores within each of the two types of classes to determine the ranking between them in schools where both tracks exist. As a result, higher tracks have students with higher mean ability.\textsuperscript{26}

Finally, the data indicate the Census region in which each school is located. We set proficiency cutoff $y^*$ per Census region to match the proficiency rate in the data with that in the Achievement Results for State Assessments data. This data contains state-specific proficiency rates, which we aggregate to Census region.\textsuperscript{27}

There are 8,853 fifth grade students in 1,772 schools in the ECLS-K sample. We delete observations missing key information, such as prior test scores, parental characteristics and track identity, leaving 7,332 students in 1,551 schools. Then, we exclude schools in which fewer than four classes or ten students are observed. The final sample includes 2,789 students in 205 schools. The last sample selection criterion costs us a significant number of observations. Given our purpose of studying the equilibrium within each school, this costly cut is necessary to guarantee that we have a reasonably well-represented sample for each school. Despite this costly cut, Table 19 shows that, among the observed 5th-graders, the summary statistics for the entire ECLS-K sample and our final sample are not very different. A separate issue is the sample attrition over survey rounds. The survey started with a nationally-representative sample of 16,665 kindergarteners, among whom 8,853 remained in the sample by Grade 5. Statistics for the first survey round are compared between the whole sample and the sample of stayers in Table 20, which shows that the sample attrition is largely random.\textsuperscript{28}

\textsuperscript{25}This variable was also used by Figlio and Page (2002) in their study of tracking.
\textsuperscript{26}See Appendix B.
\textsuperscript{27}https://inventory.data.gov. See Table 15 in this paper for regional cutoffs. Details are available upon request.
\textsuperscript{28}The only slight non-randomness is that students who are more likely to remain in the sample
Like most datasets, the ECLS-K has both strengths and weaknesses compared with other data. Administrative data sets typically contain test score information for all students in a school, which has spurred a large literature estimating reduced-form social interactions models wherein one’s own outcome is affected by both own and peer characteristics, as well as peer outcomes (Manski (1993), Moffitt (2001), Sacerdote (2011)). However, such datasets typically lack detailed measures of school and parental inputs, due to the fact that they are not based on surveys. It is precisely these input data that we use to study the endogenous interactions between the school and the parents. The tradeoff from using survey data is that it does not contain a large number of students per school. However, as far as we know, no currently available data set, other than the ECLS-K, contains the information that is essential for estimating a model like ours. Moreover, Tables 19 and 20 suggest that our final estimation sample is largely representative of the ECLS-K sample, which itself is nationally representative.

3.1 Descriptive Statistics

The first row of Table 1 shows that over 95% of schools practice ability tracking, i.e., they group students into more than one track. The most common number of tracks is three, which accounts for 46% of all schools. About 13% of schools have four tracks. To summarize the distribution of students across schools, for each school we calculate the mean and the coefficient of variation (CV) of student prior test scores and the fraction of students in the school whose prior scores were below the sample median. Rows 2-4 of Table 1 present the mean of these summary statistics across schools, by the number of tracks in the school. On average, schools with more tracks have higher dispersion and lower prior scores. For example, the average prior test score in schools with only one track is 53.4 with a CV of 0.14 and fewer than 45% of students below the median. In contrast, the average prior test score in schools with four tracks is 50.0 with a CV of 0.17 and more than 57% of students below the median. Tables 2-4 present summary statistics by the number of tracks in the school and the identity of a track. For example, entries in (Columns 7-8, Row 3) of a table refer to students who belong to the third track in a school are those with both parents (75% in the whole sample vs. 79% in the final sample) and/or college-educated parents (49% vs. 51%). Although it is comforting to see that the attrition is largely random, the slight non-randomness leads us to make some caveats, as discussed in Appendix D.
Table 1: Student Prior Test Scores in Schools by Numbers of Tracks

<table>
<thead>
<tr>
<th>Track</th>
<th>1 Track</th>
<th>2 Tracks</th>
<th>3 Tracks</th>
<th>4 Tracks</th>
<th>All Schools</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of schools</td>
<td>4.39</td>
<td>37.1</td>
<td>45.8</td>
<td>12.7</td>
<td>100.0</td>
</tr>
<tr>
<td>Mean</td>
<td>53.4</td>
<td>51.6</td>
<td>51.6</td>
<td>50.0</td>
<td>51.5</td>
</tr>
<tr>
<td>CV</td>
<td>0.139</td>
<td>0.165</td>
<td>0.160</td>
<td>0.173</td>
<td>0.163</td>
</tr>
<tr>
<td>% below median</td>
<td>44.1</td>
<td>48.3</td>
<td>49.8</td>
<td>57.1</td>
<td>50.0</td>
</tr>
</tbody>
</table>

with four tracks. Table 2 shows that students in higher ability tracks have both higher average outcome test scores and a higher probability of passing the regional proficiency cutoff. Comparing average student outcome scores and pass rates in schools with one track to those in other schools, they are similar to those of the middle-track students in schools with three or four tracks. That is, the average student allocated to the lower (higher) track in a multiple-track school has poorer (better) outcomes than an average student attending a single-track school. Table 3 shows average school effort by track. With the exception of Track 4 in schools with four tracks, the average school effort (expected hours of homework done by students per week) increases with track level, or track-specific average ability, in schools with more than one track. At the school level, the average effort level stays roughly constant with the number of tracks. Table 4 shows that parental effort shows the opposite pattern compared to school effort. Average parental effort (frequency of helping child with homework in reading) decreases with student ability or track level. For example, in schools with three tracks, while parents of low-track students on average help their children 2.6 times per week, parents of high-track students do so only 2.1 times.

Table 2: Average outcome score and percent of students passing the cutoff

<table>
<thead>
<tr>
<th>Track</th>
<th>1 track</th>
<th>2 tracks</th>
<th>3 tracks</th>
<th>4 tracks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>% pass</td>
<td>Score</td>
<td>% pass</td>
<td>Score</td>
</tr>
<tr>
<td>1</td>
<td>51.84</td>
<td>69.42</td>
<td>45.95</td>
<td>50.88</td>
</tr>
<tr>
<td>2</td>
<td>51.98</td>
<td>75.75</td>
<td>51.38</td>
<td>68.47</td>
</tr>
<tr>
<td>3</td>
<td>55.62</td>
<td>84.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>57.99</td>
<td>97.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>51.84</td>
<td>69.42</td>
<td>49.73</td>
<td>66.83</td>
</tr>
</tbody>
</table>
Table 5 summarizes parental effort and student outcomes by household characteristics. Compared to their counterpart, parents without college education, single parents and parents with lower-prior-achievement children exert more parental effort. Nevertheless, average student outcome test scores are lower in these households.

Table 3: Average teacher effort by track

<table>
<thead>
<tr>
<th>Track</th>
<th>1 track</th>
<th>2 tracks</th>
<th>3 tracks</th>
<th>4 tracks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.86</td>
<td>1.75</td>
<td>1.75</td>
<td>1.82</td>
</tr>
<tr>
<td>2</td>
<td>1.90</td>
<td>1.88</td>
<td>1.84</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>1.96</td>
<td>1.93</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>1.68</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>1.86</td>
<td>1.83</td>
<td>1.86</td>
<td>1.82</td>
</tr>
</tbody>
</table>

Table 4: Average parent effort by track

<table>
<thead>
<tr>
<th>Track</th>
<th>1 track</th>
<th>2 tracks</th>
<th>3 tracks</th>
<th>4 tracks</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.07</td>
<td>2.31</td>
<td>2.57</td>
<td>2.29</td>
</tr>
<tr>
<td>2</td>
<td>2.03</td>
<td>2.37</td>
<td>2.71</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2.11</td>
<td></td>
<td>2.78</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>2.08</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>2.07</td>
<td>2.11</td>
<td>2.27</td>
<td>2.38</td>
</tr>
</tbody>
</table>

Table 5: Parent effort and outcome test score by observed characteristics

<table>
<thead>
<tr>
<th></th>
<th>Parent effort</th>
<th>Outcome test score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than college</td>
<td>2.35</td>
<td>48.00</td>
</tr>
<tr>
<td>Parent college</td>
<td>2.12</td>
<td>54.24</td>
</tr>
<tr>
<td>Single parent hh</td>
<td>2.37</td>
<td>48.76</td>
</tr>
<tr>
<td>Two-parent hh</td>
<td>2.18</td>
<td>52.37</td>
</tr>
<tr>
<td>Grade 3 score below median</td>
<td>2.61</td>
<td>45.35</td>
</tr>
<tr>
<td>Grade 3 score above median</td>
<td>1.82</td>
<td>57.96</td>
</tr>
</tbody>
</table>
4 Estimation and Identification

4.1 Estimation

We estimate the model using maximum likelihood, where parameter estimates maximize the probability of observing the joint endogenous outcomes, given the observed distributions of household characteristics across schools.

The parameters Θ to be estimated include model parameters Θ⁰ and parameters Θζ that govern the distribution of effort measurement errors. The former (Θ⁰) consists of the following seven groups: 1) Θ_y governing student achievement production function Y(·), 2) Θ_ε governing the distribution of shocks to test score ε, 29 3) Θ_c governing school effort cost, 4) Θ_{ε,p} governing parental effort cost, 5) Θ_D governing the cost of tracking regimes, 6) ω, the weight associated with proficiency in school’s objective function, 7) Θ_T governing the distribution Pr((a, z) | x, s) of household type given observables. 30

The endogenous outcomes observed for school s (O_s) include the tracking regime \( \tilde{\mu}_s \), track-specific school effort \( \{\tilde{e}_{sj}\}_j \) and household-level outcomes: parental effort \( \tilde{e}_{si}^p \), the track to which the student is assigned \( \tau_{si} \), and student final test score \( y_{si} \). Let \( X_s = \{x_{si}\}_i \) be the observed household characteristics in school s. The vector \( X_s \) enters the likelihood through its relationship with household types \( (a, z) \), which in turn affect all of \( O_s \).

The Likelihood The likelihood for school s is

\[
L_s(\Theta) = l_{\tilde{\mu}_s}(\Theta^0) \prod_j l_{\tilde{e}_{sj}}(\Theta \setminus \Theta_D) \prod_i l_{\tau_{si}}(\Theta_y, \Theta_\epsilon, \Theta_T, \Theta_{\epsilon,p}, \Theta_\zeta),
\]

where each part of the likelihood is as follows:

A. \( l_{\tilde{\mu}_s}(\Theta^0) \) is the probability of observing the tracking regime, which depends on all model parameters \( \Theta^0 \), since every part of \( \Theta^0 \) affects a school’s tracking decision, but not on \( \Theta_\zeta \). It is given by (3).

29 Test score shocks enter the model because schools care about test scores directly.
30 The distribution that enters the model directly, i.e., \( g_s(a, z) \), does not involve additional parameters, because

\[
g_s(a, z) = \int \Pr((a, z) | x, s) dF_s(x),
\]

where \( F_s(x) \) is the distribution of \( x \) in school s.
B. $l_{sj}(\Theta \backslash \Theta_D)$ is the contribution of the observed school effort $\tilde{e}_{sj}$ in track $j$ given the tracking regime $\tilde{\mu}_s$. It depends on all $\Theta^0$ but $\Theta_D$ since the latter does not affect school effort decision given the tracking regime. It also depends on $\Theta^\xi$ because school effort is measured with error:

$$l_{sj}(\Theta \backslash \Theta_D) = \frac{1}{\sigma_\xi} \phi \left( \frac{\tilde{e}_{sj} - e^*_j(\tilde{\mu}_s | X_s, \Theta^0 \backslash \Theta_D)}{\sigma_\xi} \right),$$

where $\phi$ denotes the standard normal density.

C. $l_{si}(\Theta_y, \Theta_\epsilon, \Theta_T, \Theta_{cp}, \Theta^\zeta)$ is the contribution of household $i$, which involves integrating type-specific contributions to the likelihood over the distribution of household types:

$$l_{si}(\Theta_y, \Theta_\epsilon, \Theta_T, \Theta_{cp}, \Theta^\zeta) = \sum_{a,z} \Pr((a, z) | x_i, s, \Theta_T) l_{si}((a, z) | \Theta_y, \Theta_\epsilon, \Theta_{cp}, \Theta^\zeta),$$

where $l_{si}((a, z) | \Theta_y, \Theta_\epsilon, \Theta_{cp}, \Theta^\zeta)$ is the contribution of household $i$ if it were type $(a, z)$:

$$l_{si}((a, z) | \Theta_y, \Theta_\epsilon, \Theta_{cp}, \Theta^\zeta) = \left[ \Pr\{track = \tau_{si} | a, \tilde{\mu}_s\} \times \Pr(\tilde{e}_{si} | e^{ps}(e^*_s | q_{rsi}, a, z | \Theta_y, \Theta_{cp}, \Theta^\zeta)) \times f_\epsilon \left( [(y_{si} - Y(a, q_{rsi}, e^*_s | \cdot | \Theta_y)] | \Theta_\epsilon \right) \right].$$

The three components of $l_{si}((a, z) | \cdot)$ are

a) the probability of being assigned to $\tau_{si}$ given tracking regime $\tilde{\mu}_s$ and ability $a$, which is implied by $\tilde{\mu}_s$ and $g_s(a)$;\textsuperscript{31}

b) the contribution of the observed parental effort $\tilde{e}_{si}^{ps}$ given peer quality and the model predicted school effort $e^*_s$ in track $\tau_{si}$, which depends on parental cost parameters, the achievement parameters and the measurement error parameters; and

c) the contribution of test score given all model predicted inputs, which depends on achievement parameters and the test score distribution parameter.

4.2 Identification

Manski (1993) identifies three components in social interactions models: endogenous

\textsuperscript{31}$\Pr\{track = j | a, \mu_s\} = \frac{\mu_{sj}(a) n_j}{\sum_a \mu_{sj}(a) n_j}$, where $n_j$ is the size of track $j$, i.e., $n_j = \sum_a \mu_{sj}(a) g_s(a)$.\textsuperscript{31}
effects (the direct effect of peer outcomes on one’s own outcome), exogenous contextual effects (the effect of exogenous peer characteristics on one’s own outcome), and correlated effects (shocks that are common to both one’s peers and oneself). Separating these channels has been the focus of much empirical work. The focus of this paper is to understand the interrelated nature of school tracking decisions, track input choices, and parental effort choices, not to push the frontier of identification of social interactions models. However, it does take these inferential problems seriously and addresses them by utilizing two arguments. First, similar to other work (Caucutt (2002), Epple and Romano (1998), Epple et al. (2002)), the mechanism through which tracking operates is through peer quality. Although peer quality and one’s own ability depend on the history of inputs, they are pre-determined at the beginning of the game. Therefore, as in this other work, the “reflection problem” of separating endogenous effects from exogenous contextual effects is not relevant here. Although endogenous social interactions do not pose an identification problem in our context, we must address the problem of separating exogenous contextual effects, which operate through unobserved peer quality, and correlated effects. In our context, this is the same as accounting for selection into peer groups (Moffitt (2001)).

Intuitively, if we do not account for the correlation between own ability and peer quality when students are tracked, what looks like the effect of peers may instead be correlated unobservables (i.e. own ability and peer quality are correlated), creating an inferential problem. To address this second concern, we exploit two key pieces of information available in the ECLS-K survey. The first is student prior test scores, assumed to be sufficient statistics of their ability. The distribution of student test scores in each school therefore maps into the school-specific ability distribution. The second piece of information comes from teacher surveys, which directly classify each student’s track quality as high, middle or low. This additional information, available for multiple classes, allows us to essentially “observe” two key components of model. The first is the tracking regime used by each school. The second is the composition of peer ability in each track. Each track comprises a section of the

\[32\] For example, Bramoullé et al. (2009) circumvents the reflection problem by exploring sample finiteness. See Betts and Shkolnik (2000a) for a review of other work.

\[33\] This also means we can estimate a linear-in-means technology without having to exploit non-linearities in peer effects, a strategy identified by Blume et al. (2006).

\[34\] Assuming we observe a representative sample of classes per school.
school-specific ability distribution; the section is specified by the tracking regime. Such information directly reveals peer quality, despite the fact that a student’s ability is not directly observable and that the grouping of students into different tracks is endogenous. Therefore, we are able to separate the roles of own ability and peer quality in the technology, a common identification concern in this literature.

5 Results

5.1 Parameters

In this section, we present estimates of key parameters. Other parameter estimates can be found in Appendix B. Table 6 presents the production technology parameter estimates, where the coefficient of student’s own ability has been set to one. We find that school effort and parental effort complement each other, while the interaction between a student’s own ability and school effort is (insignificantly) negative. To understand the magnitudes of parameters in Table 6, we calculate the outcome test scores according to the estimated production function by increasing one input at a time while keeping other inputs at model averages.\textsuperscript{35} The marginal effect of increasing ability by one standard deviation (sd) above the model average causes a 74% sd increase in the outcome test score. Increasing peer quality by one sd causes a 24% sd increase in the outcome test score. Increasing school effort by one sd causes a 2% sd increase in the outcome test score; while increasing parental effort by one sd increases the outcome test score by 44% sd. Table 7 presents estimates of school-

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard error</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha_0$</td>
<td>-61.81</td>
<td>5.58</td>
<td>intercept</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>6.17</td>
<td>3.20</td>
<td>school effort</td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>16.16</td>
<td>2.14</td>
<td>parent effort</td>
</tr>
<tr>
<td>$\alpha_3$</td>
<td>0.36</td>
<td>0.08</td>
<td>track peer quality</td>
</tr>
<tr>
<td>$\alpha_4$</td>
<td>-0.03</td>
<td>0.02</td>
<td>interaction: ability and school effort</td>
</tr>
<tr>
<td>$\alpha_5$</td>
<td>1.63</td>
<td>0.86</td>
<td>interaction: school and parent effort</td>
</tr>
</tbody>
</table>

side parameters. We find a low value of $\omega$, suggesting that schools do not care much about how many students are proficient, compared to their concerns about average

\textsuperscript{35}Notice these calculations are meant to illustrate the production technology; they do not take into account any behavioral responses.
achievement. This finding may be due to the fact that the test score we use is an ECLS-K survey instrument, not a high-stakes test. This is consistent with findings from the school accountability literature, which finds that pressure, such as No Child Left Behind, leads to large gains on high-stakes tests, but much smaller gains on low-stakes exams.\textsuperscript{36} The second row shows that the cost of school effort is convex in effort levels.\textsuperscript{37} The last three rows show that the cost of tracking regimes is non-linear in the number of tracks. The most costly tracking regimes are those with four tracks (the highest possible number), followed by those with only one track (the cost of which normalized to zero). Without further information, our model is unable to distinguish between various components of the cost associated with different tracking regimes. However, we think these estimates are not unreasonable. On the one hand, increasing the number of tracks may involve developing more types of curricula as well as incur higher resistance from parents. On the other hand, pooling all students into one track may make the classroom too heterogeneous and thus difficult for the teacher to handle. If these competing costs are both convex in the number of tracks, one would expect the total cost to be higher for the one- and four-track cases. Table 8 shows the parameters governing parental effort cost and the probability of being a high-cost parent. We find that the cost of parental effort is convex. The linear type-specific cost term is 10\% higher for the high-cost type than for the low-cost type. If both types exert the same level (sample average) of parental effort, the high-cost type incur a 3\% higher cost than the low-cost type. The last four rows show the relationship between household characteristics with parental cost types. There is a strong relationship between child ability and the probability that the parent is a high-cost type. Higher-educated parents also tend to have higher effort cost. We do

\begin{table}[h]
\centering
\begin{tabular}{lll}
\hline
Parameter & Estimate & Standard error \\
\hline
$\omega$ & 0.93 & 0.91 weight on passing proficiency standard \\
$c_s^2$ & 1.71 & 0.47 quadratic school effort cost \\
$\gamma_2$ & -1.21 & 0.39 regime cost, 2 tracks \\
$\gamma_3$ & -0.91 & 0.42 regime cost, 3 tracks \\
$\gamma_4$ & 0.30 & 0.52 regime cost, 4 tracks \\
\hline
\end{tabular}
\end{table}

\textsuperscript{36}See, for example, Koretz and Barron (1998), Linn (2000), Klein et al. (2000), Carnoy and Loeb (2002), Hanushek and Raymond (2005), Jacob (2005), Wong et al. (2009), Dee and Jacob (2011), Reback et al. (2014).

\textsuperscript{37}We cannot reject that the linear cost term $c_s^1$ is zero, so we set it to zero.
not find that, conditional on their own education level and their children’s ability, single parents are more likely to have higher effort costs.\textsuperscript{38}

Based on type distribution parameters, Table 9 shows the simulated distribution of parental types over all households and by parental characteristics in our sample. Around 39% of all parents are of the high cost type. This fraction is higher among college-educated parents (44.8%) and among single parents (43.1%). On average, children of low-cost parents have lower prior test scores with a mean of 47.7 and a sd of 9, compared to those of high-cost parents (mean of 57.5 and sd of 6.9). This is consistent with the data in Table 5, where parents without college education and parents of children with lower prior achievement exert more parental effort, yet average student outcome test scores are lower in these households.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c^p$</td>
<td>0.09</td>
<td>0.01</td>
</tr>
<tr>
<td>$z_1$</td>
<td>0.07</td>
<td>0.12</td>
</tr>
<tr>
<td>$z_2$</td>
<td>0.08</td>
<td>0.12</td>
</tr>
<tr>
<td>$\theta_0$</td>
<td>-42.33</td>
<td>4.97</td>
</tr>
<tr>
<td>$\theta_1$</td>
<td>0.76</td>
<td>0.10</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>-0.46</td>
<td>7.32</td>
</tr>
<tr>
<td>$\theta_3$</td>
<td>3.56</td>
<td>11.28</td>
</tr>
</tbody>
</table>

Table 9: Parental Cost Type Distribution

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Edu$\geq$College</th>
<th>Single Parent</th>
<th>Prior Test Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Cost</td>
<td>61.3%</td>
<td>55.2%</td>
<td>56.9%</td>
<td>47.7 (9.0)</td>
</tr>
<tr>
<td>High Cost</td>
<td>38.7%</td>
<td>44.8%</td>
<td>43.1%</td>
<td>57.5 (6.9)</td>
</tr>
</tbody>
</table>

5.2 Model Fit

Overall, the model fits the data well. Table 10 shows model fit in terms of tracking patterns. The first two columns show that the model closely matches the distribution of tracking regimes across all schools. The next two columns show the fit for schools

\textsuperscript{38}Notice that parents’ own characteristics are not significantly correlated with their effort cost types. Given that student ability in our context measures human capital obtained by 5th grade, which is a cumulative result from their parents’ investment, it is not surprising that child ability contains dominant information about the type of one’s parent.
with lower spread of prior test scores, where a school is called a low-spread school if it has a coefficient of variation (CV) of prior scores below the median CV. The model slightly overpredicts the fraction of two-track schools and underpredicts that of three-track schools. The next two columns focus on schools with higher-achieving students. In particular, we rank schools by the fraction of lower-prior-score (below median score) students from high to low: the higher the ranking of a school, the higher fraction of its students are of low initial achievement. We report the tracking regime distribution among schools that are ranked below the median in this ranking, i.e., schools with relatively better students. Overall, the model captures the pattern that schools with less variation and/or higher prior test scores are more likely to have only one track and less likely to have four tracks. Table 11 shows that the model fits average outcome scores by track. The exception is that the model over-predicts the average score in the second track within two-track schools and that in the third track within four-track schools. Figure 1 shows the model predicted CDF of outcome test scores contrasted with the data counterpart. The left panel shows the case for all students. The right panel splits the distribution by parental education and by single-parenthood. As seen, the model-predicted score distributions match well with the ones in the data. Table 12 shows the model fit for school effort. Compared to the data, the model predicts a flatter profile for school effort across tracks. Table 13 shows the model fit for mean parental effort: the model matches the fact that parent effort decreases with track level, although the gradient is over-predicted in the four-track case (last two columns). Finally, Table 14 shows that the model fits well the level of parental effort and outcome scores by household characteristics.

Table 10: Tracking Regimes

<table>
<thead>
<tr>
<th></th>
<th>All schools</th>
<th>Low Spread*</th>
<th>Low fraction of low ability**</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Data Model</td>
<td>Data Model</td>
<td>Data Model</td>
</tr>
<tr>
<td>% 1 track</td>
<td>4.39 4.46</td>
<td>4.85 4.76</td>
<td>4.55</td>
</tr>
<tr>
<td>% 2 tracks</td>
<td>37.07 36.63</td>
<td>36.89 38.37</td>
<td>38.91</td>
</tr>
<tr>
<td>% 3 tracks</td>
<td>45.85 46.46</td>
<td>47.57 45.54</td>
<td>47.27</td>
</tr>
<tr>
<td>% 4 tracks</td>
<td>12.68 12.46</td>
<td>10.68 11.33</td>
<td>7.27</td>
</tr>
</tbody>
</table>

* “Low spread” schools have a below-median coefficient of variation in prior score.

** “Low fraction of low ability” schools have a below-median fraction of schools with below-median prior score.
Parents with less education, single parents, and students with lower prior score all have higher parent effort levels and lower outcome test scores.

Table 11: Outcome test score by track and number of tracks

<table>
<thead>
<tr>
<th>Track</th>
<th>1 Track</th>
<th>2 Tracks</th>
<th>3 Tracks</th>
<th>4 Tracks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Data</td>
<td>Model</td>
<td>Data</td>
<td>Model</td>
</tr>
<tr>
<td>1</td>
<td>51.84</td>
<td>50.72</td>
<td>45.95</td>
<td>46.86</td>
</tr>
<tr>
<td>2</td>
<td>51.98</td>
<td>54.85</td>
<td>51.38</td>
<td>51.38</td>
</tr>
<tr>
<td>3</td>
<td>55.62</td>
<td>56.64</td>
<td>51.45</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 12: Mean school effort by track and number of tracks

<table>
<thead>
<tr>
<th>Track</th>
<th>1 Track</th>
<th>2 Tracks</th>
<th>3 Tracks</th>
<th>4 Tracks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Data</td>
<td>Model</td>
<td>Data</td>
<td>Model</td>
</tr>
<tr>
<td>1</td>
<td>1.86</td>
<td>1.86</td>
<td>1.75</td>
<td>1.85</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>1.90</td>
<td>1.85</td>
<td>1.88</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>1.96</td>
<td>1.84</td>
<td>1.93</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 13: Mean parent effort by track and number of tracks

<table>
<thead>
<tr>
<th>Track</th>
<th>1 Track</th>
<th>2 Tracks</th>
<th>3 Tracks</th>
<th>4 Tracks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Data</td>
<td>Model</td>
<td>Data</td>
<td>Model</td>
</tr>
<tr>
<td>1</td>
<td>2.07</td>
<td>2.21</td>
<td>2.31</td>
<td>2.49</td>
</tr>
<tr>
<td>2</td>
<td>2.03</td>
<td>1.97</td>
<td>2.37</td>
<td>2.23</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>2.11</td>
<td>1.91</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>2.08</td>
</tr>
</tbody>
</table>

Table 14: Means of parent effort and outcome score, by household characteristics

<table>
<thead>
<tr>
<th>Parent effort</th>
<th>Outcome score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Data</td>
</tr>
<tr>
<td>Less than college</td>
<td>2.35</td>
</tr>
<tr>
<td>College</td>
<td>2.12</td>
</tr>
<tr>
<td>Single parent</td>
<td>2.37</td>
</tr>
<tr>
<td>Two parents</td>
<td>2.18</td>
</tr>
<tr>
<td>Low prior score</td>
<td>2.52</td>
</tr>
<tr>
<td>High prior score</td>
<td>1.78</td>
</tr>
</tbody>
</table>

6 Counterfactual Simulations

We use the estimated model to simulate two policy-relevant counterfactual scenarios. We contrast the outcomes between the baseline and each of the counterfactual cases. In particular, we present Average Treatment Effects (ATE) for different endogenous outcomes (some of which are inputs to achievement), for subgroups of students defined by their characteristics, such as prior test scores.\(^{39}\)

In the first counterfactual simulation, we quantify the effect of allowing tracking by solving the model when tracking is banned (hence all schools have only one track). We compare the changes in school effort, parental effort and student achievement. Our results indicate that failing to account for the equilibrium interactions between schools and parents could substantially bias the results. In the second counterfactual simulation, we examine the equilibrium effects of prospective changes in proficiency standards. In particular, we solve for optimal region-specific proficiency standards.

\(^{39}\)When a continuous variable is used to define subgroups, as they are in Figures 3 and 4(a), ATEs are calculated by non-parametric smoothing.
that would maximize average achievement. Unlike the tracking-ban counterfactual, here a school re-optimizes its tracking decision in response to this policy change.  

6.1 Heterogeneous Effects of Tracking

We compare the outcomes under the baseline with those when tracking is banned and every school has to put all of its students into a single track. Over 95% of the simulated schools practice ability tracking to some degree under the baseline, hence are affected by this counterfactual and experience an exogenously imposed change in peer quality within classrooms.

The first two panels in the top row of Figure 2 show average results for outcome test scores and pass rates, by decile of prior test score. The effects of a tracking ban are positive for students with lower prior test scores and negative for those with higher prior test scores, as measured in both the level of the final test scores and the pass rates. In particular, students with below-median prior scores gain 3.6% sd when ability tracking is banned, while those with above-median prior scores lose 5% sd when tracking is banned. Consistent with the first two panels, the third panel shows that the fraction of students who gain from a tracking ban declines with prior test scores.

Underlying the changes in student outcomes are the changes in the inputs, i.e., peer quality, school effort and parental effort, which are plotted in the bottom three panels of Figure 2, by decile of prior test scores. The tracking ban places all students in a school in one track, which means that lower ability students are on average placed with better students, and are made better off through the technology, ceteris paribus. The opposite holds for higher ability students. However, that is not the entire story. Both school effort and parental effort adjust to the change in peer composition imposed by this policy. Without the freedom to optimize

\[40\] Results from our counterfactual experiments are subject to the caveat that household distribution across schools are fixed. Incorporating households’ school choices into our framework is an important extension we leave for the future work.

\[41\] The ATE of banning tracking over all students is -0.04 points, or about -0.4% sd in test scores. Our result that tracking (in the short run) hurts lower-ability students and benefits higher-ability students is consistent with findings from some previous studies, e.g. Betts and Shkolnik (2000b) and Hoffer (1992).

\[42\] There are possible scenarios where the omission of cross-school sorting might lead us to over-predict the effect on achievement of a ban on tracking. For example, given that this policy lowers peer quality for high-ability students in schools that used to track students in the baseline, it is possible that some of these households may re-sort themselves to schools with more homogeneous peers.
Figure 2: Change in outcomes and inputs due to banning tracking, by decile prior score

over tracking regimes, schools that used to track students can only optimize over their effort inputs. As there is only one track, a school can only choose one effort level for all students. On average, schools increase their effort for students in all deciles (the second panel on the bottom). This change is most obvious for students with very high or very low prior test scores, who are more likely to have been tracked under the baseline. Unlike schools, which choose one effort level for all students in one track, parents can always adjust their effort levels for their own children. Indeed, the last panel shows that changes in parental effort are not only quantitatively but also qualitatively different across students with different prior test scores. Average parental effort decreases for students below the median and increases for those above the median. In particular, parents of students in the highest decile increase their inputs by the largest amount when tracking is banned, by about 2% sd.

Figure 2 highlights the trade-offs a school faces when choosing a tracking regime. Improving peer quality in one track necessarily involves reducing it in another, which will in turn lead to changes in parental effort. When low-ability students are grouped with higher-ability students, peer quality increases (decreases) for students with low (high) ability. For parents with low-ability students, who have been exerting much
higher effort than those with high-ability students (Table 14), the exogenous increase in peer quality provides strong incentives for them to reduce their own effort. For parents with high-ability students, the exogenous decrease in peer quality pushes them to increase their own effort as a remedy. However, given the concavity of the parents’ net payoff with respect to their own effort, the reduction in effort by the parents of low ability children is larger than the increase in effort by the parents of high ability children, especially among parents with very low-achieving children. To curb the reduction in parental effort among low-achieving households and encourage more effort in high-achieving households, the school increases its own effort, utilizing the fact that school effort and parental efforts are complementary to each other. Depending on the different sets of households they face, schools differ in how much effort they need to exert and how their students perform when they do not track versus when they track. These differences drive schools’ different tracking decisions.

6.1.1 The Importance of Accounting for Behavioral Changes: Effort Inputs

The test score technology plays an important role in evaluating the effect of tracking on student outcomes, which may partly explain why much of the literature studying tracking and peer quality focuses on estimating it. This might prompt one to ask whether estimates of parameters governing the technology alone would adequately characterize outcomes for students were tracking banned. To illustrate the value of estimating an equilibrium model where schools and parents may respond to changes in track peer quality, we contrast the ATE of banning tracking from our model prediction with the ATE ignoring endogenous effort responses. Let \((q^*, e^{ss}, e^{pe})\) denote inputs under the baseline scenario where tracking is endogenous, and \((q^*_{CF}, e^{ss}_{CF}, e^{pe}_{CF})\) denote counterfactual equilibrium inputs when tracking is banned.

Figure 3 graphs the ATE for test scores (y-axis) against student prior test scores (x-axis). The red (solid) line is the model-predicted ATE, taking into account the change in peer quality induced by a tracking ban, as well as school and parent effort responses, i.e. \(Y(a, q^*_{CF}, e^{ss}_{CF}, e^{pe}_{CF}) - Y(a, q^*, e^{ss}, e^{pe})\) (recall \(Y(\cdot)\) is the test score technology). The blue (short dashed) line is the ATE, ignoring school effort adjustments, i.e. \(Y(a, q^*_{CF}, e^{ss}_{CF}, e^{pe}_{CF}) - Y(a, q^*, e^{ss}, e^{pe}_{CF})\). Banning tracking pushes schools
Figure 3: ATE of banning tracking: equilibrium vs fixed effort inputs

...to increase their effort; ignoring this biases the effect of tracking ban downwards, although not by much. The bias from ignoring parental responses is much larger. The brown (long dashed) line is the ATE for test scores ignoring parental effort adjustment, i.e. $Y(a, q^{\ast}_{CF}, e^{ss}, e^{P}) - Y(a, q^{\ast}, e^{ss}, e^{P})$. When tracking is banned, lower-achieving students receive more inputs from the school, in terms of both peer quality and school effort. In response, parents of these students reduce their own effort. Failing to take into account this reduction drastically overstates the ATE of banning tracking for these students. The brown line lies far above the red line for students with the lower prior scores, especially those at the end of the distribution. The opposite is true for students with higher prior scores, whose parents increase their provision of costly effort in response to the lower peer quality. The black (dotted-dashed) line is the ATE for test scores ignoring both school and parent effort adjustments, i.e. $Y(a, q^{\ast}_{CF}, e^{ss}, e^{P}) - Y(a, q^{\ast}, e^{ss}, e^{P})$. On average, ignoring effort changes would cause one to overstate the gains from banning tracking by over 100% for students with below-median prior scores, and overstate the loss by 57% for students with above-median prior scores.
6.2 Optimal Proficiency Standards: Trade-offs

Policy changes will often create winners and losers. This is especially true in the case of educational policies that may affect ability tracking, which has qualitatively different effects on students of varying ability levels. This is because changing peer quality for some students necessarily involves changing peer quality for some other students, which is accompanied by adjustment in the effort choices by the school and by parents. By incorporating tracking regimes, school effort, and parental effort inputs into one framework, our work lends itself to a better understanding of the trade-offs educational policies can generate. Our second counterfactual experiment highlights this point by examining how changes in proficiency standards would affect the distribution of student achievement. Because schools care about the fraction of students above the proficiency standard, changes in these standards can guide the new equilibrium toward certain policy goals.\footnote{As mentioned earlier in the paper, our estimate of a school’s preference for the lower-tail of the score distribution does not reflect the pressure it faces from high-stakes tests; thus, this experiment should not be interpreted as increasing the bar for a high-stake test.} As an illustration, we search for region-specific proficiency standards that maximize the region-specific averages of student test scores.\footnote{This experiment is an illustration of the trade-offs. One could also use our framework to study the effects of other policy changes, which could generate different distributions of winners and losers.}

Figure 4 places proficiency standards in relation to the distribution of baseline outcome test scores, by region, where the left panel (4(a)) shows the density and the right panel (4(b)) shows the CDF. Each panel overlays the distribution of baseline outcome scores with the baseline proficiency standards (blue dotted line). Loosely speaking, the ranking of the regional distributions of student baseline achievement from high to low (in the sense of first order stochastic dominance) is the Northeast, the Midwest, the West, and lastly the South. This ranking lines up with that of regional proficiency standards, with the Northeast having the highest standard and the South having the lowest. In all four regions, the baseline (data) standard is approximately located at the 30th percentile of the regional distribution of the outcome scores.

The regional standards that maximize region-specific average achievement are the red solid lines in Figure 4, which are higher in all four regions than the baseline standards. When standards are lower (as in the baseline), schools have the incentive
to improve outcomes near the lower end of the distribution, which would sacrifice a much larger measure of students near the middle and top of the distribution of baseline scores. The same argument applies for the case if standards are too high. The new standards maximize the average performance by moving schools’ attention away from the low-performing students toward a location that rewards schools for improving mean test scores. In fact, the new standards are located at around the median of each region’s baseline distribution of test scores, which is also where the density of outcome scores is highest, as seen in Figure 4(a). Figure 5(a) summarizes the ATE on outcome scores by region due to the change in standards. Setting standards to maximize average achievement has the biggest ATE in the South, followed by the West, the Midwest and finally the Northeast, which is in the reverse order as the baseline regional proficiency standards. Intuitively, the ATE is increasing in the difference between the baseline and achievement-maximizing standards.

Figure 5(b) plots the ATE on the outcome score and inputs (school effort, peer quality, and parental effort) by baseline score and region. The top panel shows that in all four regions, the ATE on outcome test scores increases with student baseline outcome scores. In fact, the ATE is negative for students with low baseline scores and positive for students with high baseline scores. Underlying the pattern
of the ATE on outcome test scores is schools’ redistribution of their inputs away from low-achieving students toward high-achieving students, as shown in the second and third panels of Figure 5(b). Schools decrease (increase) their effort inputs for students with lower (higher) baseline scores. In addition, schools also change their tracking regimes such that peer quality decreases (increases) for low-achieving (high-achieving) students. Finally, the bottom panel shows that the ATE on parental effort moves in the opposite direction to school inputs, which mitigates but does not reverse the effects of school input adjustment on student outcomes.

7 Conclusion

We have developed and estimated a model of ability tracking, in which a school’s tracking regime, track-specific inputs, parental effort and student achievement are joint equilibrium outcomes. The estimated model has been shown to fit the data well.

Using the estimated model, we have shown that the effects of tracking are hetero-
geneous across students with different prior test scores. In response to the exogenous changes in peer composition under a ban on tracking, schools on average increase their effort inputs for all students. Parents with low-ability students decrease their effort, while those with high-ability students increase theirs. As a result, a ban on tracking would increase the performance for students with below-median prior scores by 3.6% sd, while decrease that for those with above-median prior scores by 5% sd.

These seemingly-small effects confound the change in peer quality and that in effort inputs, both of which have significant impacts on achievement. In fact, ignoring endogenous effort changes would cause one to overstate the gains from banning tracking by over 100% for students with below median prior scores, and overstate the loss by 57% for students with above median prior scores. Our work, therefore, re-emphasizes the importance of taking into account behavioral responses when evaluating policy changes, as argued in Becker and Tomes (1976). It is worth noting that this general principle applies regardless of whether the estimation is done on observational data or experimental data.45

Because policies that affect schools’ tracking decisions will necessarily lead to increases of peer quality for some students and decreases for some other students, our study highlights the trade-offs involved in achieving certain educational policy goals. In particular, we have shown that a change in proficiency standards that maximizes average achievement would lead schools to redistribute inputs away from low-ability students toward high-ability ones, in terms of both peer quality and track-specific effort. As a result, the performance of high-ability students increases at the cost of low-ability students.

Our work is promising for future research. Researchers have been expanding the depth and width of datasets, so it is not overly optimistic to think that a dataset containing similar information as the ECLS-K, but at a larger scale per school, will be available in the near future. Our methodology can be easily applied to such a dataset, and, due to its tractability, would remain computationally feasible even when the dataset is large. Moreover, our work also admits potentially interesting extensions for future research. One extension of particular interest would combine studies on the matching between schools and households and this paper into a single

45Carrell et al. (2013) provides a good example of this principle in the context of experimental data.
framework. This extension would form a more comprehensive view of how peer composition is determined both between and within schools. Another important extension would allow teachers to vary in quality and examine how schools match students with teachers, within a tracking framework.

Appendix

A   Functional Forms

A.1   Type Distribution

Denote observable characteristics \( x = (x^a, x^p) \), where \( x^a \) is the prior test score and \( x^p \) includes parent education level and whether or not it is a single-parent household.

Each school has three ability levels \( (a^s_l, l = 1, 2, 3) \). Let \( T^s_l \) be the \( l \)th tercile of \( F_{s,a}(\cdot) \), which is the normal distribution approximation of prior test scores of all students in school \( s \), \( \{x^a_i \}_i \). A level \( a^s_l \) is defined as the expectation of prior score within the \( l \)th tercile in school \( s \) computed using \( F_{s,a}(\cdot) \), i.e.,

\[
\begin{align*}
    a^s_1 &= \int_{-\infty}^{T^s_1} a dF_{s,a}(a | a \leq T^s_1), \\
    a^s_2 &= \int_{T^s_1}^{T^s_2} a dF_{s,a}(a | T^s_1 < a \leq T^s_2), \\
    a^s_3 &= \int_{T^s_2}^{\infty} a dF_{s,a}(a | T^s_2 < a).
\end{align*}
\]

The distribution of type conditional on \( x \) is assumed to take the form

\[
\Pr ((a^s_l, z) | x, s) = \Pr (a = a^s_l | x^a, s) \Pr (z | x^p, a^s_l),
\]

\[
\begin{align*}
    \Pr (a = a^s_1 | x^a, s) &= 1 - \Phi \left( \frac{x^a - T^s_1}{\sigma_a} \right), \\
    \Pr (a = a^s_2 | x^a, s) &= \Phi \left( \frac{x^a - T^s_2}{\sigma_a} \right), \\
    \Pr (a = a^s_3 | x^a, s) &= 1 - \Pr (a = a^s_1 | x^a) - \Pr (a = a^s_3 | x^a),
\end{align*}
\]
where $\sigma_a$ is a parameter to be estimated. The probability that a parent is of a high-cost type is given by

$$\Pr(z = z_2|x^p, a^*_i) = \Phi(\theta_0 + \theta_1 a^*_i + \theta_2 I(x^p_1 \geq \text{college}) + \theta_3 I(x^p_2 = \text{single parent})).$$  

(6)

A.2 Effort measurement system

A.2.1 Parental effort

Observed parental effort $\tilde{e}_p$ is discrete and ordered in 5 categories, as in Question HEQ.095 in the Spring 6 Parent Questionnaire.

“During this school year, how often did someone help CHILD with his/her reading, language arts or spelling homework? Would you say…

A. Never; B. Less than once a week; C. 1 to 2 times a week; D. 3 to 4 times a week; E. 5 or more times a week; F. REFUSED; G. DON’T KNOW.”

Model parental effort is unobserved, but the measurement system below maps $e^{p*}$ and an i.i.d. error term $\zeta^p \sim N(0, 1)$ into discrete ordered levels:

$$\tilde{e}_p = \begin{cases} 
\text{never} & \iff -\infty < e^{p*} + \zeta^p \leq \kappa_0 \\
< 1 & \iff \kappa_0 < e^{p*} + \zeta^p \leq \kappa_1 \\
\in [1, 2] & \iff \kappa_1 < e^{p*} + \zeta^p \leq \kappa_2 \\
\in [3, 4] & \iff \kappa_2 < e^{p*} + \zeta^p \leq \kappa_3 \\
\geq 5 & \iff \kappa_3 < e^{p*} + \zeta^p \leq \infty,
\end{cases}$$  

(7)

resulting in $\Pr(\tilde{e}_p|e^{p*})$.

A.2.2 School effort

School effort is measured in hours per week, according to Question 2 in the Spring 6 Teacher Questionnaire.

“For subjects you teach, about how much time do you expect children to spend on homework in each of the following area (Reading and Language Arts) on a typical evening?

A. I don’t teach this subject; B. None; C. 10 min; D. 20 min; E. 30 min; F. More
than 30 min.”

The observed effort is given by \( \tilde{e}_j^s = e^{**}_j + \zeta_j^s \), where \( \zeta_j^s \sim N(0, \sigma^2_z) \).

**B Additional Tables**

Table 15: Data: Proficiency cutoffs by Census region

<table>
<thead>
<tr>
<th>Region name</th>
<th>Proficiency cutoff</th>
<th>Corresponding sample percentile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northeast</td>
<td>49.69</td>
<td>42.34</td>
</tr>
<tr>
<td>Midwest</td>
<td>48.85</td>
<td>34.21</td>
</tr>
<tr>
<td>South</td>
<td>43.61</td>
<td>21.41</td>
</tr>
<tr>
<td>West</td>
<td>45.69</td>
<td>26.25</td>
</tr>
</tbody>
</table>

Table 16: Data: Prior Test Scores By Track

<table>
<thead>
<tr>
<th>Track</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>48.28</td>
<td>11.02</td>
<td>886</td>
</tr>
<tr>
<td>2</td>
<td>52.28</td>
<td>8.44</td>
<td>1,227</td>
</tr>
<tr>
<td>3</td>
<td>53.99</td>
<td>8.27</td>
<td>563</td>
</tr>
<tr>
<td>4</td>
<td>54.99</td>
<td>6.54</td>
<td>113</td>
</tr>
</tbody>
</table>

Table 17: Data: Standardized Prior Test Scores By Track

<table>
<thead>
<tr>
<th>Track</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.93</td>
<td>0.18</td>
<td>886</td>
</tr>
<tr>
<td>2</td>
<td>1.02</td>
<td>0.15</td>
<td>1,227</td>
</tr>
<tr>
<td>3</td>
<td>1.05</td>
<td>0.14</td>
<td>563</td>
</tr>
<tr>
<td>4</td>
<td>1.09</td>
<td>0.12</td>
<td>113</td>
</tr>
</tbody>
</table>

Note: Prior scores are standardized through division by school average prior score

---

46 We treat both E and F as 30 minutes per day.
### Table 18: Other Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Standard error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_a$</td>
<td>9.08</td>
<td>0.41</td>
</tr>
<tr>
<td>$\sigma_c$</td>
<td>7.31</td>
<td>0.13</td>
</tr>
<tr>
<td>$\sigma_\zeta$</td>
<td>0.55</td>
<td>0.02</td>
</tr>
<tr>
<td>$\kappa_0$</td>
<td>-0.03</td>
<td>0.61</td>
</tr>
<tr>
<td>$\kappa_1$</td>
<td>0.91</td>
<td>0.60</td>
</tr>
<tr>
<td>$\kappa_2$</td>
<td>2.11</td>
<td>0.59</td>
</tr>
<tr>
<td>$\kappa_3$</td>
<td>2.90</td>
<td>0.65</td>
</tr>
</tbody>
</table>

- sd of shock in ability distribution
- sd test score measurement error
- sd school effort measurement error
- cut-point 1, parent effort measurement
- cut-point 2, parent effort measurement
- cut-point 3, parent effort measurement
- cut-point 4, parent effort measurement

### Table 19: Summary Statistics: Full Sample vs Selected Sample (5th Grade)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Whole Sample</th>
<th></th>
<th></th>
<th>Selected Sample</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
<td>Standard Deviation</td>
<td>N</td>
<td>Mean</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>Sample Size</td>
<td>8,853</td>
<td>2,789</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parental Education</td>
<td>8,047</td>
<td>0.55</td>
<td>0.50</td>
<td>2,789</td>
<td>0.59</td>
<td>0.49</td>
</tr>
<tr>
<td>Single Parent</td>
<td>7,953</td>
<td>0.23</td>
<td>0.42</td>
<td>2,789</td>
<td>0.19</td>
<td>0.39</td>
</tr>
<tr>
<td>Prior Test Score</td>
<td>8,562</td>
<td>50.17</td>
<td>9.72</td>
<td>2,789</td>
<td>51.47</td>
<td>9.52</td>
</tr>
<tr>
<td>Outcome Test Score</td>
<td>8,751</td>
<td>50.15</td>
<td>9.65</td>
<td>2,777</td>
<td>51.68</td>
<td>9.40</td>
</tr>
<tr>
<td>Parental Effort</td>
<td>7,788</td>
<td>2.29</td>
<td>1.55</td>
<td>2,703</td>
<td>2.22</td>
<td>1.53</td>
</tr>
<tr>
<td>Teacher Effort</td>
<td>2,784</td>
<td>2.03</td>
<td>0.58</td>
<td>533</td>
<td>1.90</td>
<td>0.59</td>
</tr>
<tr>
<td>Student Obs. in the School</td>
<td>1,772</td>
<td>5.00</td>
<td>5.54</td>
<td>205</td>
<td>13.61</td>
<td>3.38</td>
</tr>
</tbody>
</table>

Note: “Whole Sample”: all fifth graders observed in the ECLS-K.
“Selected Sample”: the sample used for our empirical analyses.
Table 20: Summary Statistics: Whole Sample vs Stayers (Kindergarten)

<table>
<thead>
<tr>
<th>Variable</th>
<th>All Kindergarteners(^1a)</th>
<th>Stayers(^1b)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>Sample Size</td>
<td>16,665</td>
<td>8,494</td>
</tr>
<tr>
<td>Parental Education</td>
<td>15,705</td>
<td>0.49</td>
</tr>
<tr>
<td>Single Parent</td>
<td>14,329</td>
<td>0.25</td>
</tr>
<tr>
<td>Prior Test Score</td>
<td>13,734</td>
<td>49.19</td>
</tr>
<tr>
<td>Outcome Test Score</td>
<td>14,579</td>
<td>49.56</td>
</tr>
</tbody>
</table>

Note 1a: All kindergarteners observed in the ECLS-K.
Note 1b: Kindergarteners who are observed in the fifth grade whole sample.
Note 2: The stats are measured in the first survey round, where parental and teacher effort are both unreported.

C Nonlinear Peer Effects

There is recent interest in non-linear peer effects, where the marginal effect of a change in peer quality is not uniform across students (Sacerdote (2011)). It is important to note that, even in our preferred specification which excludes an interaction term between own ability and peer quality in the technology, the reduced form mapping between peer quality and achievement is inherently non-linear (i.e. depends on own ability), because we endogenize school and parental effort decisions.

To see this, consider a student with ability \(a_i\) in a track with peer quality \(q_j\). To derive the reduced-form achievement function, substitute her parents’ optimal effort \(e_p^*\) and optimal track effort \(e_s^*\) into (4), and differentiate with respect to peer quality:

\[
\frac{\partial Y}{\partial q} = \alpha_3 + \alpha_1 \frac{\partial e_j^*}{\partial q} + \alpha_2 \frac{\partial e_i^*}{\partial q} + \alpha_5 \left( e_i^* \frac{\partial e_j^*}{\partial q} + e_j^* \frac{\partial e_i^*}{\partial q} \right).
\]

To ease exposition, suppose \(\frac{\partial e_j^*}{\partial q} = 0\), which results in the expression

\[
\frac{\partial Y}{\partial q} = \alpha_3 + \frac{\partial e_i^*}{\partial q} \left( \alpha_2 + \alpha_5 e_j^* \right).
\]

Due to curvature of the parents’ objective, parents will have different responses to changes in peer quality depending on their children’s ability levels, i.e. \(\frac{\partial^2 e_i^*}{\partial q \partial a} \neq 0\).
This implies that $\frac{\partial^2 Y}{\partial q \partial a} \neq 0$, which is what we set out to show.

D Caveats Arising From Sample Attrition

Although the attrition is largely random, as shown in Table 20, the slight non-randomness leads us to make the following caveats. Case 1: If the non-random attrition reflects the sorting of students across different schools, our sample will represent (hence our results will apply to) only a subset of schools in the U.S.: those with students from relatively better family backgrounds. In this case, our parameter estimates will not be biased. Case 2: If the attrition reflects higher frequency of school switches among a subgroup of students, and if the out-flow of such students from one school is associated with an in-flow of similar students into the same school, then students with better family backgrounds will be over-represented in our sample within a school. This may bias our parameter estimates. Given that the sorting of students into different schools has been widely documented in the literature, we feel that concerns arising from Case 1 may be more relevant, i.e., our results may not apply to all schools in the U.S., but a subset of them.

References


47See, for example, Caucutt (2002) and Epple and Romano (1998).


44


Loveless, T., “The Resurgence of Ability Grouping and Persistence of Tracking,” 
Part II of the 2013 Brown Center Report on American Education, the Brookings Institution.[2], 2013. 2

The Review of Economic Studies, 60(3):531–542, 1993. 6, 17, 21


Moffitt, R., “Policy Interventions, Low-Level Equilibria, and Social Interactions,” 
Social Dynamics, 4:45–82, 2001. 6, 17, 22


Wong, M., T. D. Cook and P. M. Steiner, “No Child Left Behind: An interim evaluation of its effects on learning using two interrupted time series each with its own non-equivalent comparison series,” Institute for Policy Research (Working paper 09-11), 2009. 24