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

Chao Fu † Nirav Mehta ‡

September 23, 2014

Abstract
We develop and estimate an equilibrium model of ability tracking. In the model, a school chooses how to allocate students into tracks based on their ability and chooses track-specific inputs. Parents choose parental effort in response. We estimate the model using data from the ECLS-K. We use the estimated model to first examine the effects of disallowing tracking on school and parental inputs and student achievement. We then examine how policies that change proficiency standards affect equilibrium tracking, school inputs, parental effort, and student achievement. We find that behavioral responses of parents to changes in school inputs would significantly attenuate the net impact of banning tracking for both low and high ability students.

1 Introduction

Ability tracking, the practice of allocating students into different classrooms based on prior performance, is pervasive; e.g., Loveless (2013) reports that 75% of 8th-graders in

*We thank Yuseob Lee, George Orlov, and Atsuko Tanaka for excellent research assistance. We thank John Bound, Betsy Caucutt, Tim Conley, Steven Durlauf, Lance Lochner, Jesse Rothstein, Jeffrey Smith, Steven Stern, and Chris Taber for insightful discussions. We benefit from the comments from participants at the AEA 2014 winter meeting and Labo(u)r Day, and seminars at Amherst, ASU, Berkeley, Brock, Queen’s, Rochester, 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
the National Assessment of Educational Progress data were in tracked classes. However, ability tracking is controversial because it may benefit students of certain ability levels while hurting the 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 student outcomes.

Several complications need to be addressed to answer these questions. First, changing peer composition in one classroom necessarily involves student re-allocation, hence changes peers in some other classroom(s). As such, it is important not to treat classrooms in isolation when studying the treatment effect of changing peers. Second, one needs to understand how school and parental inputs are chosen given a tracking regime to infer what these input levels and student achievement would be if tracking regimes, hence peer compositions, were changed. Finally, knowledge about how schools choose tracking regimes is necessary 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 an ideal world, carefully designed random experiments can conquer these complications and allow one to learn about the effects of tracking and the effects of policies that might affect schools’ tracking decisions. However, in practice, it is infeasible to run randomized control trials for every set of school characteristics (including student composition), and for every potential policy scenario. 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.

In the model, children from different types of households are educated in one school. 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 invested by the school and by her parent, and the quality of her peers.1 A parent maximizes her child’s achievement by choosing costly parental effort in response to 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 and track-specific effort inputs to maximize its own objective, which increases in the total achievement of its students and the fraction of students satisfying a proficiency requirement. Our framework naturally allows policies to produce winners and losers

---

1See Epple and Romano (2011) for a recent review on the literature of peer effects in education.
due to the differential impacts that tracking regimes may have on students of different ability levels and parental backgrounds.

We estimate our model via maximum likelihood using data from the Early Childhood Longitudinal Study (ECLS-K). The data 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. Most important to our research, parents report the frequencies with which they help their children with homework, and teachers report the overall level of ability among students in each of their classes as well as the class-specific workload.

Using the estimated model, we conduct two sets of policy evaluations. In the first, we quantify the effects of ability tracking on the distribution of student test scores by comparing outcomes from the baseline model with counterfactual outcomes where no schools are allowed to track students. Over 95% of schools practice ability tracking under the baseline and therefore are affected by this policy. The ban on tracking increases ability dispersion within classrooms of these affected schools, which leads schools and parents to change their equilibrium effort levels. While schools increase their effort overall, the behavioral responses of parents vary by their children’s ability: Parents with low-achieving children reduce their effort while parents with high-achieving children increase theirs. With a ban on tracking, students with below-median prior achievement on average gain 3.6% of a standard deviation (sd) in outcome test score, while those with above-median prior score on average lose 5% sd. We also find evidence that estimating the test score production function per se is not sufficient to characterize the effects of changes in peer quality; the equilibrium interactions between schools and parents are important. If the effort adjustments by schools and by 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%. Put another way, 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.

In the second counterfactual experiment, we show how changes in proficiency standards impact tracking regimes, school and parent effort, and thus student achievement; and how these standards can be used to achieve certain goals. In particular, we search for region-specific proficiency standards that maximize average student test scores in each census region, which are found to be higher than their baseline levels. Under these
new proficiency standards, schools adjust their effort inputs and tracking regimes such that resources are shifted from low-achieving students towards high-achieving students, leading to decreases (increases) in student test scores for low-achieving (high-achieving) students.

Most research on ability tracking focuses on measuring how student test scores (or other outcomes of interest) vary with classroom ability composition, or peer group. There is considerable heterogeneity in results from empirical work assessing the effect of ability tracking on both the level and distribution of achievement.² Argys et al. (1996) find that tracking reduces performance of low ability students, Betts and Shkolnik (2000). 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 highlight the importance of explicitly taking into account the fact that enrollee 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.

Faced with the methodological challenges discussed in Manski (1993), Moffitt (2001), and Brock and Durlauf (2001), Fruehwirth (2013) builds an equilibrium model of student effort choices, which motivates her use of a particular instrument - the student accountability policy in North Carolina - to identify peer achievement spillover.³ Most related to our work, she uses this instrumental variable to eliminate the problem of nonrandom assignment in identifying peer achievement spillover. This is a valid approach as long as students are not reassigned in response to the policy. She provides support for this assumption by showing that observable compositions of peer groups do not appear to change in response to student accountability.⁴ 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 the direct social interaction within a class (see Blume, Brock, Durlauf

²See Betts (2011) for an extensive review of this literature.
³This policy requires that a student score above a certain level to be automatically promoted to the next grade.
⁴Assuming random assignment to classrooms within schools, Fruehwirth (2014) finds positive effects of peer parental education on student achievement.
and Ioannides (2011) for a comprehensive review on social interactions). Instead, we focus on the interaction between the school and enrollee households. In particular, we model schools’ decisions about student assignment and track-specific inputs precisely 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 achieve more when schools practice tracking (the status quo) than they do when tracking is banned (the counterfactual scenario). Walters (2014) studies the demand for charter schools. 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.\(^5\) 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. We emphasize 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 explains our estimation strategy. Section 4 describes the data. Section 5 presents the estimation results. Section 6 conducts counterfactual experiments. The last section concludes. The appendix contains further details and additional tables.

2 Model

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

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

\(^5\)Mehta (2013) also endogenizes school input choices.
different parental effort costs \((z \in \{low, high\})\). 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 equally within a school. Let \(g_s(a, z)\), \(g_s(a)\) and \(g_s(z|a)\) denote, respectively, the school-\(s\) specific 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 input \((e_{sj}^s)\), and parental effort \((e_{pi}^p)\), according to \(Y(a_i, q_j, e_{sj}^s, e_{pi}^p)\). Test score \(y_{ji}\) measures student achievement with noise \(\epsilon_{ji} \sim F(\cdot)\), such that

\[
y_{ji} = Y(a_i, q_j, e_{sj}^s, e_{pi}^p) + \epsilon_{ji}. \tag{1}
\]

2.2 Parent’s Problem

A parent values her child’s achievement, the utility from which is assumed to be logarithmic. Given the track-specific school input \((e_{sj}^s)\) and the peer quality \((q_j)\) of the track to which her child is assigned, a parent \(i\) chooses her own effort to maximize the utility from her child’s achievement, net of her effort cost \(C^p(e_{pi}^p, z_i)\):

\[
u(e_{sj}^s, q_j, a_i, z_i) = \max_{e_{pi}^p \geq 0} \left\{ \ln \left( Y(a_i, q_j, e_{sj}^s, e_{pi}^p) \right) - C^p(e_{pi}^p, z_i) \right\},
\]

Denote the optimal parental choice \(e_{pi}^p(e_{sj}^s, q_j, a_i, z_i)\).
2.3 School’s Problem

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

**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

Given a tracking regime \( \mu \), the optimal choice of track-specific effort \( e^s \equiv \{ e^s_j \}_j \) solves the following problem

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

where \( \omega \geq 0 \) measures the additional valuation to a school from a student passing the proficiency standard (\( y^* \)). \( C^s(e^s_j) \) 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 \( y^* \). 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, \(\{\mu_j(a_i)\}\). The overall objective of a school is thus the integration of individual students’ contributions. There are four constraints a school faces. The first two are the test score technology and the optimal response of the parent. The last two identity constraints define the size \(n_j\) and the average student quality \(q_j\) of a track. Let \(e^{s*}(\mu)\) be the optimal solution to (2).

### 2.3.2 Optimal Tracking Regime

A school’s operational cost (pecuniary and non-pecuniary) may vary with tracking regimes, captured by the function \(D(\mu)\). Balancing benefits and costs of tracking, a school solves the following problem:\(^6\)

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

where \(\eta_\mu\) is the idiosyncratic shifter associated with regime \(\mu\), which is i.i.d. across schools. \(M_s\) is the support of tracking regime for school \(s\), specified in Section 3.1.2.

### 2.4 Equilibrium

**Definition 2** A subgame perfect Nash equilibrium in school \(s\) consists of \(\{e^{p*}(\cdot), e^{s*}(\cdot), \mu^*, e^{\cdot*}\}\), such that

1) For each \((e_j^s, q_j, a_i, z_i)\), \(e^{p*}(\cdot)\) solves parent’s problem;
2) \((e^{s*}(\mu^*), \mu^*)\) solves school’s problem.

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

\(^6\)We assume that 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.
3 Empirical Implementation and Estimation

3.1 Further Empirical Specifications

3.1.1 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 \((a^*_1, a^*_2, a^*_3)\). Household types are unobservable to the researcher but may be correlated with observable household characteristics \(x\), which include a noisy measure of student ability, parental education and an indicator of single parenthood. Let \(\Pr((a^*, z) | x, s)\) be the distribution of \((a^*, z)\) given \(x\) in school \(s\). See the appendix for details.

Remark 1 The assumption that ability distributions are discrete and school-specific allows us to tractably model 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.

3.1.2 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, measured as 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 value from \(\left\{0, \frac{1}{K_s}, \frac{2}{K_s}, ..., 1\right\}\).\(^7\) Second, ability composition within a track cannot be “disjoint” in the sense that a track cannot mix low-ability students with 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 tracking regime depends only on the number of tracks in a regime, and is given by

\[
D(\mu) = \gamma|h|,
\]

\(^7\)Larger schools typically have more classrooms, hence finer grids of \(M_s\). As such, school size enters the model through the support of tracking regimes \(M_s\).
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, where $\gamma_1$ 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 track $\tilde{\mu}$ in school $s$ is given by

$$\frac{\exp (V_s(\tilde{\mu}|\Theta))}{\sum_{\mu'} \exp (V_s(\mu'|\Theta))}. \quad (3)$$

### 3.1.3 Achievement Function

Student achievement is governed by

$$Y(a, q, e^s, e^p) = \alpha_0 + a + \alpha_1 e^s + \alpha_2 e^p + \alpha_3 q + \alpha_4 e^s a + \alpha_5 e^s e^p, \quad (4)$$

where the coefficient on one’s own ability $a$ has been normalized 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.

### 3.1.4 Cost Functions

The functions for both the parent effort and the 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

$$C^s(e^s) = c_1^s e^s + c_2^s (e^s)^2.$$ 

**Remark 2** It is worth noting that the model we present here is relatively parsimonious. We have estimated more general versions of the model to include some additional features. Besides a more flexible test score production function, these additional features include: on the household side, we allowed for heterogeneity in both the cost and the effectiveness of parental effort. On the school side, we allowed for a nested logit counterpart of (3) in school’s regime choices, as well as school’s additional focus on the upper tail of student achievement. In likelihood ratio tests, we cannot reject that the simpler
model fits the data as well as the more complicated versions. The conclusions from our counterfactual experiments are robust to the inclusion of these additional features (results available upon request).

3.1.5 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 $j$ ($\tilde{e}_j^s$) is given by

$$\tilde{e}_j^s = e_j^s + \zeta_j^s,$$

where $\zeta_j^s \sim N(0, \sigma_{\zeta_j}^2)$.

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. Let $\Pr(\tilde{e}_j^p|e_j^p)$ denote the probability of observing $\tilde{e}_j^p$ when model effort is $e_j^p$.

3.2 Estimation

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

The parameters $\Theta$ to be estimated include model parameters $\Theta^0$ and parameters $\Theta^e$ that govern the distribution of measurement errors. The former ($\Theta^0$) consists of the following seven groups: 1) $\Theta_y$ governing student achievement production function $Y(\cdot)$, 2) $\Theta_\epsilon$ the distribution of shocks to test score $\epsilon$, 4) $\Theta_c^s$ governing school effort cost, 5) $\Theta_c^p$ governing parental effort cost, 4) $\Theta_D$ governing the cost of tracking regimes, 6) $\omega$, the weight associated with proficiency in school’s objective function, 7) $\Theta_T$ governing the distribution $P((a, z) | x)$ of household type given observables.

The endogenous outcomes observed for school $s$ ($O_s$) include the tracking regime $\tilde{\mu}_s$, track-specific school effort $\{\tilde{e}_s^j\}_j$ and household-level outcomes: parental effort $\tilde{e}_s^p$, the track to which the student is assigned $\tau_{si}$, and student final test score $y_{si}$. Let

$\text{8The distribution that enters the model directly, i.e., } g_s(a, z), \text{ does not involve additional parameters, because}$

$$g_s(a, z) = \int P((a, z) | x) dF_s(x),$$

where $F_s(x)$ is the distribution of $x$ in school $s$. 


Let \( X_s = \{x_{si}\}_i \) be the observed household characteristics in school \( s \). The vector \( X_s \) enters the likelihood via its correlation with household types \((a, z)\), which in turn affect all of \( O_s \).

**The Likelihood**  

The likelihood for school \( s \) is

\[
L_s(\Theta) = l_{\tilde{\mu}_s}(\Theta^0) \prod_j l_{s_j}(\Theta_\Delta) \prod_i l_{s_i}(\Theta_y, \Theta_\epsilon, \Theta_T, \Theta_\epsilon^p, \Theta^\zeta),
\]

where each part of the likelihood is as follows:

- \( l_{\tilde{\mu}_s}(\Theta^0) \) 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).

- \( l_{s_j}(\Theta_\Delta) \) is the contribution of the observed school effort \( \tilde{e}_{s_j} \) on 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 \( \Theta^\zeta \) as the observed effort is measured with error:

\[
l_{s_j}(\Theta_\Delta) = \frac{1}{\sigma_\zeta^s} \phi \left( \frac{\tilde{e}_{s_j}^s - e_{s_j}^{ss}(\tilde{\mu}_s|X_s;\Theta^0_\Delta)}{\sigma_\zeta^s} \right).
\]

where \( \phi \) denotes the standard normal density.

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

\[
l_{s_i}(\Theta_y, \Theta_\epsilon, \Theta_T, \Theta_\epsilon^p, \Theta^\zeta) = \sum_{a, z} P((a, z)|x_{i, s};\Theta_T) l_{s_i}((a, z)|\Theta_y, \Theta_\epsilon, \Theta_\epsilon^p, \Theta^\zeta),
\]

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

\[
l_{s_i}((a, z)|\Theta_y, \Theta_\epsilon, \Theta_\epsilon^p, \Theta^\zeta) = \left[ \Pr\{\text{track} = \tilde{\tau}_{si}|\tilde{\mu}_s\} \times \Pr(\tilde{e}_{si}|\epsilon_{si}(\tilde{\tau}_{si}, q_{s_{si}}, a, z|\Theta_y, \Theta_\epsilon^p)) \times f_{\epsilon_y}[(y_{si} - Y(a, q_{s_{si}}, e_s^{ss}(\cdot)|\Theta_y)|\Theta_\epsilon]\right].
\]

The three components of \( l_{s_i}((a, z)|\cdot) \) are
1. the probability of being assigned to $\tilde{\tau}_{si}$ given tracking regime $\tilde{\mu}_s$ and ability $a$, which is implied by $\tilde{\mu}_s$ and $g_s(a)$;³

2. the contribution of the observed parental effort $\tilde{e}_{si}^p$ given peer quality and the model predicted school effort $e_{\tilde{\tau}_{si}}^s$ on track $\tilde{\tau}_{si}$, which depends on parental cost parameters, the achievement parameters and the measurement error parameters; and

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

4 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 survey round (kindergarten). These students were then followed through the 8th grade, resulting 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.¹⁰

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 (used as the measure of 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).¹¹ Assuming that homework loads on students increase teachers’ effort cost, we use homework loads reported by the teacher to measure the school’s effort

³$\Pr\{\text{track} = j | a, \mu_s\} = \frac{\mu_{xj}(a)n_j}{\sum_j \mu_{xj}(a)n_j}$, where $n_j$ is the size of track $j$, i.e., $n_j = \sum_a \mu_{xj}(a)g_s(a)$.

¹⁰We focus on reading instead of math because the former involves a much larger sample size.

¹¹Ferreyra and Liang (2012) use time spent doing homework as a measure of student effort.
invested in each class.\footnote{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 B.1 for the questions.}

For the tracking regime, we use teachers’ reports on the ability level of their classes, which are available for different classes in the same school.\footnote{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.} The question for reading classes is: “What is the reading ability level of this child’s reading class, relative to the children in your 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. 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 on the same track. The size of each track is calculated as the number of classes on 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.\footnote{See Appendix B.2.}

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.\footnote{https://inventory.data.gov. See Table 15 in this paper for regional cutoffs.}

There are 8853 fifth grade students in 1772 schools in the ECLS-K sample. We delete observations missing key information, such as prior test scores, parental characteristics and track identity, leaving 7332 students in 1551 schools. Then, we exclude schools in which fewer than four classes and/or ten students are observed. The final sample includes 2789 students in 205 schools. The last sample selection criterion costs us a significant number of observations. However, 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 in each school. Appendix D shows that the summary statistics for the entire ECLS-K sample and our final sample are not very different.
4.1 Descriptive Statistics

Ability tracking is prevalent in the data; and most schools have two or three tracks. The first row of Table 1 shows that over 95% of schools practice ability tracking, i.e., they group all students into more than one track. About 13% of schools have four tracks. The most common number of tracks is three, which accounts for 46% of all schools. To summarize the distribution of students across schools, we calculate, for each school, the mean and the coefficient of variation (CV) of student prior test scores, and the fraction of students in the school who scored 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 student 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.

Table 1: Student Prior Test Scores in Schools by Numbers of Tracks

<table>
<thead>
<tr>
<th></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>

The following three tables 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 with four tracks. Table 2 shows that students in higher ability tracks, which are defined by their prior test scores, have both higher average outcome test scores and a higher probability of passing the regional proficiency cutoff. Comparing student outcome average scores and passing 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, an 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 the 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
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>44.92</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>51.45</td>
<td>64.54</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>57.99</td>
<td></td>
<td>97.87</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>

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 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, the 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></td>
<td>1.90</td>
<td>1.88</td>
<td>1.84</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>1.96</td>
<td>1.93</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>1.68</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></td>
<td>2.11</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>

5 Results

5.1 Parameters

In this section, we present parameter estimates of major interest. Other parameter estimates can be found in the appendix. Table 6 presents the production technology parameter estimates, where the coefficient of student’s own ability is normalized 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. 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. School effort is much less effective: increasing it by one sd causes a 2% sd increase in the outcome test score. Parental effort is quite

\[16\text{Notice these calculations are meant to illustrate the production technology; and they do not take into account any behavioral responses.}\]
effective: increasing it by one sd increases the outcome test score by 44% sd.

Table 6: Production technology parameters

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

Table 7 presents estimates of school-side parameters. We find a low value of $\omega$, suggesting that schools do not care much about students at the lower end of academic achievement, other than how they contribute to average 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.\(^{17}\) The second row shows that school effort cost is convex in effort levels.\(^{18}\) The last four rows show that the cost of various 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-


\(^{18}\)We cannot reject that the linear cost term $c_s^i$ is 0, so we fix it at 0.
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 correlation between household characteristics with parental cost types. There is a strong correlation between child ability and the probability that the parent is of a high-cost type. Higher-educated parents also tend to have higher effort cost. We do not find that, conditional on their own education level and their children’s ability, single parents are more likely to have higher effort cost.\(^{19}\)

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 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 fact shown in Table 5, where parents without college education and parents with lower-prior-achievement children exert more parental effort, yet the average student outcome test scores are lower in these households.

\(^{19}\)Notice that parents’ own characteristics are not significantly correlated with their effort cost types. Given that student abilities in our context measure their human capital achieved by 5th grade, which is a cumulative result from their parents’ investment, it is not surprising that child ability contains dominant information on the type of one’s parent.
Table 9: Parental Cost Type Distribution

<table>
<thead>
<tr>
<th>Cost Type</th>
<th>All Edu&gt;=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>
</tr>
<tr>
<td>High Cost</td>
<td>38.7%</td>
<td>44.8%</td>
<td>43.1%</td>
</tr>
</tbody>
</table>

5.2 Model Fit

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 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 10: Tracking Regimes

<table>
<thead>
<tr>
<th>All schools</th>
<th>Low Spread*</th>
<th>Low fraction of low ability**</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Data</td>
<td>Model</td>
</tr>
<tr>
<td>% 1 track</td>
<td>4.39</td>
<td>4.46</td>
</tr>
<tr>
<td>% 2 tracks</td>
<td>37.07</td>
<td>36.63</td>
</tr>
<tr>
<td>% 3 tracks</td>
<td>45.85</td>
<td>46.46</td>
</tr>
<tr>
<td>% 4 tracks</td>
<td>12.68</td>
<td>12.45</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.

Table 11 shows that the model fits well the 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 shows the cases by parental education and by single-parenthood. As seen, the model-predicted score distributions match well with the ones in the data.

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

<table>
<thead>
<tr>
<th>Track</th>
<th>1 Track Data</th>
<th>1 Track Model</th>
<th>2 Tracks Data</th>
<th>2 Tracks Model</th>
<th>3 Tracks Data</th>
<th>3 Tracks Model</th>
<th>4 Tracks Data</th>
<th>4 Tracks Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>51.84 50.72</td>
<td>45.95 46.86</td>
<td>44.92 45.76</td>
<td>45.40 46.32</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>51.98 54.85</td>
<td>51.38 51.38</td>
<td>51.44 50.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>55.62 56.64</td>
<td>51.45 55.70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>57.99 58.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 levels, 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. Parents with less education, single
parents, and students with lower prior score all have higher parent effort levels and lower outcome test scores.

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

<table>
<thead>
<tr>
<th>Track</th>
<th>1 Track Data</th>
<th>Model</th>
<th>2 Tracks Data</th>
<th>Model</th>
<th>3 Tracks Data</th>
<th>Model</th>
<th>4 Tracks Data</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.86</td>
<td>1.86</td>
<td>1.75</td>
<td>1.85</td>
<td>1.75</td>
<td>1.84</td>
<td>1.82</td>
<td>1.84</td>
</tr>
<tr>
<td>2</td>
<td>1.90</td>
<td>1.85</td>
<td>1.88</td>
<td>1.86</td>
<td>1.84</td>
<td>1.86</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>1.96</td>
<td>1.84</td>
<td></td>
<td>1.93</td>
<td>1.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>1.68</td>
<td></td>
<td>1.84</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 Data</th>
<th>Model</th>
<th>2 Tracks Data</th>
<th>Model</th>
<th>3 Tracks Data</th>
<th>Model</th>
<th>4 Tracks Data</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.07</td>
<td>2.21</td>
<td>2.31</td>
<td>2.49</td>
<td>2.57</td>
<td>2.62</td>
<td>2.29</td>
<td>2.75</td>
</tr>
<tr>
<td>2</td>
<td>2.03</td>
<td>1.97</td>
<td>2.37</td>
<td>2.23</td>
<td>2.71</td>
<td>2.44</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>2.11</td>
<td>1.91</td>
<td></td>
<td>2.78</td>
<td>2.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>2.08</td>
<td></td>
<td>1.94</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Household Characteristics</th>
<th>Parent effort Data</th>
<th>Parent effort Model</th>
<th>Outcome score Data</th>
<th>Outcome score Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than college</td>
<td>2.35</td>
<td>2.38</td>
<td>48.00</td>
<td>49.24</td>
</tr>
<tr>
<td>College</td>
<td>2.12</td>
<td>2.16</td>
<td>54.24</td>
<td>52.67</td>
</tr>
<tr>
<td>Single parent</td>
<td>2.37</td>
<td>2.36</td>
<td>48.76</td>
<td>49.53</td>
</tr>
<tr>
<td>Two parents</td>
<td>2.18</td>
<td>2.23</td>
<td>52.37</td>
<td>51.67</td>
</tr>
<tr>
<td>Low prior score</td>
<td>2.52</td>
<td>2.48</td>
<td>45.35</td>
<td>47.79</td>
</tr>
<tr>
<td>High prior score</td>
<td>1.78</td>
<td>2.03</td>
<td>57.96</td>
<td>54.72</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 subgroups of students defined by their characteristics, such as prior test scores.

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. Unlike in the tracking-ban counterfactual, a school re-optimizes its tracking decision in response. In particular, we solve for optimal region-specific proficiency standards that maximize the average achievement.

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 one 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 of Figure 2 on the top show results for outcome test scores and pass rates by decile of prior test scores. The effects from a ban of tracking 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.\footnote{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., and Betts and Shkolnik (2000) and Hoffer (1992).} Consistent with the first two panels, the third panel shows that the fraction of students who gain from a ban of tracking 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. 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 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. The 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 major 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 parental effort adjustment. 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 in Effort Inputs

The test score technology plays an important role in evaluating the effect of tracking on student outcomes. This may 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, we contrast the ATE of banning tracking from our model prediction with the ATE ignoring endogenous effort responses. Let \((q^*, e^{ss}, e^{ps})\) denote inputs under the baseline scenario of endogenous tracking, and \((q^*_CF, e^{ss}_CF, e^{ps}_CF)\) 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. \(Y(a, q^*_CF, e^{ss}_CF, e^{ps}_CF) - Y(a, q^*, e^{ss}, e^{ps})\) (recall \(Y(\cdot)\) is the test score technology). The blue (short dashed) line is the ATE, ignoring school effort adjustments, i.e. \(Y(a, q^*_CF, e^{ss}_CF, e^{ps}_CF) - Y(a, q^*, e^{ss}, e^{ps})\). Banning tracking pushes schools 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. $Y(a, q^*_s, e^{s*}, e^{p*}) - Y(a, q^*, e^{s*}, e^{p*})$. 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. $Y(a, q^*_s, e^{s*}, e^{p*}) - Y(a, q^*, e^{s*}, e^{p*})$.

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

Schools care about the fraction of students above the proficiency standard, which will in turn affect their decisions. In this counterfactual experiment, we search for the
proficiency standards that, when imposed on schools, will achieve certain educational goals in the new equilibrium.\textsuperscript{21} In particular, we search for region-specific proficiency standards that maximize the region-specific averages of student test scores.

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

\textsuperscript{21}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 in high-stake tests.
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 to 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.

In most cases, policy changes will create both winners and losers. This is especially true in the case of educational policies that may affect ability tracking, as highlighted in our second counterfactual experiment. 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.
Figure 5: The effect of increasing in proficiency standards to maximize average achievement
7 Conclusion

We have developed and estimated a model of ability tracking, in which a school’s tracking regime, track-specific input, parental efforts and student achievements 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 heterogeneous across students with different prior achievement, and that ignoring endogenous effort responses to changes in tracking regimes can seriously bias the effect of tracking. We have illustrated how changes in proficiency standards can be used to achieve certain goals, via their impacts on schools’ decision on tracking regimes and effort inputs, as well as the trade-offs involved in achieving these goals.

There are potential extensions to our paper that are worth pursuing for future research. One extension that is of particular interest is to combine studies on the matching between schools and households and this paper into one coherent framework. This extension would form a more comprehensive view of how peer composition is determined both at the school and at the class levels.
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 \( F_{s,a}(x) \), which is the normal distribution approximation of prior test scores of all students in school \( s \), \((\{x^o_{si}\})\). A level \( a^s_l \) is defined as the expectation of prior score within the \( l^{th} \) tercile in school \( s \) computed using \( F_{s,x}(x) \), i.e.,

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

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

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

In particular, the ability distribution is given by

\[
\begin{align*}
    \Pr(a = a^s_1 \mid x^a, s) &= 1 - \Phi \left( \frac{x^a - T^s_1}{\sigma_a} \right) \\
    \Pr(a = a^s_3 \mid x^a, s) &= \Phi \left( \frac{x^a - T^s_2}{\sigma_a} \right) \\
    \Pr(a = a^s_2 \mid x^a, s) &= 1 - \Pr(a = a^s_1 \mid x^a) - \Pr(a = a^s_3 \mid x^a),
\end{align*}
\]

where \( \sigma_a \) is a parameter to be estimated. The probability that a parent is of a high-cost
type is given by

\[
\text{Pr}(z = z_2|x, a_i) = \Phi(\theta_0 + \theta_1 a_i + \theta_2 I(x_1 \geq \text{college}) + \theta_3 I(x_2 = \text{single parent})).
\]  

(5)

A.2 Parent effort measurement system

Observed parental effort is discrete and ordered: \( \tilde{e}_p \in \{0, 0.5, 1.5, 3.5, 5\} \). Model 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:

\[
\begin{align*}
\tilde{e}_p = 0 & \iff -\infty < e_p^* + \zeta^p \leq \kappa_0 \\
= 0.5 & \iff \kappa_0 < e_p^* + \zeta^p \leq \kappa_1 \\
= 1.5 & \iff \kappa_1 < e_p^* + \zeta^p \leq \kappa_2 \\
= 3.5 & \iff \kappa_2 < e_p^* + \zeta^p \leq \kappa_3 \\
= 5 & \iff \kappa_3 < e_p^* + \zeta^p \leq \infty,
\end{align*}
\]

resulting in \( \text{Pr}(\tilde{e}_p|e_p^*) \).

B Data details

<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>

B.1 Effort measures

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...

1. Never
2. Less than once a week
3. 1 to 2 times a week
4. 3 to 4 times a week
5. 5 or more times a week
6. REFUSED
7. DON’T KNOW

**Question 2 in the Spring 6 Teacher-Level Questionnaire**

For subjects you teach, about how much time do you expect children to spend on homework in each of the following areas on a typical evening? CIRCLE ONE NUMBER ON EACH LINE. CIRCLE N/A IF YOU DO NOT TEACH THE SUBJECT.

a. Reading and Language Arts.

1. I don’t teach this subject.
2. None.
3. 10 min.
4. 20 min.
5. 30 min.
6. More than 30 min.

**B.2 Prior test scores by track**
Table 16: 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>48.2826</td>
<td>11.0239</td>
<td>886</td>
</tr>
<tr>
<td>2</td>
<td>52.2786</td>
<td>8.4345</td>
<td>1227</td>
</tr>
<tr>
<td>3</td>
<td>53.9920</td>
<td>8.2680</td>
<td>563</td>
</tr>
<tr>
<td>4</td>
<td>54.9899</td>
<td>6.5394</td>
<td>113</td>
</tr>
</tbody>
</table>

Table 17: 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>0.9315</td>
<td>0.1807</td>
<td>886</td>
</tr>
<tr>
<td>2</td>
<td>1.0198</td>
<td>0.1469</td>
<td>1227</td>
</tr>
<tr>
<td>3</td>
<td>1.0471</td>
<td>0.1403</td>
<td>563</td>
</tr>
<tr>
<td>4</td>
<td>1.0880</td>
<td>0.1219</td>
<td>113</td>
</tr>
</tbody>
</table>

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

C Parameter Estimates

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 sd of shock in ability distribution</td>
</tr>
<tr>
<td>$\sigma_\epsilon$</td>
<td>7.31</td>
<td>0.13 sd test score measurement error</td>
</tr>
<tr>
<td>$\sigma_\zeta_s$</td>
<td>0.55</td>
<td>0.02 sd school effort measurement error</td>
</tr>
<tr>
<td>$\kappa_0$</td>
<td>-0.03</td>
<td>0.61 cut-point 1, parent effort measurement</td>
</tr>
<tr>
<td>$\kappa_1$</td>
<td>0.91</td>
<td>0.60 cut-point 2, parent effort measurement</td>
</tr>
<tr>
<td>$\kappa_2$</td>
<td>2.11</td>
<td>0.59 cut-point 3, parent effort measurement</td>
</tr>
<tr>
<td>$\kappa_3$</td>
<td>2.90</td>
<td>0.65 cut-point 4, parent effort measurement</td>
</tr>
</tbody>
</table>

D Sample Selection
Table 19: Summary Statistics Before and After the Imposition of Sample Restrictions

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parental Education</td>
<td>8047</td>
<td>0.5523</td>
<td>0.4973</td>
<td>2789</td>
<td>0.5887</td>
<td>0.4922</td>
</tr>
<tr>
<td>Single Parent</td>
<td>7953</td>
<td>0.2321</td>
<td>0.4222</td>
<td>2789</td>
<td>0.1915</td>
<td>0.3935</td>
</tr>
<tr>
<td>Prior Test Score</td>
<td>8562</td>
<td>50.1726</td>
<td>9.7178</td>
<td>2789</td>
<td>51.4649</td>
<td>9.5216</td>
</tr>
<tr>
<td>Outcome Test Score</td>
<td>8751</td>
<td>50.1542</td>
<td>9.6469</td>
<td>2777</td>
<td>51.6809</td>
<td>9.4012</td>
</tr>
<tr>
<td>Parental Effort</td>
<td>7788</td>
<td>2.2901</td>
<td>1.5522</td>
<td>2703</td>
<td>2.2155</td>
<td>1.5308</td>
</tr>
<tr>
<td>Teacher Effort</td>
<td>2784</td>
<td>2.0337</td>
<td>0.5835</td>
<td>533</td>
<td>1.8978</td>
<td>0.5919</td>
</tr>
</tbody>
</table>

References


Fruehwirth, J. C. Does the education of peers mothers and fathers matter? mechanisms of parental spillovers in the classroom (2014).


