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

Preliminary: Please do not circulate or cite.

Chao Fu † Nirav Mehta ‡

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.

1 Introduction

Ability tracking, the practice of allocating students into different classrooms based on prior performance, is pervasive yet controversial (Yee [2013]). It is pervasive because schools are typically endowed with heterogeneous sets of students, and schools may want to create more homogeneous environments within a classroom to facilitate learning. It is controversial because tracking *may* benefit students of certain ability levels

*We thank Yuseob Lee, George Orlov, and Atsuko Tanaka for excellent research assistance. All errors are ours.

†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

while reducing the achievement of other students. There is considerable policy interest in learning how ability tracking affects different types of students, and how policy changes, such as increasing proficiency standards, would affect schools' tracking choices and student outcomes.

Three complications need to be addressed in order to answer these questions. First, a change in peer composition in one classroom necessarily involves student re-allocation, hence changes of peers in some other classroom(s). A study of the effects of peer composition may neglect to measure treatment effects for other groups if it treats each classroom in isolation. Second, one needs to understand how school and parental inputs are chosen given a tracking regime, in order to infer what these input levels would be if tracking regimes (hence peer compositions) were changed, which would in turn affect student achievement. Finally, the 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 principle, one could quantify the effect of ability tracking by running a randomized experiment where different tracking regimes, including non-tracking, are imposed on various schools, and then compare outcomes across treatment groups. One could also experiment with various policy changes to learn their effects. Practically, however, it is infeasible to run randomized control trials for every set of school characteristics (including student composition), and every alternative policy scenario.

As a feasible alternative, we adopt a structural approach to confront the above complications . We develop and estimate a model that treats tracking regimes, track-specific school effort, parental effort and student outcomes as joint outcomes from a subgame perfect Nash equilibrium. In the model, there is a continuum of households of different types whose children are educated in one school. A household type is defined by the ability of the child and the effectiveness and cost types of the parent in helping her child to learn, where parental types are households' private information. A child's achievement depends on ones own ability, effort invested by the school and by ones parent, and the quality of ones peers. The net effect of peer quality differs across students, depending on their own ability levels. A parent maximizes her child's achievement by choosing costly parental effort in response to the track her child is assigned to (which determines peer quality) and the effort invested by the school. A school maximizes a weighted average of the total achievement of its students, the fraction of students whose achievement satisfies a proficiency requirement, and the fraction of students with superior performance. To achieve its goal, the school chooses a tracking regime and track-specific effort inputs, taking into account its costs and responses by different types of parents. A distribution of treatment effects arises naturally at both the household level and the school level.

We estimate our model 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 score, class membership, and end-of-the-year test score. Most important to our research, a parent reports the frequency with which she helps her child with homework, and a teacher reports the overall level of ability among students in her class as well as the class-specific workload. We also observe the Census region in which each school is located. We estimate parameters of the model using maximum likelihood.

Using the estimated model, we conduct two types of policy evaluations. It is important to note that our framework naturally allows policies to produce winners and losers due to the differential impact tracking regimes may have on students of different abilities and parental backgrounds. 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 are therefore affected by this policy. The ban on tracking increases ability dispersion within classrooms of these affected schools, which leads to changes in school and parent effort. We find evidence of heterogeneous effects of tracking: students with below-median prior achievement on average lose 4.1% of a standard deviation (sd) in outcome test score when schools are allowed to track by ability. Students with above-median prior score on average gain 4.6% sd when schools are allowed to track by ability. We also find evidence that estimating the technology mapping own and peer characteristics to outcomes is not sufficient: the equilibrium interactions between schools and parents are important. Ignoring the behavioral responses of parents by holding their input levels constant when banning ability tracking would overstate the gain from tracking for students with above-median prior achievement by over 60%.

In the second policy evaluation, we quantify how changes in proficiency standards would affect tracking regimes, school and parent effort, and student achievement. We then simulate model outcomes when standards in each of four Census regions are increased by 10% and 25% from their regional baseline. The achievement of students with below-median prior test scores suffers in both scenarios, but moreso in the latter, where standards are unattainably high for them. Students in the top decile of prior test score, on the other hand, benefit when standards are increased. Their gains double when standards are further increased from 10% higher bar to a 25% higher bar.

Most research on ability tracking focuses on measuring how the test scores (or other outcomes of interest) of individual students 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: Argys et al. [1996] find that tracking reduces performance of low ability students, Betts and Shkolnik [2000] 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, and 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 by school type (e.g. public, Catholic). See Betts [2011] for an extensive review of this literature.

While our work focuses on how peer groups are determined within a school, a different literature studies how households sort themselves into different school-level peer groups. 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. 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. Mehta [2013] also endogenizes school input choices. Our work complements this literature by taking a first step toward studying a school's tracking decision, which determines 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.

2 Model

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

2.1 Players

A school s is endowed with a continuum of households of measure one.[1] Households are of different types defined by the ability levels of the child (a) and parent types ($z = [z_e, z_c]$), where z_e is the effectiveness of parental effort and z_c is a cost shifter for parental effort. Student ability a is known to the household and the school, but z is a household's private information. 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. Let Q_s be the average student ability in the school. In the following, we suppress the school index s.

2.2 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 school's choices, parents choose their own parental effort.
Stage 3: Student test score is realized.

2.3 Production Function

The achievement of a student i in track j depends on the student's ability a_i, the average ability of ones classmates (q_j), the dispersion of student ability in the class measured by the coefficient of variation (ξ_j), track-specific effort input $\left(e_j^s\right)$, parental effort (e_i^p), parental efficiency (z_{ie}), and the overall student ability in the school (Q), according to $Y(a_i, q_j, \upsilon_j, e_j^s, e_i^p, z_{ie}, Q)$.[2] Test score y_{ji} measures student achievement

[1] We normalize all schools to have a measure 1 of households without any loss of generality because, as shown later, achievement production technology is constant returns to scale in terms of class size and the school's objective is also invariant to school size.

[2] See Appendix A.1 for functional forms.

with noise $\epsilon_{ji} \backsim F_{\epsilon}(\cdot)$, such that

$$y_{ji} = Y(a_i, q_j, \xi_j, e_j^s, e_i^p, z_{ie}, Q) + \epsilon_{ji}. \tag{1}$$

2.4 Parent's Problem

A parent cares about her child's achievement, the utility from which is assumed to be logarithmic. Given the track-specific school input $\left(e_j^s\right)$ and the peer quality (q_j) of the track to which her child is assigned, a parent chooses her effort to maximize utility net of her effort cost $C^p\left(e_i^p, z_{ic}\right)$:

$$\max_{e_i^p \geq 0} \left\{ \ln\left(Y(a_i, q_j, \xi_j, e_j^s, e_i^p, z_{ie}, Q)\right) - C^p\left(e_i^p, z_{ic}\right) \right\},$$

Denote the optimal parental choice $e^{p*}(e_j^s, q_j, \upsilon_j, a_i, z_i, Q)$ and the maximized utility $u(e_j^s, q_j, \upsilon_j, a_i, z_i, Q)$.

2.5 School's Problem

A school cares about the average test score of its students. In addition, it may also care about the tails of the score distribution: the fraction of students who can pass a proficiency standard y^*, and the fraction of students who exceed a higher threshold y^{**}. It chooses a tracking regime and track-specific inputs. Tracking specifies how students are allocated across classrooms based on student ability only. 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 $\mu = \{\mu_j(:)\}_j$.*

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 1). As such, the problem can be solved backwards.

Given a tracking regime μ, the optimal choice of track-specific effort $e^s \equiv \left\{e_j^s\right\}_j$ solves the following problem

$$\max_{e^s \geq 0} \left\{ \begin{array}{c} \int_i \left\{ \sum_j \left[\begin{array}{l} E\left(y_{ji} + \omega_1 I(y_{ji} > y^*) + \omega_2 I(y_{ji} > y^{**})\right) \\ -C^s(e_j^s) \end{array} \right] \mu_j(a_i) \right\} di \\ s.t.\ y_{ji} = Y(a_i, q_j, \xi_j, e_j^s, e_i^p, z_{ie}, Q) + \epsilon_{ji} \\ e_{ji}^p = e^{p*}(e_j^s, q_j, \xi_j, a_i, z_i, Q) \\ n_j = \sum_a \mu_j\left(a\right) g_s\left(a\right) \\ q_j = \frac{1}{n_j} \sum_a \mu_j\left(a\right) g_s\left(a\right) a \\ \xi_j = \frac{\sqrt{\frac{1}{n_j} \sum_a \mu_j(a) g_s(a)(a-q_j)^2}}{q_j}. \end{array} \right\}. \tag{2}$$

The integrand gives student i's expected net contribution to the school's objective, denominated in units of test scores. The terms in the bracket gives student i's expected net contribution conditional on her being on track j, where the expectation is taken over both the shock to the test score ϵ_{ji}, and the distribution of parent type z_i given student ability a_i, $(g_s\left(z|a\right))$. In particular, she contributes by her test score y_{ji} and an additional ω_1 if y_{ji} is above the proficiency bar y^*, and ω_2 as well if y_{ji} is above the high threshold y^{**}. School effort is costly and given by $C^s(e_j^s)$, the per-student effort cost on track j. 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 $\left\{\mu_j(a_i)\right\}_j$. When evaluating its objective, the school integrates over the student body. There are four constraints the school face. 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 $e^{s*}\left(\mu\right)$ be the optimal solution to (2) and $V_s\left(\mu\right)$ the maximized value.

A school's operational cost (pecuniary and non-pecuniary) may vary with tracking regimes, captured by the function $D\left(\mu\right)$. Balancing benefits and costs, a school solves the following problem

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

where $\eta_{s\mu}$ is the idiosyncratic shifter associated with regime μ for school s, and M_s is the support of tracking regime for school s, specified in Section 3.1.2.

2.6 Equilibrium

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

We solve the model using backward induction. First, solve the parent's problem for any given $(e_j^s, q_j, \xi_j, a_i, z_i, Q)$.[3] Second, for a given μ, solve the track-specific school inputs e^s. Finally, optimize over tracking regimes to obtain the optimal μ^* and the associated $(e^{p*}(\cdot), e^{s*})$.[4]

3 Empirical Implementation and Estimation

3.1 Further Empirical Specifications

3.1.1 Household Type

There are 12 types of households in a school: four types of parents (two types of z_e and two types of z_c), and three school-specific student ability levels (a_1^s, a_2^s, a_3^s).[5] Household type is 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 $P((a^s, z)|x, s)$ be the distribution of (a^s, z) given x in school s.

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

[3]The parent's problem has an analytical solution.
[4]Note that we must solve for the value of each $\mu \in M_s$ to evaluate the likelihood.
[5]Our 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 vital to our understanding why schools make different tracking decisions.

value from $\left\{0, \frac{1}{K_s}, \frac{2}{K_s}, ..., 1\right\}$.[6] 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 are identical across classrooms in the same track.

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

$$D(\mu) = \gamma_{|\mu|},$$

where $|\mu| \in \{1, 2, 3, 4\}$ is the number of tracks in regime μ. $\gamma = [\gamma_1, \gamma_2, \gamma_3, \gamma_4]$ is the vector of tracking cost, with γ_1 normalized to 0.

The permanent idiosyncratic shifter in the school objective function, $\eta_{s\mu}$, follows a generalized extreme-value distribution (nested logit), where all regimes with the same number of tracks share a nest. Let $m(\mu) \in \{1, .., M_s\}$ denote the number of tracks in regime μ. From the researcher's point of view, conditional on a set of parameter values Θ, the probability of observing a particular track $\check{\mu}$ in school s is given by

$$\frac{\exp\left(\frac{V_s(\check{\mu}|\Theta)}{\lambda}\right)\left(\sum_{\mu' \in M_s | m(\mu') = m(\check{\mu})} \exp\left(\frac{V_s(\mu'|\Theta)}{\lambda}\right)\right)^{\lambda - 1}}{\sum_{m=1}^{M}\left(\sum_{\mu' \in M_s | m(\mu') = m} \exp\left(\frac{V_s(\mu'|\Theta)}{\lambda}\right)\right)^{\lambda}}, \tag{3}$$

where $(1 - \lambda)$ measures the correlation between $\eta_{s\mu}$ within a nest; when $\lambda = 1$, all $\eta_{s\mu}$ are i.i.d. extreme-value distributed.

3.1.3 Measurement Errors

We assume that both the school effort e^s and the parent effort e^p are measured with idiosyncratic errors, such that the observed effort $(\widetilde{e})$ are given by

$$\widetilde{e}^s_j = e^s_j + \varepsilon^s_j \text{ and } \widetilde{e}^p_i = e^p_i + \varepsilon^p_i,$$

[6] 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 $\varepsilon_j^s \sim N\left(0, \sigma_{\varepsilon^s}^2\right)$, $\varepsilon_i^p \sim N\left(0, \sigma_{\varepsilon^p}^2\right)$.

3.2 Estimation

The parameters Θ to be estimated include model parameters Θ^0 and parameters Θ^ε that govern the distribution of measurement errors. The former (Θ^0) consists of the following seven groups: 1) Θ_y governing student achievement production function $Y(\cdot)$, 2) Θ_ϵ the distribution of shocks to test score ϵ, 4) Θ_{c^s} governing school effort cost, 5) Θ_{c^p} governing parental effort cost, 4) Θ_D governing the cost of tracking regimes, 6) ω, the importance weight in school's objective function, 7) Θ_T governing the distribution $P\left((a,z)\,|x\right)$ of household type given observables.[7]

We estimate Θ via maximum likelihood (MLE). The parameter estimates maximizes the probability of observing the joint endogenous outcomes given the observed exogenous household variables x and school variable w. The endogenous outcomes observed for school s (O_s) include the tracking regime μ_s, track-specific school effort $\left\{\widetilde{e}_{sj}^s\right\}_j$ and household-level outcomes: parental effort $\widetilde{e}_{si}^p$, the track to which the student is assigned τ_{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 via its correlation with household types (a,z), which in turn affects all of O_s.

The likelihood for school s is given by

$$L_s(\Theta) = l_{\mu_s}\left(\Theta^0\right)\prod_j l_{sj}\left(\Theta \backslash \Theta_D\right)\prod_i l_{si}\left(\Theta_y, \Theta_\epsilon, \Theta_T, \Theta_{c^p}, \Theta^\varepsilon\right),$$

where each part of the likelihood is as follows:
$l_{\mu_s}\left(\Theta^0\right)$ is the probability of observing the tracking regime, which depends on all model parameters Θ^0, since every part of Θ^0 affects a school's tracking decision, but not on Θ^ε. It is given by (3).
$l_{sj}\left(\Theta\backslash\Theta_D\right)$ is the contribution of the observed school effort $\left(\widetilde{e}_{sj}^s\right)$ on track j given the tracking regime μ_s. It depends on all Θ^0 but Θ_D since the latter does not affect school effort decision given the tracking regime. It also depends on Θ^ε as the observed effort

[7]The distribution that enters the model directly, i.e., $g_s(a,z)$, does not involve additional parameters, because

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

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

is measured with error:

$$l_{sj}\left(\Theta\backslash\Theta_D\right) = \frac{1}{\sigma_{\varepsilon^s}}\phi\left(\frac{\widetilde{e}^s_{sj} - e^{s*}_j(\mu_s|X_s;\Theta^0\backslash\Theta_D)}{\sigma_{\varepsilon^s}}\right).$$

where ϕ denotes the standard normal density.

$l_{si}\left(\Theta_y,\Theta_T,\Theta_{c^p},\Theta^\varepsilon\right)$ is the contribution of household i, which involves an integration of type-specific contributions to the likelihood over the distribution of household types.

$$l_{si}\left(\Theta_y,\Theta_\epsilon,\Theta_T,\Theta_{c^p},\Theta^\varepsilon\right) = \sum_{a,z} P((a,z)\left|x_i,s;\Theta_T\right) l_{si}\left((a,z)\left|\Theta_y,\Theta_\epsilon,\Theta_{c^p},\Theta^\varepsilon\right.\right),$$

where $l_{si}\left((a,z)\left|\Theta_y,\Theta_{c^p},\Theta^\varepsilon\right.\right)$ is the contribution of household i if it were type (a,z). It consists of 1) the probability of being assigned to τ_{si} *given* tracking regime μ_s *and* ability a, which in itself does not depend on parameters as it is directly implied by μ_s and a^s;[8] 2) the contribution of the observed parental effort $\widetilde{e}_s i^p$ *given* peer quality and the model predicted school effort $e_\tau s i^s$ on track $\tau_s i$, 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.

$$l_{si}\left((a,z)\left|\Theta_y,\Theta_\epsilon,\Theta_{c^p},\Theta^\varepsilon\right.\right) = \left[\begin{array}{l} \Pr\{track=\tau_{si}|a,\mu_s\}\times \\ \frac{1}{\sigma_{\varepsilon^p}}\phi\left(\frac{\widetilde{e}^p_{si}-e^{p*}(e^s_{\tau_{si}},q_{\tau_{si}},\upsilon_{\tau_{si}},a,z,Q|\Theta_y,\Theta_{c^p})}{\sigma_{\varepsilon^p}}\right)\times \\ f_{\epsilon_y}\left[\left(y_{si}-Y(a,q_{\tau_{si}},\upsilon_{\tau_{si}},e^s_{\tau_{si}},e^p(\cdot),z_e,Q|\Theta_y)\right)|\Theta_\epsilon\right] \end{array}\right]$$

4 Data

We use 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

[8] $\Pr\{track=j|a,\mu_s\} = \frac{\mu_{sj}(a)n_j}{\sum_j \mu_{sj}(a)n_j}$, where n_j is the size of track j.

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.[9] We restrict the sample to schools with at least 10 students in the sample. The final sample size is 205 schools with a total of 2,789 students.

The data are rich enough to allow us to model the interactions between schools and parents. For students, we observe their prior test score (used as the measure of their ability), class membership (to identify their ability track), and end-of-the-year test score, 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), education (which affects effectiveness of parental input), household composition (single parent or not – affects cost of providing parental input). 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.

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.[10] 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, and the number of classes in tracks as the sizes of various tracks. Although the relative ability ranking is clear cross answers a), c) and b) or d), the relative ranking between b) and d) is less clear. We use the average student prior test scores within each of the two types of classes to determine the ranking between them. As a result, higher tracks have students with higher mean ability.

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.[11] We choose the higher cutoff y^{**} to match the national fraction of students meeting performance standards (32%) as defined by the National Assessment of Educational Progress (NAEP). This

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

[10]The ECLS-K sampling scheme follows many students at the same school. As such, we have the above information on classes for several classes at each school.

[11]See Table 18 for regional cutoffs.

corresponds to the 68th percentile prior test score in our sample.[12]

4.1 Descriptive Statistics

For each school, we calculate the mean and the standard deviation of student prior test scores, and the fraction of students in the school who scored below the sample median. Table 1 presents the mean of these summary statistics across schools, by the number of tracks in a school. Rows 1-2 of Table 1 show, respectively, the cross-school mean of school-level mean and coefficient of variation (CV) of students' prior test scores. Row 3 shows the cross-school average of the fraction of low-achieving students. On average, schools with more tracks have higher dispersion and lower student prior score.

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

	1 Track	2 Tracks	3 Tracks	4 Tracks	All Schools
CV	0.14	0.17	0.16	0.18	0.16
Mean	53.3	51.3	51.4	49.8	51.2
% below median	44.5	49.9	50.7	58.4	51.1
% of schools	4.39	37.1	45.8	12.7	100.0

The following three tables present summary statistics by the number of tracks in the school and the identity of a track. For example, (Column 4, Row 3)th entry of a table refers to students who belong to the third track in a school with four tracks. Table 2 shows that students in higher ability tracks have both higher average outcome test scores and a higher proportion of students passing the proficiency cutoff. Tables 3 and 4 show that, while average teacher effort (expected hours of homework done by students per week) increases as we look at tracks with higher mean ability, average parental inputs (time spent per week helping child with English coursework) decrease. Table 5 shows that lower-educated parents, single-parent households, and students with lower prior score all have higher average levels of parental inputs and lower outcome achievement.

[12] http://dashboard.ed.gov/statecomparison.aspx?i=c4&id=0&wt=40

Table 2: Average outcome score and percent of students passing the cutoff by track, by number of tracks at school

	1 track		2 tracks		3 tracks		4 tracks	
Track	Score	% pass	Score	% pass	Score	% pass	Score	% pass
1	51.84	69.42	45.95	50.88	44.92	42.85	45.40	33.38
2			51.98	75.75	51.38	68.47	51.44	61.00
3					55.62	84.39	51.45	64.54
4							57.99	97.87

Table 3: Average teacher effort by track, by number of tracks at school

Track	1 track	2 tracks	3 tracks	4 tracks
1	1.86	1.75	1.75	1.82
2		1.90	1.88	1.84
3			1.96	1.93
4				1.68

Table 4: Average parent effort by track, by number of tracks at school

Track	1 track	2 tracks	3 tracks	4 tracks
1	2.07	2.31	2.57	2.29
2		2.03	2.37	2.71
3			2.11	2.78
4				2.08

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

	Parent effort	Outcome test score
Less than college	2.35	48.00
Parent college	2.12	54.24
Single parent hh	2.37	48.76
Two-parent hh	2.18	52.37
Grade 3 score below median	2.61	45.35
Grade 3 score above median	1.82	57.96

5 Results

5.1 Parameters

Parameter estimates are in Appendix C.[13]

The production technology parameter estimates imply that while the average quality of one's classmates contributes to one's own achievement, the overall quality of students in the school does not have a significant impact (in fact, it is marginally negative), which suggests that peer effects occur mainly within a class. Parental efforts have a significant impact on student achievement and are complementary to student ability. School effort is complementary to both student ability and parental efforts. More importantly, the effectiveness of school effort decreases with the dispersion of student ability levels within a track. Intuitively, it is harder to for teachers to improve student achievement when the class is composed of students with very different ability levels.

The values of the estimated ω_1 and ω_2 are small. The low value of ω_1 means that schools do not care much about students at the lower end of academic achievement. This finding may be due to the fact that the test score we use is not from a high-stakes test, and 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.[14]

The type-specific component of parental effort cost for the high-cost household is about 10% higher than that of the low-cost household. Effort from high-efficacy parents is 13% more productive than that of low-efficacy parents. Parents with higher-ability children are more likely to be the high-efficacy type, but their effort also tend to be more costly. Single-parent households have a substantially higher probability of being the high-cost type, while college-educated parents are more likely to be more high efficacy in helping their children study.

5.2 Model Fit

The following tables show that the model can reproduce key patterns in the data:

[13] These estimates are still preliminary.

[14] 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. [2011].

- There are fewer average tracks at schools with less variation in student ability and higher prior test scores.
- Outcome test scores are increasing in track.
- School effort is increasing in track.
- Parent effort is decreasing in track.
- Parents with less education, single parents, and students with lower prior score all have higher parent effort levels and lower outcome test scores.

Table 6: Distribution of schools by number of tracks, overall and by certain characteristics

	All schools		Low Spread*		Low fraction of low ability**	
	Data	Model	Data	Model	Data	Model
% 1 track	4.39	4.64	4.85	5.71	4.55	5.47
% 2 tracks	37.07	36.86	36.89	39.65	40.91	40.14
% 3 tracks	45.85	45.54	47.57	43.64	47.27	43.62
% 4 tracks	12.68	12.96	10.68	11.00	7.27	10.77

* "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 7: Outcome test score by track and number of tracks

	1 Track		2 Tracks		3 Tracks		4 Tracks	
Track	Data	Model	Data	Model	Data	Model	Data	Model
1	51.84	50.93	45.95	46.07	44.92	43.57	45.40	42.91
2			51.98	56.43	51.38	51.24	51.44	48.88
3					55.62	57.67	51.45	55.43
4							57.99	58.85

Table 8: School effort by track and number of tracks

	1 Track		2 Tracks		3 Tracks		4 Tracks	
Track	Data	Model	Data	Model	Data	Model	Data	Model
1	1.86	1.85	1.75	1.84	1.75	1.82	1.82	1.81
2			1.90	1.86	1.88	1.86	1.84	1.86
3					1.96	1.86	1.93	1.86
4							1.68	1.85

Table 9: Parent effort by track and number of tracks

	1 Track		2 Tracks		3 Tracks		4 Tracks	
Track	Data	Model	Data	Model	Data	Model	Data	Model
1	2.07	2.15	2.31	2.53	2.57	2.70	2.29	2.85
2			2.03	1.96	2.37	2.24	2.71	2.47
3					2.11	1.90	2.78	2.11
4							2.08	1.93

6 Counterfactual Policy Evaluations

We use the estimated model to evaluate policy-relevant counterfactual scenarios. First, we quantify the effect of tracking by solving the model where we ban schools from tracking (hence all schools have only one track), and compare the changes in school effort, parental effort and student achievement. Our results indicate failing to account for equilibrium interactions between schools and parents could substantially bias the results.

We then examine the equilibrium effects of prospective changes in proficiency standards. We investigate the effects of adopting stricter performance standards by increasing regional proficiency cutoffs by 10% and 25%. Unlike the banning-tracking counterfactual, a school re-optimizes its tracking decision in both these counterfactual scenarios.

We contrast simulated outcomes between the baseline and each of the counterfactual policies. In particular, we present Average Treatment Effects (ATE) for subgroups of students defined by their prior test scores, by their parental education, and by single-parenthood.

	Parent effort		School effort		Outcome score	
	Data	Model	Data	Model	Data	Model
Low edu.	2.35	2.44	1.88	1.84	48.00	47.92
College	2.12	2.16	1.87	1.85	54.23	52.86
Single parent	2.37	2.33	1.90	1.84	48.76	47.70
Two parent	2.18	2.26	1.87	1.85	52.36	51.57
Low prior score	2.52	2.59	1.87	1.84	45.35	45.27
High prior score	1.77	1.97	1.88	1.85	57.96	56.39

6.1 Heterogeneous Effects of Tracking on Student Achievement

We compare the outcomes under the baseline with those when tracking is not allowed. Over 95% of the schools practice ability tracking to some degree in the baseline, hence are affected by this counterfactual. In response to this exogenous changes in peer quality within classrooms, these affected schools and the attendant parents adjust their effort inputs.

The below table reports ATE by decile of prior test score, where each row represents a decile.[15] The first column is the fraction of students in that decile who gain from the ban of ability tracking, which turns out to be a very small number (0.7%) for students in the highest decile of prior test scores. The second column reports the ATE by decile. The ATE of banning ability tracking is decreasing in ones prior score. The ATE for students with below-median prior scores is 0.39 points and that for students with above-median prior scores is -0.43. In particular, students with the lowest prior scores would on average gain 0.875 points (about 9% sd in test scores), while students in the top decile of prior scores would lose on average by 0.75 points, or 8% sd.[16]

These results follow the fact that peer effects have a modest, though positive, effect on student achievement. Banning tracking places all students in a school in one track, which means that lower ability students are now with better students, on average, and are made better off through the technology, ceteris paribus. The opposite holds for higher ability students.

Figure 1 shows results for outcome test score, pass rate, parental effort, school

[15]The expected ATE of banning tracking over the whole population is -0.05 points, or about -0.5% sd in test scores.

[16]Standard deviation for outcome test score is 9.40 points. Standard deviation for parent effort is 1.53 hours per week. Standard deviation for school effort is 0.57 hours per week.

	Frac. gain	ATE	ATE (losers)	ATE (gainers)
1	0.996	0.875	-0.409	0.880
2	0.914	0.524	-0.237	0.596
3	0.813	0.350	-0.304	0.500
4	0.699	0.141	-0.388	0.368
5	0.650	0.041	-0.415	0.287
6	0.464	-0.111	-0.409	0.233
7	0.198	-0.333	-0.459	0.176
8	0.158	-0.413	-0.527	0.193
9	0.039	-0.560	-0.587	0.112
10	0.007	-0.750	-0.758	0.346

Table 10: ATE: Outcome score by decile of prior score, banning tracking

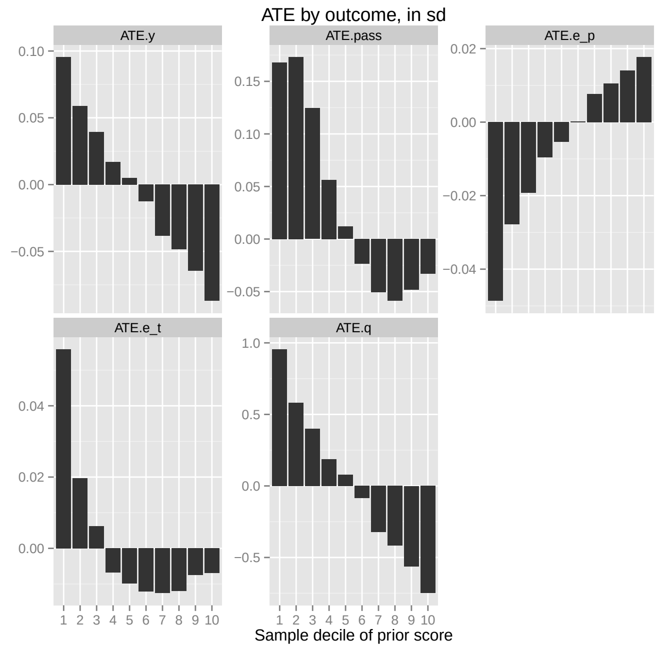
	Baseline	CF	ATE
1	0.140	0.169	0.029
2	0.316	0.346	0.030
3	0.461	0.482	0.021
4	0.631	0.640	0.009
5	0.732	0.734	0.002
6	0.819	0.815	-0.004
7	0.915	0.906	-0.009
8	0.938	0.929	-0.010
9	0.970	0.963	-0.008
10	0.988	0.983	-0.005

Table 11: ATE: Pass rates by decile of prior score, banning tracking

effort, and peer quality by decile of prior score. Students in the highest decile suffer the most when tracking is banned, and their parents increase the provision of costly effort the most in response.

Given that the technology plays an important role in evaluating the effect of tracking on student outcomes, one might ask whether an estimate of the technology, as opposed to estimation of the equilibrium model, would be sufficient to characterize tracking outcomes. Table 12 reports changes in parent effort responses by decile of prior achievement. Parents of students in the highest decile increase their inputs by the largest amount when tracking is banned, by over 2% sd. Given the estimated productivity of parental effort, ignoring parental effort responses may drastically overstate the negative effects of banning tracking on the low-ability students. Holding the

Figure 1: Change in outcomes due to banning tracking, by decile prior score



level of school effort constant at its average, were we to ignore the remediating effect of parent effort increases for above-median prior score students, we would find that banning tracking decrease test scores for these students by more than 0.25 *additional points* (or 2.5% sd) lower if we computed the change in test score produced holding parental effort at the tracking-equilibrium levels. This means one would overstate the negative effect of tracking on students with below-median prior score by 60%.

6.2 Changing Proficiency Standards

The second policy evaluation examines how changing proficiency standards would affect the distribution of achievement. We increase proficiency bars by 10% and by 25%.

Increasing standards by 10% causes an average gain of 0.02 points. Students with below-median prior achievement – those less likely to already be above the proficiency standard – on average experience a loss of 0.02 points, while those with above-median prior scores on average receive 0.06 higher outcome test scores.

Table 14 shows that the ATE is increasing in decile of prior test score, except for students in the highest decile, who were already extremely likely to pass the proficiency cutoff anyways.

	Frac inc. ep	Avg. inc. ep
1	0.004	-0.071
2	0.050	-0.041
3	0.122	-0.028
4	0.251	-0.014
5	0.304	-0.008
6	0.432	0.000
7	0.763	0.012
8	0.809	0.016
9	0.939	0.021
10	0.993	0.026

Table 12: ATE: Parent effort changes by decile of prior score

	Less than college	College
Two-parent hh	-0.014	-0.003
Single-parent hh	-0.021	-0.008

Table 13: ATE: Parent effort changes by household characteristics, banning tracking

Figure 2 shows results for outcome test score, pass rate, parental effort, school effort, and peer quality by decile of prior score. Schools increase their effort provision and peer quality for students with above-median prior scores, tapering off increases for students with the highest prior scores. The parents of these students increase their provision of costly effort due to the complementarities between their effort and school effort.

Increasing standards even more, by 25%, however, does not produce similar results. Average achievement now decreases by about 0.01 points. Table 16 shows that gains are now negative for some students who benefitted from the 10% increase in standards. Because this counterfactual exercises poses a bar that is unattainable without costs that would exceeds the benefits, schools do not increase effort or substantially change tracking practices. Figure 3 shows results for outcome test score, pass rate, parental effort, school effort, and peer quality by decile of prior score. As opposed to the 10% increase in proficiency standards, schools now increase inputs (effort and peer quality) the most for students with the highest prior performance, who have a chance of passing the new, higher, proficiency bar.

Figure 2: Change in outcomes due to 10% higher bar, by decile prior score

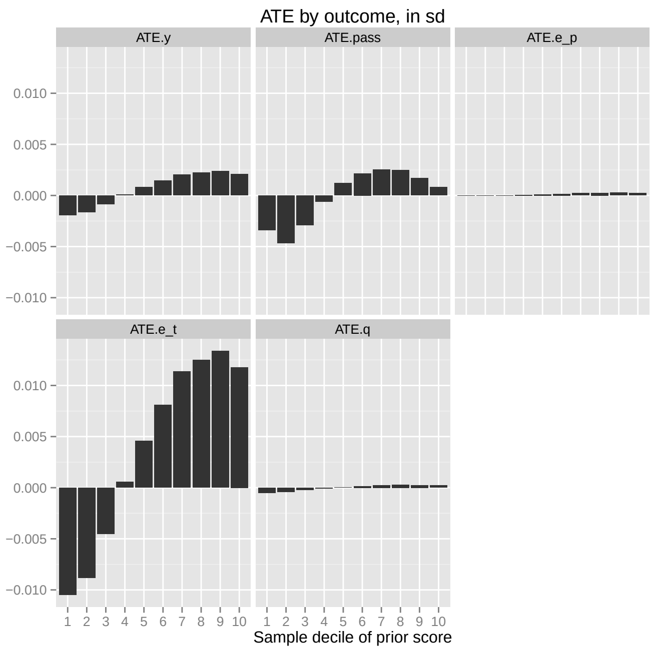
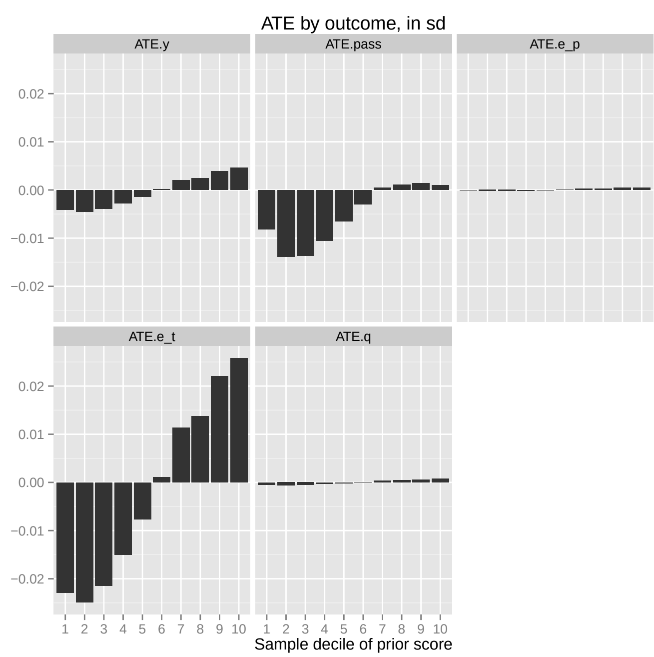


Figure 3: Change in outcomes due to 25% higher bar, by decile prior score



	Frac. gain	ATE	ATE (losers)	ATE (gainers)
1	0.029	-0.061	-0.064	0.031
2	0.111	-0.052	-0.062	0.026
3	0.284	-0.028	-0.052	0.033
4	0.455	0.001	-0.036	0.046
5	0.639	0.023	-0.023	0.050
6	0.860	0.043	-0.023	0.054
7	0.986	0.060	-0.057	0.061
8	0.986	0.066	-0.011	0.067
9	1.000	0.069		0.069
10	0.996	0.060	-0.023	0.060

Table 14: ATE: Outcome score by decile of prior score, 10% higher bar

	Baseline	CF	ATE
1	0.140	0.138	-0.002
2	0.316	0.313	-0.003
3	0.461	0.459	-0.002
4	0.631	0.631	-0.000
5	0.732	0.733	0.001
6	0.819	0.820	0.001
7	0.915	0.916	0.001
8	0.938	0.939	0.001
9	0.970	0.971	0.001
10	0.988	0.988	0.000

Table 15: ATE: Pass rates by decile of prior score, 10% higher bar

	Frac. gain	ATE	ATE (losers)	ATE (gainers)
1	0.000	-0.129	-0.129	
2	0.022	-0.140	-0.144	0.045
3	0.032	-0.124	-0.130	0.067
4	0.168	-0.085	-0.115	0.061
5	0.318	-0.044	-0.094	0.063
6	0.496	0.004	-0.081	0.091
7	0.737	0.062	-0.042	0.099
8	0.820	0.075	-0.027	0.098
9	0.957	0.117	-0.029	0.124
10	0.993	0.135	-0.088	0.137

Table 16: ATE: Outcome score by decile of prior score, 25% higher bar

	Baseline	CF	ATE
1	0.140	0.135	-0.005
2	0.316	0.307	-0.008
3	0.461	0.453	-0.008
4	0.631	0.625	-0.006
5	0.732	0.728	-0.004
6	0.819	0.817	-0.002
7	0.915	0.915	0.000
8	0.938	0.939	0.001
9	0.970	0.971	0.001
10	0.988	0.988	0.001

Table 17: ATE: Pass rates by decile of prior score, 25% higher bar

Appendix

A Functional Forms

A.1 Achievement Function and Cost Functions

Achievement:

$$\begin{aligned} Y(a,q,\upsilon,e^s,e^p,z_e,Q) &= \alpha_0+\alpha_1 a+\alpha_2 e^s+\alpha_3\check{e}^p+\alpha_4 q+\alpha_5 Q+\alpha_6 a e^s \\ &\quad +\alpha_7 a\check{e}^p+\alpha_8 e^s\check{e}^p+\alpha_9 e^s\upsilon, \\ \check{e}^p &= e^p z_e. \end{aligned} \tag{4}$$

Cost of parental effort:

$$C^P\left(e^p,z_c\right)=z_c e^p+c_2^p\left(e^p\right)^2.$$

Cost of school effort:

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

A.2 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 $\left(a_l^s, l=1,2,3\right)$. Let T_l^s be the l^{th} tercile of prior test scores among all students in school s $\left(\{x_{si}^a\}_i\right)$. A level a_l^s is defined as the average prior scores with the l^{th} tercile in school s, i.e.,

$$\begin{aligned} a_1^s &= \sum \frac{I\left(x_{si}^a\le T_1^s\right)x_{si}^a}{I\left(x_{si}^a\le T_1^s\right)}, \\ a_2^s &= \sum \frac{I\left(T_1^s< x_{si}^a\le T_2^s\right)x_{si}^a}{I\left(T_1^s< x_{si}^a\le T_2^s\right)}, \\ a_3^s &= \sum \frac{I\left(x_{si}^a> T_2^s\right)x_{si}^a}{I\left(x_{si}^a> T_2^s\right)}. \end{aligned}$$

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

$$P\left(\left(a_l^s, z\right)|x, s\right) = \Pr\left(a = a_l^s|x^a, s\right)\Pr\left(z_c|x^p, a_l^s\right)\Pr\left(z_e|x^p, a_l^s\right).$$

In particular, ability distribution is given by

$$\begin{aligned}
\Pr\left(a = a_1^s|x^a, s\right) &= 1 - \Phi\left(\frac{x^a - T_1^s}{\sigma_a}\right) \\
\Pr\left(a = a_3^s|x^a, s\right) &= \Phi\left(\frac{x^a - T_2^s}{\sigma_a}\right) \\
\Pr\left(a = a_2^s|x^a, s\right) &= 1 - \Pr\left(a = a_1^s|x^a\right) - \Pr\left(a = a_3^s|x^a\right),
\end{aligned}$$

where σ_a is a parameter to be estimated. Parental type distribution is given by

$$\begin{aligned}
\Pr\left(z_c = z_{c1}|x^p, a_l^s\right) &= \Phi\left(\theta_0^c + \theta_1^c a_l^s + \theta_2^c I\left(x_1^p \geq \text{college}\right) + \theta_3^c I\left(x_2^p = \text{single parent}\right)\right) \quad (5) \\
\Pr\left(z_c = z_{c2}|x^p, a_l^s\right) &= 1 - \Pr\left(z_c = z_{c1}|x^p, a_l^s\right),
\end{aligned}$$

and

$$\begin{aligned}
\Pr\left(z_e = z_{e1}|x^p, a_l^s\right) &= \Phi\left(\theta_0^e + \theta_1^e a_l^s + \theta_2^e I\left(x_1^p \geq \text{college}\right) + \theta_3^e I\left(x_2^p = \text{single parent}\right)\right) \quad (6) \\
\Pr\left(z_e = z_{e2}|x^p, a_l^s\right) &= 1 - \Pr\left(z_e = z_{e1}|x^p, a_l^s\right).
\end{aligned}$$

We restrict $\theta_2^c = \theta_3^p = 0$.

B Data details

Table 18: Proficiency cutoffs by Census region

Region name	Proficiency cutoff	Corresponding sample percentile
Northeast	49.69	42.34
Midwest	48.85	34.21
South	43.61	21.41
West	45.69	26.25

C Parameter Estimates

Table 19: Parameter Estimates

Production technology			
	α_0	-33.9468	intercept
	α_1	1.1119	own ability
	α_2	0.8309	school effort
	α_3	8.8416	parent effort
	α_4	0.2265	track peer quality
	α_5	-0.0205	school peer quality
	α_6	0.0118	interaction: ability and school effort
	α_7	0.0277	interaction: ability and parent effort
	α_8	0.8985	interaction: school and parent effort
	α_9	-0.0775	interaction: school effort and CV track ability
Parent objective and types			
	c_2^p	0.0583	quadratic parent effort cost
	z_{c1}	0.0908	linear parent effort cost, low cost type
	z_{c2}	0.1099	linear parent effort cost, high cost type
	θ_0^c	-21.1039	cost type intercept
	θ_1^c	0.2799	cost type, ability
	θ_2^c	1.7712	cost type, single parent indicator
	θ_3^c	0.0000	cost type, college indicator
	z_{e1}	1.0000	low efficiency level, normalized to one
	z_{e2}	1.1323	high efficiency level
	θ_0^e	-25.7799	efficiency type, intercept
	θ_1^e	0.4412	efficiency type, ability
	θ_2^e	0.0000	efficiency type, single parent indicator
	θ_3^e	9.7199	efficiency type, college indicator
School objective			
	ω_1	0.1411	weight on passing proficiency standard
	ω_2	0.0068	weight on exceeding upper level achievement
	c_1^s	0.0000	linear school effort cost
	c_2^s	0.9406	quadratic school effort cost
	γ_1	0.0000	regime cost, 1 track
	γ_2	-1.0836	regime cost, 2 tracks
	γ_3	-0.5122	regime cost, 3 tracks
	γ_4	0.9972	regime cost, 4 tracks
Shocks			
	σ_a	4.0467	sd of shock in ability distribution
	σ_ϵ	5.2161	sd test score measurement error
	σ_{ϵ^s}	0.5601	sd school effort measurement error
	σ_{ϵ^p}	1.0000	sd parent effort measurement error
	λ	0.8516	1 within nest correlation, regime pref. shock

References

Laura M Argys, Daniel I Rees, and Dominic J Brewer. Detracking america's schools: Equity at zero cost? *Journal of Policy analysis and Management*, 15(4):623–645, 1996. 4

J.R. Betts. The economics of tracking in education. *Handbook of the Economics of Education*, 3:341–381, 2011. 4

Julian R Betts and Jamie L Shkolnik. The effects of ability grouping on student achievement and resource allocation in secondary schools. *Economics of Education Review*, 19(1):1–15, 2000. 4

Martin Carnoy and Susanna Loeb. Does external accountability affect student outcomes? a cross-state analysis. *Educational Evaluation and Policy Analysis*, 24(4): 305–331, 2002. 15

E.M. Caucutt. Educational vouchers when there are peer group effects – size matters. *International Economic Review*, 43(1):195–222, 2002. 4

Thomas S Dee and Brian Jacob. The impact of no child left behind on student achievement. *Journal of Policy Analysis and Management*, 30(3):418–446, 2011. 15

Esther Duflo, Pascaline Dupas, and Michael Kremer. Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya. *The American Economic Review*, 101:1739–1774, 2011. 4

D. Epple and R.E. Romano. Competition between private and public schools, vouchers, and peer-group effects. *American Economic Review*, 88(1):33–62, 1998. 4

D. Epple, E. Newlon, and R. Romano. Ability tracking, school competition, and the distribution of educational benefits. *Journal of Public Economics*, 83(1):1–48, 2002. 4

M.M. Ferreyra. Estimating the effects of private school vouchers in multidistrict economies. *The American Economic Review*, pages 789–817, 2007. 4

David N Figlio and Marianne E Page. School choice and the distributional effects of ability tracking: does separation increase inequality? *Journal of Urban Economics*, 51(3):497–514, 2002. 4

Adam Gamoran. The variable effects of high school tracking. *American Sociological Review*, pages 812–828, 1992. 4

Eric A Hanushek and Margaret E Raymond. Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, 24(2): 297–327, 2005. 15

Brian A Jacob. Accountability, incentives and behavior: The impact of high-stakes testing in the chicago public schools. *Journal of public Economics*, 89(5):761–796, 2005. 15

Stephen P Klein, Laura S Hamilton, Daniel F McCaffrey, Brian M Stecher, et al. *What do test scores in Texas tell us?* Rand Santa Monica, CA, 2000. 15

Daniel M Koretz and Sheila I Barron. *The Validity of Gains in Scores on the Kentucky Instructional Results Information System (KIRIS).* ERIC, 1998. 15

Robert L Linn. Assessments and accountability. *Educational researcher*, 29(2):4–16, 2000. 15

Nirav Mehta. Competition in public school districts: Charter school entry, student sorting, and school input determination. *Working paper*, 2013. 4

T.J. Nechyba. Mobility, targeting, and private-school vouchers. *American Economic Review*, 90(1):130–146, 2000. 4

Randall Reback, Jonah Rockoff, and Heather L Schwartz. Under pressure: Job security, resource allocation, and productivity in schools under nclb. Technical report, National Bureau of Economic Research, 2011. 15

Manyee Wong, Thomas D Cook, and Peter 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 (Documento de trabajo 09-11)*, 2009. 15

Vivian Yee. Grouping students by ability regains favor in classroom. *The New York Times*, June 2013. URL http://www.nytimes.com/2013/06/10/education/grouping-students-by-ability-regains-favor-with-educators.html?_r=0&pagewanted=print. 1