Ability Tracking, Parental Effort and Student Achievement*

Chao Fu† Nirav Mehta‡

Abstract

We develop and estimate an equilibrium model of ability tracking. A school allocates students into tracks based on their ability and then chooses track-specific inputs to test score production. Parents choose parental effort in response. We estimate the model using data from the ECLS-K. We use the estimated model to quantify the effects of allowing ability tracking, making proficiency standards more stringent, or punishing schools for having students below the proficiency bar – all three policies may change equilibrium tracking, school inputs, parental effort and student achievement.

1 Introduction

Ability tracking or streaming, the practice of grouping students based on prior achievement, is controversial yet pervasive ((Yee 2013)). It is controversial because the effects of ability tracking may be heterogeneous, affecting students of different abilities differently. It is pervasive because public schools are often endowed with heterogeneous sets of students and may want to create more homogeneous classroom environments to facilitate learning. Policymakers would like to know how ability tracking affects different types of students. They may also want to know how policy changes, such as increasing proficiency standards, or holding schools more accountable for students below a proficiency bar, would affect school tracking choices and student outcomes.

*We thank Yuseob Lee and George Orlov 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

Most research on ability tracking focuses on measuring how test scores (or other outcomes of interest) change when student peer groups are changed.[1] Effectively, they are studies of contextual peer effects where peer groups (classroom ability composition) are chosen by schools. Like other work on peer effects, authors in this literature typically focus on the inferential problem caused by the non-random assignment of peer groups (or tracking regimes). Ideally, the estimated effects of tracking recovered by such studies return a composite policy effect, combining estimates of the technology of educational production and school and household inputs choices to this technology.

Although these studies may be useful for learning how changes in peer composition may affect certain types of students, they cannot provide a comprehensive picture of the equilibrium effects of tracking or the above policy changes for three reasons. First, when considering a change in peer composition, the students must come from somewhere, and wherever they come from will in theory be affected by the re-allocation. Therefore, an analysis that treats a change in peer composition as exogenous may neglect to measure treatment effects for other groups of affected students. Second, without a theory of how school and parental inputs are chosen given a tracking regime, we cannot infer what inputs to educational production and the distribution of student achievement would be were tracking banned. These input choices are necessary to quantify how allowing schools to track affects the distribution of achievement, even were the technology of education production known. Third, without a theory of how schools choose tracking regimes, we cannot predict how tracking regime choices and subsequent school and parent input choices would change in response to a policy change.

In principle, one could quantify the effect of ability tracking on the distribution of academic achievement by running a randomized experiment where there are control schools that do not track, treatment schools that do track, and then compare outcomes for students with similar observable characteristics, such as prior achievement. One could also use an experiment to quantify the effects of policy changes such as making proficiency standards more stringent or holding schools more accountable for failing students. Practically, it is infeasible to run randomized control trials for every set of school characteristcs and alternative policy scenario. To address this issue, we develop a framework that includes interactions between schools and households and heterogeneity in both household characteristics and the distribution of treatment effects.

We develop and estimate a model that treats tracking regimes, track-specific school

[1]See (Betts 2011) for an extensive review of this literature.

efforts, parental effort and student outcomes as joint outcomes from a subgame perfect Nash equilibrium.[2] 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 costs of the parent in helping her child to learn, the latter two are private information of the household. A child's achievement depends on one's own ability, the efforts invested by the school and by one's parent, and the quality of one's peers. Our test score production technology allows outcome scores to depend on both a direct effect of peer quality and the distance between one's own ability and the quality of one's peers, the effect of which may vary by one's ability level. A parent maximizes her child's achievement by choosing parental effort in response to the track her child is assigned to, hence peer quality, and the effort invested by the school. The value of a school's objective increases with both the total achievement of its students and the fraction of students whose achievement satisfies a minimum requirement, and decreases with its effort. The value also varies across tracking regimes due to different costs associated with implementing these regimes. The school maximizes its objective by choosing a tracking regime and track-specific effort inputs, taking into account expected parental responses, where the expectation is taken over the distribution of unobserved household types.

Although we believe our work is the first to endogenize school tracking regime choices, (Epple, Newlon and Romano 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 be able to attract higher ability students who would otherwise attend private schools. There is also a related literature where authors develop equilibrium models to study sorting between schools and its effects on peer composition ((Caucutt 2002), (Epple and Romano 1998), (Ferreyra 2007), (Mehta 2013), (Nechyba 2000)). However, only (Caucutt 2002) and (Mehta 2013) endogenize school input choices. Our work differs from these papers because we focus on how peer composition is determined *within* public schools and also have rich data on household characteristics and parental inputs, which allows to model and estimate how parents respond to school choices.

We estimate our model using data from the Early Childhood Longitudinal Study - 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, teach-

[2] To the best of our knowledge, ours is the first paper to explicitly model a school's choice of tracking regime.

ers, 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 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. 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). We also observe class-specific teacher inputs, the expected hours of reading homework per week.

Using the estimated model, we conduct three counterfactual experiments. It is important to note that our framework naturally allows all of these policies to produce winners and losers due to the differential impact tracking regimes may have on students of differing abilities and parental backgrounds. In the first, we quantify how allowing for ability tracking affects the distribution of student test scores by comparing outcomes from the baseline model with simulated equilibrium outcomes when no schools are allowed to track students. By disallowing ability tracking, the vast majority of schools have tracks with different peer group composition, which subsequently affects school efforts and parental efforts. We find evidence of heterogeneous affects of tracking: students with below-median prior achievement on average gain 4.5% of a standard deviation in outcome test score when schools are allowed to track by ability. Students with above-median prior achievement on average lose 4.6% sd when schools are allowed to track by ability. We also find that ignoring the behavioral responses of parents by holding their input levels constant when banning ability tracking would overstate the gain from tracking by 675%.

In the second policy experiment, we quantify how tracking regimes will change with policies that impose stricter proficiency standards, such as No Child Left Behind (NCLB), and the effects of those changes on student performance on standardized tests. We do this by simulating the new equilibrium outcomes when changes are made to the minimum score required by NCLB and/or the minimum fraction of students that are required to pass these requirements. We find that achievement for students with below-median prior achievement stays constant, while that for students with above-median prior achievement increases slightly (1% sd). Schools adopt policies that benefit higher ability students, such as having larger tracks and choosing higher inputs in tracks with more high-ability students. Intuitively, students with low abilities are far enough below the bar that increasing effort is not worthwhile from the school's perspective.

In the third counterfactual, we examine how holding schools more accountable for students below the proficiency standard affects the distribution of achievement by quadrapuling the weight on the pass rate in school objectives, holding average school utility constant. Now students with below-median prior achievement have an average increase of 3.7% sd of outcome test scores, which is associated with a 50% sd increase in school inputs for those students.

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 discrete types. Types differ in 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. 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 efforts.
Stage 3: Student achievement 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 one's classmates (q_j), track-specific input $\left(e_j^s\right)$, parental effort $\left(e_i^p\right)$, parental efficiency (z_{ie}), governed by function $Y(a_i, q_j, e_j^s, e_i^p, z_{ie})$. Test score y_{ji} measures student achievement with noise $\epsilon_{ji} \backsim F_\epsilon(\cdot)$, such that

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

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:

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

where $C^p\left(e_i^p, z_{ic}\right)$ is the parent effort cost function. Denote the optimal parental choice $e^p(e_j^s, q_j, a_i, z_i)$ and the maximized utility $u(e_j^s, q_j, a_i, z_i)$.

2.5 School's Problem

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 $\mu = \{\mu_j(:)\}_j$.*

A school cares about the total test score of its students and the fraction of students who can pass a certain threshold grade y^*. The latter reflects the pressure due to, for example, NCLB. A school chooses a tracking regime and track-specific inputs, both of which are costly. The cost of effort spent on track j, governed by function $C^s\left(e_j^s, n_j\right)$, depends on the intensity of the effort $\left(e_j^s\right)$ and the size of the track (n_j). Given a tracking regime μ, the optimal choice of e^s conditional on μ solves the following problem

$$\max_{e^s \geq 0} \left\{ \begin{array}{c} E \int_i \left(\sum_j \left(y_{ji} + \omega I(y_{ji} > y^*) - C^s(e^s_j) \right) \mu_j(a_i) \right) di \\ s.t.\ y_{ji} = Y(a_i, q_j, e^s_j, e^p_{ji}, z_{ie}) + \epsilon_{ji} \\ e^p_{ji} = e^p(e^s_j, q_j, a_i, z_i) \\ n_j = \sum_a \mu_j(a)\, g_s(a) \\ q_j = \frac{1}{n_j} \sum_a \mu_j(a)\, g_s(a)\, a. \end{array} \right\}, \quad (2)$$

where we have normalized the importance weight on total test score to be 1 such that all benefits and costs are measured in terms of average test score. $I(\cdot)$ is the indicator function and ω is the extra importance attached to "passing the bar." 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, $(g_s(z|a))$. The first two constraints faced by the school 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) and $V_s(\mu)$ the maximized value. The optimal tracking regime is obtained via solving the following problem

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

where $D(\mu)$ is the function governing the cost of tracking regime, and $\eta_{s\mu}$ is the idiosyncratic shifter associated with regime μ for school s. M_s is the support of tracking regime for school s, which we specify 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^s_j, q_j, a_i, z_i), $e^{p*}(e^s_j, q_j, a_i, z_i)$ 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^s_j, q_j, a_i, z_i).[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] 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.

3 Empirical Implementation and Estimation

3.1 Further Empirical Specifications

3.1.1 Household Type

There are assumed to be 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 is correlated with observable household characteristics x, which includes a noisy measure of student ability and household demographics. Let $P\left((a^s, z)\,|x, s\right)$ be the distribution of (a^s, z) given x in school s.

Remark 1 *In the model, the only household-level information available to the school is student ability a, so it is not a constraint, per se, that tracking is based on a only. In reality, a school may also observe household variables x that are correlated with z. Our implicit assumption is that a school cannot discriminate between students based on x.*

3.1.2 Tracking Regime

The support of tracking regimes (M_s) is assumed to be finite and school specific. In particular, we assume that 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. Suppose school s has K classrooms, we assume that the size of a particular track can only take value from $\left\{0, \frac{1}{K}, \frac{2}{K}, ..., 1\right\}$. We also assume that the composition of a track cannot be "disjoint," so that a class cannot mix low-ability students with high-ability students without any middle-ability student. Subject to these two constraints, M_s contains all possible ways to allocate students across the K classrooms. In the case a track contains multiple classrooms $\left(n_j > \frac{1}{K}\right)$, the composition of students are identical across classrooms on the same track.

The permanent idiosyncratic shifter in the school objective function, $\eta_{s\mu}$, is assumed to follow a generalized extreme-value distribution (nested logit), where all regimes with the same number of tracks are nested as one group. Let $n(\mu) \in \{1, .., N\}$ denote the

[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.

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

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

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 errors, such that the observed efforts $(\widetilde{e})$ are given by

$$\begin{aligned}\widetilde{e}_j^s &= e^s\varepsilon_j^s\\ \widetilde{e}_i^p &= e_i^p\varepsilon_i^p,\end{aligned}$$

where $\varepsilon_j^s \sim N\left(0,\sigma^2_{\varepsilon^s}\right)$, $\varepsilon_i^p \sim N\left(0,\sigma^2_{\varepsilon^p}\right)$.

3.2 Estimation

The parameters Θ to be estimated include model parameters Θ^M and parameters Θ^ε that govern the distribution of measurement errors. The former $\left(\Theta^M\right)$ 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 $P((a,z)|x)$ of household type given observables.[6]

We estimate Θ via maximum likelihood (MLE). The parameter estimates maximizes the probability of observing the joint endogenous outcomes given the observed

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

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

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

exogenous variables x. The endogenous outcomes observed for school s include the tracking regime μ_s, track-specific school efforts $\left\{\widetilde{e}^s_{sj}\right\}_j$ and household-level outcomes: parental effort $\widetilde{e}^p_{si}$, the track to which the student is assigned to 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

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

where each part of the likelihood is as follows:
$l_{\mu_s}\left(\Theta^M\right)$ is the probability of observing the tracking regime, which depends on all model parameters Θ^M since every part of Θ^M affects a school's tracking decision, but it does not depend on Θ^ε,

$$l_{\mu_s}\left(\Theta^M\right) = \frac{\exp\left(V_s\left(\mu_s|X_s;\Theta^M\right)\right)}{\sum_{\mu\in M_s}\exp\left(V_s\left(\mu|X_s;\Theta^M\right)\right)}.$$

$l_{sj}\left(\Theta\backslash\Theta_D\right)$ is the contribution of the observed school effort $\left(\widetilde{e}^s_{sj}\right)$ on track j given the tracking regime μ_s. It depends on all Θ^M 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 is measured with error:

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

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

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

where $l_{si}\left((a,z)\,|\Theta_y, \Theta_{c^p}, \Theta^\varepsilon\right)$ 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;[7] 2) the contribution of the observed parental effort $\widetilde{e}_s i^p$ *given* peer quality and the model predicted school effort $e^s_{\tau_{si}}$ on track τ_{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.

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

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 cross section for each school.

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. 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.[8] Because the

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

[8] The question for reading classes is:" What is the reading ability level of this childs reading class, relative to the children in your school at this child's grade?"

1. Primarily high ability

ECLS-K sampling scheme follows many students at the same school we have the above information on classes for several classes at each school. Therefore, we can also make inferences about the proportion of students in low, mixed, and high ability classes. We focus on 5th grade math classes. 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.

We use information from the Achievement Results for State Assessments [9] because ECLS-K data do not obtain information on the passing score y^*. The national data provide the distribution of students by school-level pass rates. We then calibrate y^* to match the school-level pass rate distribution in our data with the national one. Our cutoff of 44.09 corresponds to the 18th percentile test score in our sample data. Were we to use restricted-use sample data we would know the state each school was in and could in principle find state-specific proficiency cutoffs. However, given that we have an analytic sample of 205 schools, there are not enough schools to populate that distribution in each state.

4.1 Descriptive Statistics

Table 1 shows the distribution of the number of for all schools, schools with lower ability spreads[10], and higher prior test score distribution[11]. Schools with less variation in student ability and higher levels of ability are less likely to use ability tracking – they have higher proportions with only one track (which means they are not splitting students by ability level) and lower proportions of of four tracks.

Tables 2 and 3 show that students in higher ability tracks have both higher average outcome test scores and a higher proportion of students passing the proficiency cutoff.

Tables 4 and 5 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)

2. Primarily average ability
3. Primarily low ability
4. Widely mixed ability

[9] `https://explore.data.gov/`

[10] This is the difference between the highest and lowest ability at that school, divided by the middle level ability.

[11] Defined as having a below-median fraction of students with below-median prior achievement.

	All schools	Lower ability spread	Higher prior test scores
% with 1 track	4.39	4.85	4.55
% with 2 tracks	37.07	37.86	40.91
% with 3 tracks	45.85	49.51	47.27
% with 4 tracks	12.68	7.77	7.27
Total	100.00	100.00	100.00

Table 1: Distribution of schools by number of tracks, overall and by certain characteristics

Track	2 tracks	3 tracks	4 tracks
1	47.11	44.25	42.95
2	53.27	50.65	48.35
3		55.01	48.56
4			54.32

Table 2: Average outcome test scores by track, by number of tracks at school

Track	2 tracks	3 tracks	4 tracks
1	56.68	46.67	42.8
2	82.62	74.79	65.32
3		86.49	68.33
4			93.5

Table 3: Percent of students passing the cutoff by track, by number of tracks at school

decrease.

Track	2 tracks	3 tracks	4 tracks
1	1.75	1.75	1.82
2	1.9	1.88	1.84
3		1.96	1.93
4			1.68

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

Track	2 tracks	3 tracks	4 tracks
1	2.39	2.49	2.17
2	2.06	2.34	2.58
3		2.11	2.48
4			1.94

Table 5: Average parent effort by track, by number of tracks at school

Table 6 shows that lower-educated parents, single-parent households, and students with lower prior achievement all have higher average levels of parental inputs and lower outcome achievement.

	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

Table 6: Parent effort and outcome test score by observed characteristics

5 Results

5.1 Parameters

Parameter estimates are in Appendix A.3.[12]

[12]These estimates are still preliminary.

5.2 Model Fit

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

- 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 ability.
- Teacher effort is increasing in track ability.
- Parent effort is decreasing in track ability.
- Parents with less education, single parents, and students with lower prior achievement all have higher parent effort levels and lower outcome test scores.

	All schools		Lower ability spread		Higher prior test scores	
	Data	Model	Data	Model	Data	Model
#track 1 %	4.39	8.09	4.85	8.61	4.55	8.17
#track 2 %	37.07	33.44	37.86	34.42	40.91	34.18
#track 3 %	45.85	45.77	49.51	45.04	47.27	45.32
#track 4 %	12.68	12.69	7.77	11.93	7.27	12.33

Table 7: Distribution of schools by number of tracks, overall and by certain characteristics

	2 Tracks		3 Tracks		4 Tracks	
Track	Data	Model	Data	Model	Data	Model
1	47.11	47.72	44.25	46.39	42.95	45.88
2	53.27	54.59	50.65	51.31	48.35	49.51
3			55.01	55.67	48.56	53.83
4					54.32	56.17

Table 8: Outcome test score by track and number of tracks

6 Counterfactual Policy Evaluations

We use the estimated model to evaluate several policy-relevant counterfactual scenarios. First, we quantify the effect of tracking by solving the model where we ban schools from

Track	Data	Model	Data	Model	Data	Model
1	1.75	1.82	1.75	1.81	1.82	1.8
2	1.9	1.87	1.88	1.86	1.84	1.84
3			1.96	1.88	1.93	1.87
4					1.68	1.88

Table 9: Teacher effort by track and number of tracks

Track	Data	Model	Data	Model	Data	Model
1	2.39	2.44	2.49	2.52	2.17	2.55
2	2.06	2.05	2.34	2.23	2.58	2.34
3			2.11	1.98	2.48	2.08
4					1.94	1.95

Table 10: Parent effort by track and number of tracks

tracking (hence all schools have one track only), and compare the changes in school effort, parental effort and student achievement. Our results indicate that equilibrium interactions between schools and parents are quantitatively important, and that failing to account for them could substantially bias results.

We then examine the effects of two prospective policies that have been suggested as ways to increase student achievement. In the second counterfactual simulation we examine how adopting stricter performance standards, which we model by increasing the pass criterion for a high-stakes test, might affect the distribution of achievement. In the third counterfactual we quantify how increasing school accountablity for failing students affects the distribution of test score achievement. Unlike the first, a school re-optimizes its tracking decision in the last two experiments.

For each scenario, we report the difference in the endogenous variable for student, under the current scenario and the baseline, which is the model evaluated at the estimated parameters where we allow for schools to choose their desired tracking regime (except for the first counterfactual). We then compute Average Treatment Effects (ATE) for certain subgroups by averaging across students with certain characteristics, such as prior test score, parental education, and number of parents in the student's household. We consider the endogenous variables of outcome test score and pass rate.

	Data e_p	Model e_p	Data e^s	Model e^s	Data y	Model y
Low edu.	2.35	2.38	1.88	1.84	48.00	47.23
College	2.12	2.14	1.87	1.84	54.24	53.77
Single parent	2.37	2.31	1.90	1.84	48.76	48.40
Two parent	2.18	2.22	1.87	1.84	52.37	51.72
Low prior score	2.52	2.47	1.87	1.83	45.35	47.23
High prior score	1.78	2.01	1.88	1.85	57.96	54.92

6.1 Heterogeneous Effects of Tracking on Student Achievement

The below table reports ATE by decile of prior test score, where each row represents a decile.[13] The first column is the fraction of students in that decile who gain from the counterfactual (in this case, banning ability tracking), relative to the baseline where schools make tracking decisions. Very few (2.2%) of students in the lowest decile of prior test score would benefit from their school moving from their endogeonous tracking equilibrium to a scenario where they are not allowed to track.

The second column reports the ATE by decile. We can see that students with the lowest prior achievement would on average lose 1.117 points (about 12% sd in test scores) if their schools moved to a non-tracking equilibrium.[14] As we look at students with higher prior achievement, they have increasing benefits to tracking, where the median student would be indifferent and students in the top decile of prior achievement would benefit on average by 0.790 points, or 8.4% sd. The ATE for students with below-median prior achievement is -0.42 points and that for students with above-median prior achievement is 0.43.

These results are not surprising, given the estimated parameters in test score technology. The effects of peers differ across students depending whether or not one is above or below class-average: students benefit/suffer from being above/below class-average. Banning tracking places all students in a school in one track, which means that lower/higher ability students are now necessarily below/above class average, and are made worse/better off through the technology, ceteris paribus.

Given that the technology plays an important role in evaluating the effect of tracking

[13] The expected ATE of banning tracking over the whole population is 0.01 points, or about 0.1% sd in test scores.

[14] 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	0.022	-1.117	-1.148	0.295
2	0.115	-0.519	-0.615	0.222
3	0.222	-0.331	-0.495	0.244
4	0.348	-0.128	-0.351	0.291
5	0.434	-0.015	-0.275	0.324
6	0.568	0.114	-0.218	0.367
7	0.771	0.306	-0.196	0.455
8	0.849	0.401	-0.256	0.517
9	0.921	0.530	-0.251	0.597
10	0.921	0.790	-0.477	0.898

Table 11: ATE by decile of prior achievement

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. The last two columns in the next table report changes in parent effort responses by decile of prior achievement. Parents of students in the lowest decile increase their inputs by the largest amount when tracking is banned, by about 20% sd. Given the estimated productivity of effort inputs, this implies that we would drastically overstate the negative effects of banning tracking on the low-ability students, without taking parent effort responses into account. For example, were we to ignore the remediating effect of parent effort increases for below-median prior score students, we would find that banning tracking decrease test scores for these students by a further 2.665 points (or 28% 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 achievement by 675%.

6.2 Increasing Performance Standards

Now we evaluate a policy where performance standards are increased to the sample median of 51.435, to mimic an increase in the stringency of standards. We can see that schools shift towards policies that benefit higher ability students by reducing the amount of tracking. For example, the fraction of schools that do not track increases from 8.19% to 8.44% and the fraction of schools that have four tracks drops from 12.29% to 12.20%.

Such a policy reduces the incentive for schools to focus on lower-ability students,

	Frac inc. ep	Avg. inc. ep
1	0.986	0.316
2	0.928	0.148
3	0.832	0.099
4	0.706	0.045
5	0.656	0.018
6	0.504	-0.009
7	0.268	-0.054
8	0.179	-0.076
9	0.090	-0.104
10	0.036	-0.166

Table 12: Parent effort changes by decile of prior achievement

	Less than college	College
Two-parent hh	0.046	-0.002
Single-parent hh	0.073	0.012

Table 13: Parent effort changes by household characteristics

while increasing the incentive to help higher ability students.

Students with below-median prior achievement would have about the same achievement as before, while those with above-median prior achievement would gain only about 1% sd in test scores. The next table shows that the fraction of students who gain from more stringent performance standards is increasing in decile of prior achievement, where all students in the top quintile of prior achievement would benefit. School effort for students in the top half of prior achievement increases by 0.07 hours per week, or 12.3% sd, while that for students with below-median prior achievement is essentially unchanged. Intuitively, students with the lowest prior achievement are so far below the cutoff that increasing costly effort is much less effective for them than it is for students

	Baseline	CF
% 1 track	8.19	8.44
% 2 tracks	32.23	32.32
% 3 tracks	47.29	47.04
% 4 tracks	12.29	12.20

Table 14: Distribution of number of tracks at schools

with higher prior achievement.

	Frac. gain	ATE	ATE (losers)	ATE (gainers)
1	0.129	-0.070	-0.086	0.036
2	0.437	-0.013	-0.048	0.033
3	0.599	0.008	-0.040	0.041
4	0.878	0.048	-0.037	0.060
5	0.953	0.065	-0.047	0.071
6	0.982	0.081	-0.011	0.083
7	0.989	0.096	-0.038	0.098
8	0.996	0.102	-0.016	0.102
9	1.000	0.110		0.110
10	1.000	0.110		0.110

Table 15: ATE by decile of prior achievement

Pass rates fall for students in the lowest decile of prior achievement, but only by a marginal amount.

	Baseline	CF	ATE
1	0.411	0.407	-0.004
2	0.603	0.602	-0.001
3	0.680	0.680	0.000
4	0.790	0.792	0.002
5	0.833	0.835	0.002
6	0.871	0.873	0.002
7	0.919	0.921	0.002
8	0.937	0.938	0.002
9	0.961	0.962	0.001
10	0.983	0.984	0.001

Table 16: Pass rates by decile of prior achievement

Tables 17 and 18 show that if we instead increase the proficiency standard to a score of 47, the effects are similar, though not quite as extreme. Students with below-median prior achievement gain on average 0.02 points (0.2% sd in test scores), while those with above-median prior achievement gain slightly more (0.4% sd in test scores).

	Frac. gain	ATE	ATE (losers)	ATE (gainers)
1	0.269	-0.018	-0.033	0.021
2	0.703	0.009	-0.016	0.020
3	0.882	0.019	-0.012	0.024
4	0.986	0.033	-0.013	0.034
5	0.996	0.038	-0.014	0.039
6	1.000	0.042		0.042
7	0.996	0.044	-0.004	0.045
8	1.000	0.045		0.045
9	1.000	0.045		0.045
10	1.000	0.041		0.041

Table 17: ATE by decile of prior achievement

	Baseline	CF	ATE
1	0.411	0.410	-0.001
2	0.603	0.603	0.000
3	0.680	0.681	0.001
4	0.790	0.792	0.001
5	0.833	0.834	0.001
6	0.871	0.872	0.001
7	0.919	0.920	0.001
8	0.937	0.937	0.001
9	0.961	0.961	0.001
10	0.983	0.983	0.000

Table 18: Pass rates by decile of prior achievement

6.3 Holding Schools More Accountable

Now we examine a policy that punishes schools for failing students, given the status-quo proficiency cutoff. We implement this by quadrupling the weight of the pass rate in the school's objective, where we reduce the weight on average test score to keep the average level of the school objective constant. We see that schools now adopt policies that favor students with lower prior performance, in contrast to the previous scenario, though the effects are small. Tracking increases, as opposed to the last counterfactual.

Students with below-median prior achievement would gain 3.7% sd in test scores, while those with above-median prior achievement would essentially receive the same test scores. School effort would increase by about 0.28 hours per week (50% sd) for

	Baseline	CF
1 track %	8.19	7.39
2 tracks %	32.23	31.87
3 tracks %	47.29	48.10
4 tracks %	12.29	12.63

Table 19: Distribution of number of tracks at schools

students with below-median prior achievement, while that for students above-median prior achievement would remain the same.

	Frac. gain	ATE	ATE (losers)	ATE (gainers)
1	1.000	0.530		0.530
2	0.996	0.414	-0.015	0.416
3	0.996	0.399	-0.073	0.400
4	0.961	0.285	-0.050	0.299
5	0.918	0.222	-0.042	0.246
6	0.867	0.171	-0.060	0.207
7	0.646	0.088	-0.072	0.175
8	0.559	0.053	-0.089	0.165
9	0.349	-0.031	-0.112	0.122
10	0.129	-0.227	-0.278	0.116

Table 20: ATE by decile of prior achievement

	Baseline	CF	ATE
1	0.411	0.440	0.029
2	0.603	0.626	0.024
3	0.680	0.702	0.021
4	0.790	0.804	0.013
5	0.833	0.842	0.009
6	0.871	0.877	0.006
7	0.919	0.922	0.003
8	0.937	0.938	0.002
9	0.961	0.961	0.000
10	0.983	0.983	-0.000

Table 21: Pass rates by decile of prior achievement

	Less than college	College
Two-parent hh	-0.079	-0.019
Single-parent hh	-0.113	-0.047

Table 22: Parent effort changes by household characteristics

Appendix

A Functional Forms

A.1 Achievement Function and Cost Functions

Achievement:

$$\begin{aligned} Y(a, q, e^s, e^p, z_e) &= \alpha_0 + \alpha_1 a + \alpha_2 e^s + \alpha_3 \check{e}^p + \alpha_4 q + \alpha_5 (a-q)^2 + \alpha_6 (a-q)^2 I(a>q) \\ &\quad + \alpha_7 a e^s + \alpha_8 a \check{e}^p + \alpha_9 e^s \check{e}^p + \alpha_{10} a^2, \\ \check{e}^p &= e^p z_e. \end{aligned} \tag{3}$$

Cost of parental effort:

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

Cost of school effort:

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

Cost of tracking regime:

$$D(\mu) = \gamma(|\mu|),$$

where $|\mu|$ is the number of tracks in regime μ.

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

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

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

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

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

$$
\begin{aligned}
\Pr\left(z_c = z_{c1} | x^p, a_l^s\right) &= \Phi\left(\theta_0^c + \theta_1^c a_l^s + \theta_2^c I\left(x_1^p \geq \text{college}\right) + \theta_3^c I\left(x_2^p = \text{single parent}\right)\right) \quad (4) \\
\Pr\left(z_c = z_{c2} | x^p, a_l^s\right) &= 1 - \Pr\left(z_c = z_{c1} | x^p, a_l^s\right),
\end{aligned}
$$

and

$$
\begin{aligned}
\Pr\left(z_e = z_{e1} | x^p, a_l^s\right) &= \Phi\left(\theta_0^e + \theta_1^e a_l^s + \theta_2^e I\left(x_1^p \geq \text{college}\right) + \theta_3^e I\left(x_2^p = \text{single parent}\right)\right) \quad (5) \\
\Pr\left(z_e = z_{e2} | x^p, a_l^s\right) &= 1 - \Pr\left(z_e = z_{e1} | x^p, a_l^s\right).
\end{aligned}
$$

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

A.3 Parameter Estimates

Production technology	
α_0	-81.6023
α_1	0.9355
α_2	8.962
α_3	26.1358
α_4	0.0737
α_5	-0.1428
α_6	0.2872
α_7	-0.0181
α_8	-0.1106
α_9	0
α_{10}	0.0058
Parent objective	
c_2^p	0.0396
Parent cost type	
z_{c1}	0.4576
z_{c2}	0.4773
θ_0^c	-3.8478
θ_1^c	0.0628
θ_2^c	1.9693
Parent efficiency of effort type	
z_{e2}	1.1632
θ_0^e	-2.9098
θ_1^e	0.0049
θ_2^e	0
θ_3^e	4.7961
School objective	
ω	0.8482
c_1^s	0
c_2^s	0.3809
δ_1	-0.0023
δ_5	1.0133
$D(2)$	0
$D(3)$	0.1245
$D(4)$	1.3775
Shocks	
σ_a	9.065
σ_ϵ	6.5892
σ_{ϵ^s}	0.5563
σ_{ϵ^p}	1.4641
λ	1

References

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

Caucutt, E.M., "Educational Vouchers When There Are Peer Group Effects – Size Matters," *International Economic Review*, 2002, *43* (1), 195–222.

Epple, D. and R.E. Romano, "Competition Between Private and Public Schools, Vouchers, and Peer-Group Effects," *American Economic Review*, 1998, *88* (1), 33–62.

, E. Newlon, and R. Romano, "Ability Tracking, School Competition, and the Distribution of Educational Benefits," *Journal of Public Economics*, 2002, *83* (1), 1–48.

Ferreyra, M.M., "Estimating the Effects of Private School Vouchers in Multidistrict Economies," *The American Economic Review*, 2007, pp. 789–817.

Mehta, Nirav, "Competition in Public School Districts: Charter School Entry, Student Sorting, and School Input Determination," *Working paper*, 2013.

Nechyba, T.J., "Mobility, Targeting, and Private-School Vouchers," *American Economic Review*, 2000, *90* (1), 130–146.

Yee, Vivian, "Grouping Students by Ability Regains Favor in Classroom," June 2013.