Ability Tracking, Teacher Discretion and Student's Achievement

Cong Li

October 25, 2022

Motivation

Motivation

- Tracking as a low cost approach to improving students performance is crucial and pervasive(Glewwe and Muralidharan, 2016).
- Tracking regime (tracking time and tracking criteria) varies between countries.
  - ability tracking within school, in the US, China and Canada (Dieterle et al., 2015).
  - tracking across school (vocational or academic track), in Europe.
- Tracking effects on students performance are controversial (Betts, 2011).
  - proponents: more homogeneous group, tailor instruction $\Rightarrow$ more effective
  - opponents (de-tracking): mis-classification $\Rightarrow$ aggravates initial differences

Motivation

- In recent years, lots of work in understanding the determinants of skill formation in children (Heckman and Mosso, 2014).
- Peer composition is as important as teacher quality, class size, and parental involvement as a determinant of student achievement.
- The debate over standardized testing and the mix of standardized measures and discretionary evaluations (Diamond and Persson, 2016).
  - teachers know more about students $\Rightarrow$ reduce the effects of "a bad test day"
  - equal criteria $\Rightarrow$ increase fairness in some way

Research Question

1. (Tracking effects) What is the impact of tracking within a school on students' academic performance?
   - + teacher discretion
   - + multiple tracking
2. (Skill formation function) How is a student's skill determined at different ages, taking into account innate ability, peer influences, and teacher's instructional level?
3. (Policy relevant questions or counterfactual) To achieve the policy goal, how to design an "optimal" tracking rule?
   - tracking frequency
   - tracking time
   - tracking criteria

Contribution

1. (Empirical part) I collect the novel, rich and detailed panel data from a single high school in China, and use the RD design, to evaluate the tracking effects.
2. (Theoretical part) I am the first one to develop and estimate a new dynamic skill production model, that simultaneously endogenizes peer effects and teachers' mismatch effects.
3. (Policy implications) With the estimated model in hand, I use it to do some counterfactual analysis and simulation.
    - tracking rule (time + criteria)
    - increase/decrease the tracking frequency

Literature Review

- Tracking
  - Figlio and Page (2002), Fu and Mehta (2018), Duflo, Dupas, and Kremer (2011) (Tracking effects)
  - Meghir and Palme (2005), Piopiunik (2014), Dustmann, Puhani, and Schönberg (2017) (Tracking regime)
- Peer effects
  - Epple and Romano (2011), Sacerdote (2011)
- Teacher discretion
  - Diamond and Persson (2016), Apperson, Bueno, and Sass (2016),Lavy and Sand (2018)
- Skill production technology
  - Cunha and Heckman (2007, 2008), Cunha, Heckman, and Schennach (2010), Agostinelli and Wiswall (2016)

Motivating Example

	college	tier1 [1]	tier2	tier3
HHHH	**0.96**	**0.35**	**0.56**	**0.05**
LLHH/HHLL	0.89	0.06	0.78	0.11
LLLH/LLHL	0.85	0.02	0.57	0.26
LLLL	**0.81**	**0**	**0.37**	**0.42**

- After three years, there is a large difference in college and top 1 tier program admission rates between HHHH and LLLL. [unconditional comparison]

[1] Highest Quality

Motivating Example

	college	tier1(Highest)	tier2	tier3
H***	0.9	0.09	0.58	0.22
L***	0.86	0.06	0.5	0.29
X%	**5%**	**50%**	**16%**	**-24%**

two students, A and B

- similar entry score, at age 15
- A went to High, and B went to Low track
- A's college entrance excore is 0.3 sd[2] higher than B's, at age 18

[2]standard deviation is 93.13, and full credit is 750.

Background and Data Intro

Background

1. Education system in China
   - 9 years compulsory education
     - 6 years primary school + 3 years junior middle school

Background

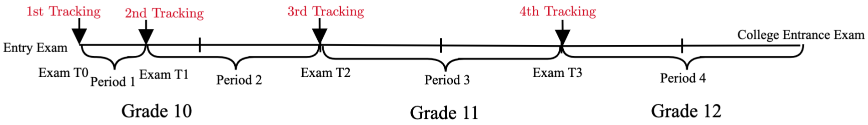
1. Education system in China
    - 9 years compulsory education
        - 6 years primary school + 3 years junior middle school
    - senior high school
        - pass the Senior High School Entrance Exam
        - three years
        - National College Entrance Exam (Gaokao) then college application
    - vocational school or join the labor force

Background

1. Education system in China
    - 9 years compulsory education
        - 6 years primary school + 3 years junior middle school
    - senior high school
        - pass the Senior High School Entrance Exam
        - three years
        - National College Entrance Exam (Gaokao) then college application
    - vocational school or join the labor force
2. Special tracking rule
    - four tracking over three years
    - teacher discretion

Background

1. Education system in China
   - 9 years compulsory education
     - 6 years primary school + 3 years junior middle school
   - senior high school
     - pass the Senior High School Entrance Exam
     - three years
     - National College Entrance Exam (Gaokao) then college application
   - vocational school or join the labor force
2. Special tracking rule
   - four tracking over three years
   - teacher discretion
   - Tracking timeline



Data Introduction

- Individual level panel data (individual - class track - time)
  - cohort 2017 (age 15 - 18)
  - class track trajectory
  - test score by subject of student i at class j in time t at period p: $s_{ij,pt}$
    - $i \in \{1, 2..., 1979\}, j =$ H or L
    - $p \in \{0, 1, 2, 3\}$
    - $t \in \{0, 1, ..., 27\}$: four tracking exams included
  - high school entrance exam score: s_{ij0}
  - characteristics: gender, age, urban/rural
- Census 2010 data

Transition Matrix and Class Track Trajectory

		Period 1	
		L	H
Period 0	L	857	151(15.0%)
	H	291(35.1%)	537

Transition Matrix and Class Track Trajectory

		Period 1 L	Period 1 H
Period 0	L	857	151(15.0%)
	H	291(35.1%)	537

- Tracking decision is made with more informative information.
  - Transition rates decrease over time. ▸ other periods

Transition Matrix and Class Track Trajectory

		Period 1 L	Period 1 H
Period 0	L	857	151(15.0%)
	H	291(35.1%)	537

- Tracking decision is made with more informative information.
  - Transition rates decrease over time. other periods
- Lots of persistence, but there are still turnovers
  - HHHH (27.65%) or LLLL (35.58%).
  - About 26.53% students changed once of class track type.
  - 9.35% students changed twice and about 0.89% with class track type HLHL or LHLH.

Empirical Evidence

The First Tracking (Non-discretionary)

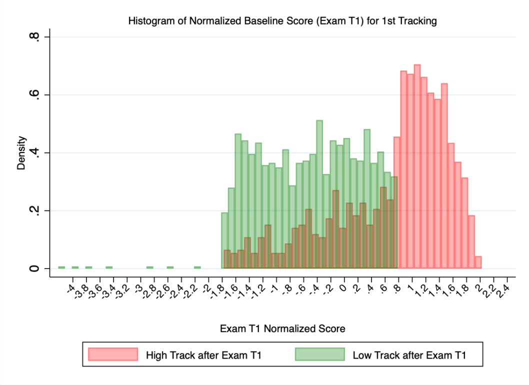
- At the beginning of Grade 10.
- One-side sharp cutoff.

$$Pr(j_1 = H) = \begin{cases} 1 & \text{if } s_0 > \text{cutoff} \\ p(s_0) & \text{if } s_0 \leq \text{cutoff} \end{cases}$$

The First Tracking (Non-discretionary)

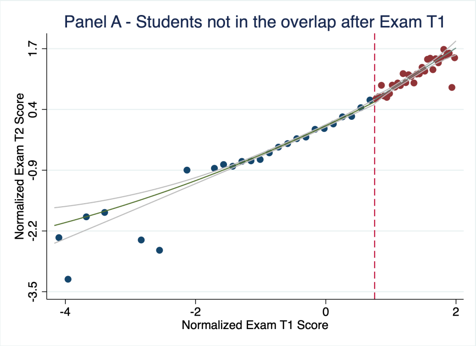
- At the beginning of Grade 10.
- One-side sharp cutoff.

$$Pr(j_1 = H) = \begin{cases} 1 & \text{if } s_0 > \text{cutoff} \\ p(s_0) & \text{if } s_0 \leq \text{cutoff} \end{cases}$$

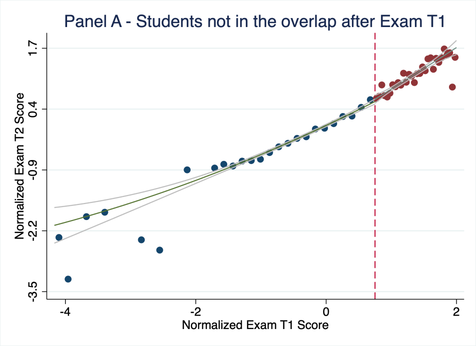


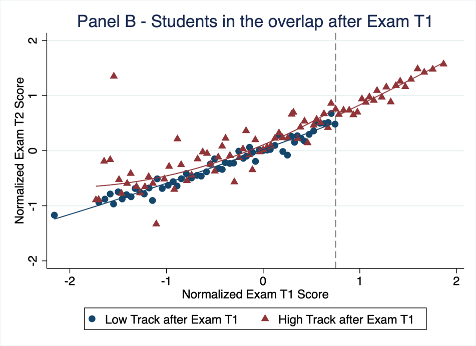
▸ Robustness check

RD Results of the First Tracking

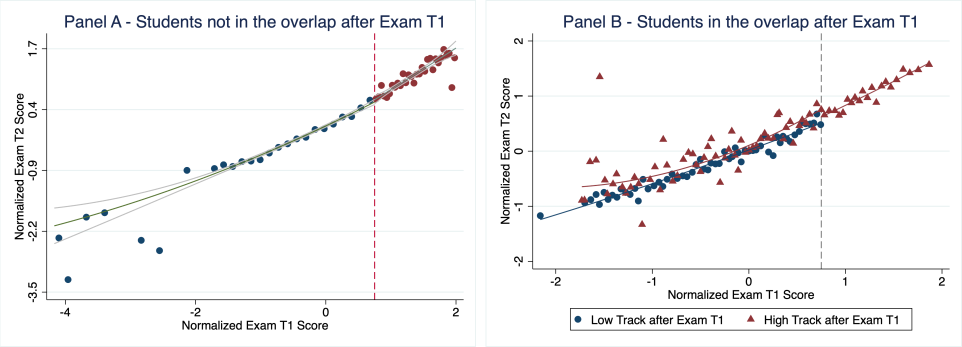


RD Results of the First Tracking





RD Results of the First Tracking



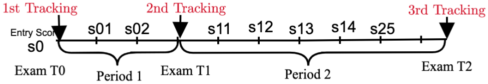
- No RD effects in first tracking.
- Students who are in high track with lower entry score performs slightly better than in low.

The Second Tracking (Discretionary)

- After a semester, teachers have a better understanding of their students.
- Tracking is not based on a score threshold and students' char. ▸ Robustness check

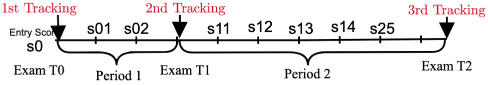
The Second Tracking (Discretionary)

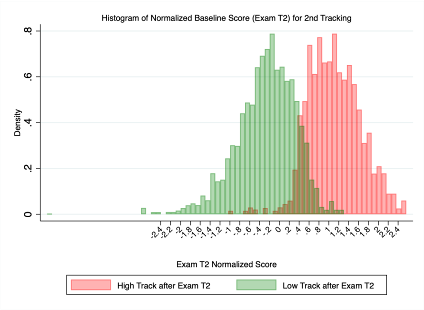
- After a semester, teachers have a better understanding of their students.
- Tracking is not based on a score threshold and students' char. ▸ Robustness check



The Second Tracking (Discretionary)

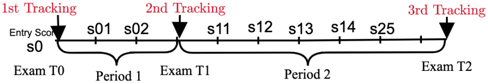
- After a semester, teachers have a better understanding of their students.
- Tracking is not based on a score threshold and students' char. ▸ Robustness check

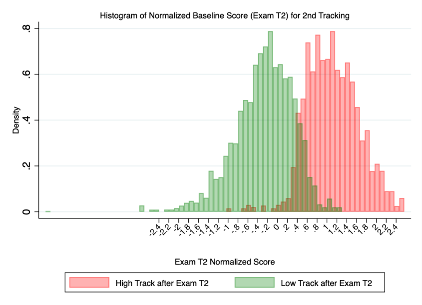


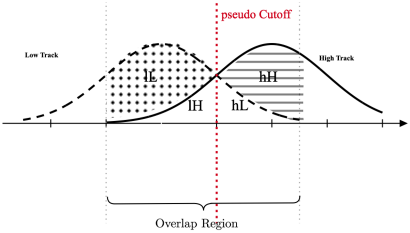


The Second Tracking (Discretionary)

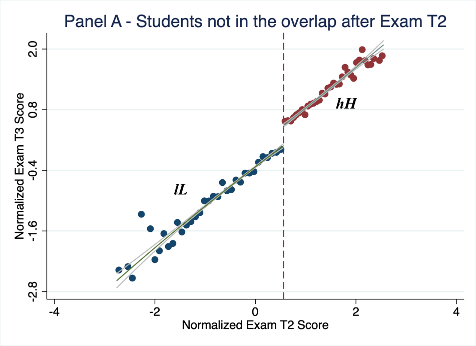
- After a semester, teachers have a better understanding of their students.
- Tracking is not based on a score threshold and students' char. ▸ Robustness check



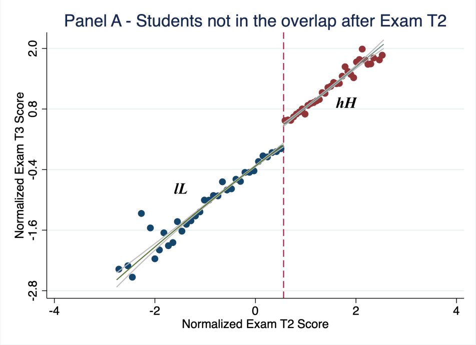


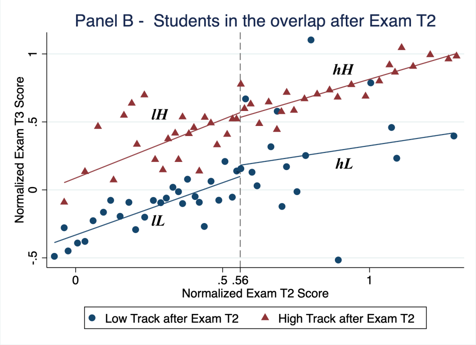


RD Results of the Second Tracking

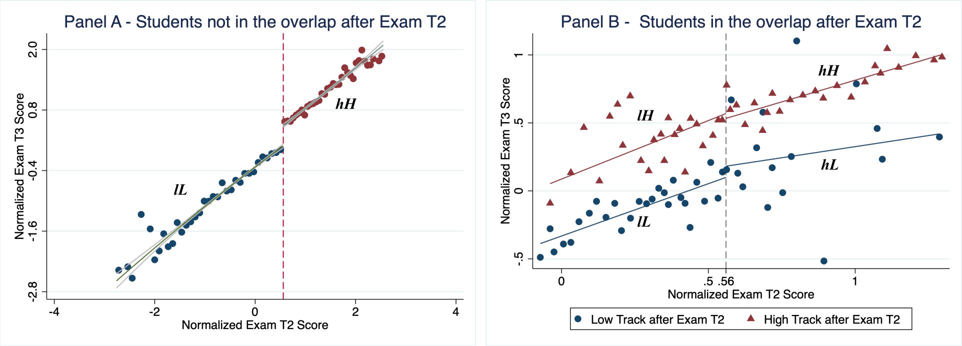


RD Results of the Second Tracking





RD Results of the Second Tracking



- Positive RD effects indicate that test score is not the only tracking criteria.
- Students with type lH performed much better than in low track.
- Exclusion of students in the overlap makes the effects under-estimated. ▸ RD results

Model and Identification Issues

From Empirical Patterns to the Model

Empirical observation

1. RD results from the first tracking
2. blurred tracking criteria, lower turnover rate overtime
3. class track trajectory and academic performance evolution

From Empirical Patterns to the Model

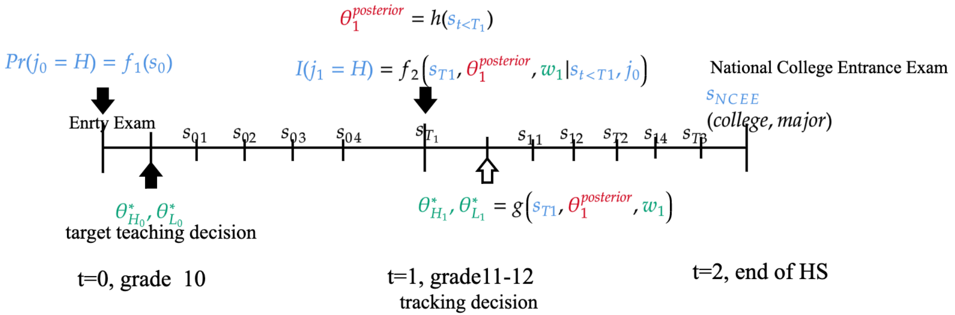
Empirical observation

1. RD results from the first tracking
2. blurred tracking criteria, lower turnover rate overtime
3. class track trajectory and academic performance evolution

Model feature

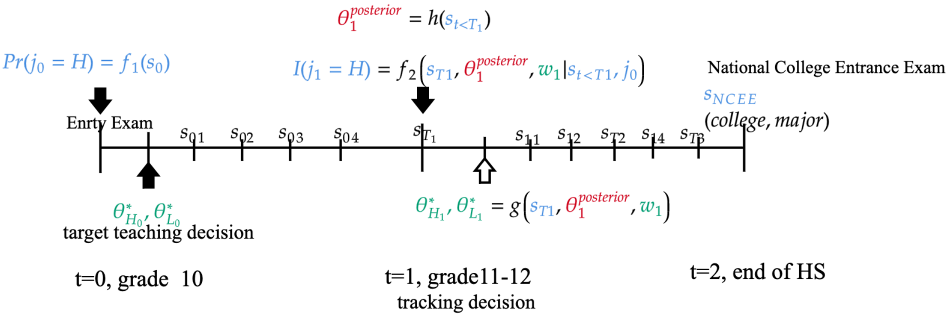
1. $+$ peer effects, $+$ teacher's target teaching level decision
2. teacher learning about students' ability and using discretionary tracking criteria
3. dynamic skill formation function

Timeline



Notation:

Timeline



Notation:

- $\theta_1^{posterior}$: teacher's posterior belief of students ability at period 1
- w_1: weight on the posterior belief in second tracking criteria
- $\theta^*_{H_0}, \theta^*_{L_0}, \theta^*_{H_1}, \theta^*_{L_1}$: target teaching level set by teachers in period 0, and period 1

parameters data unobservables

Timeline

There are three stages in this model.

- Stage 0 (grade 10): School allocates students to different tracks based on s_0. Teachers choose s_{j0}^* to maximize their expected payoff at the end of period 1. (Enrollment)

Timeline

There are three stages in this model.

- Stage 0 (grade 10): School allocates students to different tracks based on s_0. Teachers choose s^*_{j0} to maximize their expected payoff at the end of period 1. (Enrollment)
- Stage 1 (grade 11-12): Teachers update their belief learning from signals ($s_{t<T_1}$) and make the tracking decision based on θ_1^{pos} and s_{T_1}. Then, the new teachers choose s^*_{j1} based on new class composition. (Discretionary Tracking)

Timeline

There are three stages in this model.

- Stage 0 (grade 10): School allocates students to different tracks based on s_0. Teachers choose s^*_{j0} to maximize their expected payoff at the end of period 1. (Enrollment)
- Stage 1 (grade 11-12): Teachers update their belief learning from signals ($s_{t<T_1}$) and make the tracking decision based on θ_1^{pos} and s_{T_1}. Then, the new teachers choose $s^*_{j_1}$ based on new class composition. (Discretionary Tracking)
- Stage 2 (end of HS): Students take the National College Entrance Exam, apply for the college and find a job.

parameters data unobservables

Skill Formation Function

The framework follows the approach in (Cunha, Heckman, and Schennach, 2010) and (Agostinelli and Wiswall, 2016), adding peer effects and mismatch from $ln\theta^*$.

$$ln\theta_{i,t+1} = lnA_t + \overbrace{\gamma_{1,t} ln\theta_{i,t}}^{\text{self productive}} + \overbrace{\gamma_{2,t} ln\theta_{-i,t}}^{\text{peer effect}} \underbrace{- \gamma_{3,t} |ln\theta^*_t - ln\theta_{i,t}|}_{\text{teacher effect}} + \eta_{i,t} \tag{1}$$

Skill Formation Function

The framework follows the approach in (Cunha, Heckman, and Schennach, 2010) and (Agostinelli and Wiswall, 2016), adding peer effects and mismatch from $ln\theta^*$.

$$ln\theta_{i,t+1} = lnA_t + \overbrace{\gamma_{1,t} ln\theta_{i,t}}^{\text{self productive}} + \overbrace{\gamma_{2,t} ln\theta_{-i,t}}^{\text{peer effect}} \underbrace{-\gamma_{3,t} |ln\theta_t^* - ln\theta_{i,t}|}_{\text{teacher effect}} + \eta_{i,t} \quad (1)$$

lnA_t: TFP term

$\eta_{i,t}$: production shock including students' efforts, parental inputs etc, i.i.d. and $\perp\!\!\!\perp ln\theta_{i,t}$

$ln\theta_{i,t}$: student i's ability at time t

$ln\theta_{-i,t}$: average peer's ability, $ln\theta_{-i,t} = (\sum_{j=1}^{N} ln\theta_{j,t} - ln\theta_{i,t})/(N-1)$, N is class size

$ln\theta_t^*$: teachers' target teaching level

parameters data unobservables ▸ Justification

Measurement Function of $ln(\theta)$

$$s_{i,t,m} = \underbrace{\mu_{t,m}}_{\text{location}} + \underbrace{\lambda_{t,m}}_{\text{scale}} ln(\theta_{i,t}) + \epsilon_{i,t,m} \tag{2}$$

Measurement Function of $ln(\theta)$

$$s_{i,t,m} = \underbrace{\mu_{t,m}}_{\text{location}} + \underbrace{\lambda_{t,m}}_{\text{scale}} ln(\theta_{i,t}) + \epsilon_{i,t,m} \tag{2}$$

$\mu_{t,m}$: base score of measure m when no knowldege of m

$\lambda_{t,m}$: discrimination level of measure m

$\epsilon_{i,t,m}$: individual measurement error at time t and $E(\epsilon_{i,t,m}) = 0$ for all t, m

m: Chinese, Math, English, Chemistry, Biology and Physics test score

Measurement Function of $ln(\theta)$

$$s_{i,t,m} = \underbrace{\mu_{t,m}}_{\text{location}} + \underbrace{\lambda_{t,m}}_{\text{scale}} ln(\theta_{i,t}) + \epsilon_{i,t,m} \tag{2}$$

$\mu_{t,m}$: base score of measure m when no knowldege of m

$\lambda_{t,m}$: discrimination level of measure m

$\epsilon_{i,t,m}$: individual measurement error at time t and $E(\epsilon_{i,t,m}) = 0$ for all t, m

m: Chinese, Math, English, Chemistry, Biology and Physics test score

$$\underbrace{\widetilde{s_{i,t,m}}}_{\text{error contaminated}} = \frac{s_{i,t,m} - \mu_{t,m}}{\lambda_{t,m}} = ln(\theta_{i,t}) + \widetilde{\epsilon_{i,t,m}}$$

parameters data unobservables

Identification of Initial Conditions

Normalization: $\underbrace{E(ln(\theta_0)) = 0}_{\text{fix initial latent ability}}$ and $\underbrace{\lambda_{0,1} = 1}_{\text{normalize scale as 1}}$ $\Rightarrow s_{0,1} = \mu_{0,1} + ln(\theta_0) + \epsilon_{0,1}$

Identification of Initial Conditions

Normalization: $\underbrace{E(ln(\theta_0)) = 0}_{\text{fix initial latent ability}}$ and $\underbrace{\lambda_{0,1} = 1}_{\text{normalize scale as 1}}$ $\Rightarrow s_{0,1} = \mu_{0,1} + ln(\theta_0) + \epsilon_{0,1}$

Assumption 1:

$$\underbrace{Cov(\epsilon_{0,m}, \epsilon_{0,m'}) = 0}_{\text{uncorrelated measurement error bet various measures}} \quad , \text{ for all } m \neq m'$$

$$\underbrace{Cov(\epsilon_{0,m}, ln(\theta_0)) = 0}_{\text{measurement error is uncorrelated with ability}} \quad , \text{ for all } m$$

Identification of Initial Conditions

Normalization: $\underbrace{E(ln(\theta_0)) = 0}_{\text{fix initial latent ability}}$ and $\underbrace{\lambda_{0,1} = 1}_{\text{normalize scale as 1}}$ $\Rightarrow s_{0,1} = \mu_{0,1} + ln(\theta_0) + \epsilon_{0,1}$

Assumption 1:

$$\underbrace{Cov(\epsilon_{0,m}, \epsilon_{0,m'}) = 0}_{\text{uncorrelated measurement error bet various measures}} \quad , \text{ for all } m \neq m'$$

$$\underbrace{Cov(\epsilon_{0,m}, ln(\theta_0)) = 0}_{\text{measurement error is uncorrelated with ability}} \quad , \text{ for all } m$$

Then $\lambda_{0,m} = \frac{Cov(s_{0,m}, s_{0,m'})}{Cov(s_{0,1}, s_{0,m'})}, \mu_{0,m} = E(s_{0,m})$ for all $m \neq 1, m \neq m'$.

Identification of Initial Conditions

Normalization: $\underbrace{E(ln(\theta_0)) = 0}_{\text{fix initial latent ability}}$ and $\underbrace{\lambda_{0,1} = 1}_{\text{normalize scale as 1}}$ $\Rightarrow s_{0,1} = \mu_{0,1} + ln(\theta_0) + \epsilon_{0,1}$

Assumption 1:

$$\underbrace{Cov(\epsilon_{0,m}, \epsilon_{0,m'}) = 0}_{\text{uncorrelated measurement error bet various measures}} \quad , \text{ for all } m \neq m'$$

$$\underbrace{Cov(\epsilon_{0,m}, ln(\theta_0)) = 0}_{\text{measurement error is uncorrelated with ability}} \quad , \text{ for all } m$$

Then $\lambda_{0,m} = \frac{Cov(s_{0,m}, s_{0,m'})}{Cov(s_{0,1}, s_{0,m'})}, \mu_{0,m} = E(s_{0,m})$ for all $m \neq 1, m \neq m'$.

$$\widetilde{s_{0,m}} = \frac{s_{0,m} - \mu_{0,m}}{\lambda_{0,m}} = ln(\theta_0) + \widetilde{\epsilon_{0,m}} \tag{3}$$

parameters data unobservables

Sequential Identification

Assumption 2: (Production Function Restriction) $\underbrace{lnA_t = 0}_{\text{TFP is 0}}$, $\underbrace{\gamma_{1,t} + \gamma_{2,t} + \gamma_{3,t} = 1}_{\text{CRTS}}$

Sequential Identification

Assumption 2: (Production Function Restriction) $\underbrace{lnA_t = 0}_{\text{TFP is 0}}, \underbrace{\gamma_{1,t} + \gamma_{2,t} + \gamma_{3,t} = 1}_{\text{CRTS}}$

$$\begin{aligned} s_{1,m} &= \mu_{1,m} + \lambda_{1,m} ln\theta_1 + \epsilon_{1,m} \\ &= (\mu_{1,m} + \lambda_{1,m} lnA_0) + \lambda_{1,m}(\gamma_{1,0} ln\theta_0 + \gamma_{2,0} ln\theta_{-0} - \gamma_{3,0}|ln\theta_0^* - ln\theta_0| + \epsilon_{0,m}) + \epsilon_{1,m} \\ &= \beta_{0,0} + \beta_{1,0}\widetilde{s_{0,m}} + \beta_{2,0}(\sum_{j=1}^{N} \widetilde{s_{j,0,m}} - \widetilde{s_{0,m}})/(N-1) + \beta_{3,0}|ln\theta_0^* - \widetilde{s_{0,m}}| + \pi_{0,m} \end{aligned} \tag{4}$$

Sequential Identification

Assumption 2: (Production Function Restriction) $\underbrace{lnA_t = 0}_{\text{TFP is 0}}, \underbrace{\gamma_{1,t} + \gamma_{2,t} + \gamma_{3,t} = 1}_{\text{CRTS}}$

$$\begin{aligned}
s_{1,m} &= \mu_{1,m} + \lambda_{1,m} ln\theta_1 + \epsilon_{1,m} \\
&= (\mu_{1,m} + \lambda_{1,m} lnA_0) + \lambda_{1,m}(\gamma_{1,0} ln\theta_0 + \gamma_{2,0} ln\theta_{-0} - \gamma_{3,0}|ln\theta_0^* - ln\theta_0| + \epsilon_{0,m}) + \epsilon_{1,m} \\
&= \beta_{0,0} + \beta_{1,0}\widetilde{s_{0,m}} + \beta_{2,0}(\sum_{j=1}^{N} \widetilde{s_{j,0,m}} - \widetilde{s_{0,m}})/(N-1) + \beta_{3,0}|ln\theta_0^* - \widetilde{s_{0,m}}| + \pi_{0,m}
\end{aligned} \tag{4}$$

- Error in variables: $\pi_{0,m} \sim \epsilon_{0,m} \sim \widetilde{s_{0,m}} \Rightarrow$ IV based on multiple measures

Sequential Identification

Assumption 2: (Production Function Restriction) $\underbrace{lnA_t = 0}_{\text{TFP is 0}}$, $\underbrace{\gamma_{1,t} + \gamma_{2,t} + \gamma_{3,t} = 1}_{\text{CRTS}}$

$$\begin{aligned}
s_{1,m} &= \mu_{1,m} + \lambda_{1,m} ln\theta_1 + \epsilon_{1,m} \\
&= (\mu_{1,m} + \lambda_{1,m} lnA_0) + \lambda_{1,m}(\gamma_{1,0} ln\theta_0 + \gamma_{2,0} ln\theta_{-0} - \gamma_{3,0}|ln\theta_0^* - ln\theta_0| + \epsilon_{0,m}) + \epsilon_{1,m} \\
&= \beta_{0,0} + \beta_{1,0}\widetilde{s_{0,m}} + \beta_{2,0}(\sum_{j=1}^{N} \widetilde{s_{j,0,m}} - \widetilde{s_{0,m}})/(N-1) + \beta_{3,0}|ln\theta_0^* - \widetilde{s_{0,m}}| + \pi_{0,m}
\end{aligned} \tag{4}$$

- Error in variables: $\pi_{0,m} \sim \epsilon_{0,m} \sim \widetilde{s_{0,m}} \Rightarrow$ IV based on multiple measures
- Exactly identified system: $(\lambda_{1,m}, \mu_{1,m}, \gamma_{1,0}, \gamma_{2,0}, \gamma_{3,0}) \Rightarrow$ Assumption 2 +

$$\beta_{0,0} = \mu_{1,m} + \lambda_{1,m} lnA_0 \qquad \beta_{1,0} = \lambda_{1,m}\gamma_{1,0}$$

$$\beta_{2,0} = \lambda_{1,m}\gamma_{2,0} \qquad \beta_{3,0} = -\lambda_{1,m}\gamma_{3,0}$$

Learning Process of Student's Ability

Assumption 3: $\underbrace{ln(\theta_0) \sim N(0, \sigma^2_{\theta_0}), \epsilon_m \sim N(0, \sigma^2_{\epsilon_m})}_{\text{normal distribution of signal and noise}}$

Learning Process of Student's Ability

Assumption 3: $\underbrace{ln(\theta_0) \sim N(0, \sigma^2_{\theta_0}), \epsilon_m \sim N(0, \sigma^2_{\epsilon_m})}_{\text{normal distribution of signal and noise}}$

With Bayesian rule, $ln(\theta_0)|\widetilde{s_{0,m_s}}, s = 1, 2, ..., k_0 \sim N(\mu^{pos}, \sigma^{2,pos})$:

Learning Process of Student's Ability

Assumption 3: $\underbrace{ln(\theta_0) \sim N(0, \sigma^2_{\theta_0}), \epsilon_m \sim N(0, \sigma^2_{\epsilon_m})}_{\text{normal distribution of signal and noise}}$

With Bayesian rule, $ln(\theta_0)|\widetilde{s_{0,m_s}}, s = 1, 2, ..., k_0 \sim N(\mu^{pos}, \sigma^{2,pos})$:

$$\begin{aligned} \mu^{pos} &= \frac{\sigma^2_{\theta_0}}{\frac{\sigma^2_\epsilon}{k_0} + \sigma^2_{\theta_0}} \left(\frac{\sum_{s=1}^{k_0} \widetilde{s_{0,m_s}}}{k_0}\right) \\ \sigma^{2,pos} &= \left(\frac{1}{\sigma^2_{\theta_0}} + \frac{k_0}{\sigma^2_\epsilon}\right)^{-1} \end{aligned} \tag{5}$$

parameters data unobservables

Tracking Decision

- The First Tracking

$$Pr(j_0 = H) = \begin{cases} 1 & \text{if } s_0 > cutoff^1 \\ p(s_0) & \text{if } s_0 \leq cutoff^1 \end{cases}$$

Tracking Decision

- The First Tracking

$$Pr(j_0 = H) = \begin{cases} 1 & \text{if } s_0 > cutoff^1 \\ p(s_0) & \text{if } s_0 \leq cutoff^1 \end{cases}$$

- The Second Tracking ▸ More Evidence
  - weighted average of the posterior belief of ability (w_1) and tracking exam score $(1-w_1)$

$$Eln(\theta_1)^{adj} = w_1\mu^{pos} + (1 - w_1)\widetilde{s_{1,m}} \quad (6)$$

Tracking Decision

- The First Tracking

$$Pr(j_0 = H) = \begin{cases} 1 & \text{if } s_0 > cutoff^1 \\ p(s_0) & \text{if } s_0 \leq cutoff^1 \end{cases}$$

- The Second Tracking ▸ More Evidence
  - weighted average of the posterior belief of ability (w_1) and tracking exam score $(1-w_1)$

$$Eln(\theta_1)^{adj} = w_1 \mu^{pos} + (1 - w_1)\widetilde{s_{1,m}} \tag{6}$$

  - cutoff is determined by the seats of high track

$$I(j_1 = H) = \begin{cases} 1 & \text{if } Eln(\theta_1)^{adj} > cutoff^2 \\ 0 & \text{if } Eln(\theta_1)^{adj} \leq cutoff^2 \end{cases}$$

parameters data unobservables

Target Teaching Level Decision

$$\max_{ln\theta^*_{j,t}} EE_{\eta_{j,t}} P(ln\theta_{ij,t+1}) = \max_{ln\theta^*_{j,t}} EE_{\eta_{j,t}} (ln\theta_{ij,t+1})^{\alpha}$$

- rational expectation of skill formation function

Target Teaching Level Decision

$$\max_{ln\theta^*_{j,t}} EE_{\eta_{j,t}} P(ln\theta_{ij,t+1}) = \max_{ln\theta^*_{j,t}} EE_{\eta_{j,t}} (ln\theta_{ij,t+1})^{\alpha}$$

- rational expectation of skill formation function
- α : convexity of payoff function ($\alpha = 2$)
- no closed-form solution for $ln\theta^*_j$

$$ln\theta^*_j \begin{cases} = \overline{ln\theta_{ij,t-1}} & \text{if } \alpha = 1 \\ > \overline{ln\theta_{ij,t-1}} & \text{if } \alpha > 1 \\ < \overline{ln\theta_{ij,t-1}} & \text{if } \alpha \in (0,1) \end{cases}$$

Identification Summary

1. Two-dimensional skill production technology (γs) + measurement equation (λs, μs)

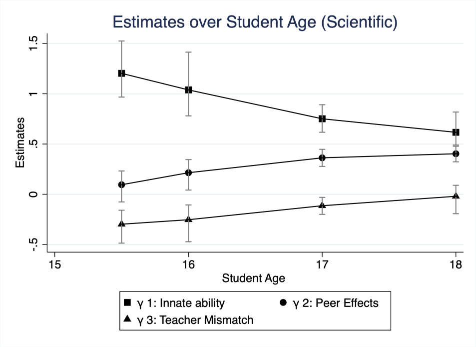
    - CRTS assumption
    - "Guess-and-Verify" + sequential IV regression
        - multiple measures of two-dimension skills over multiple periods

Identification Summary

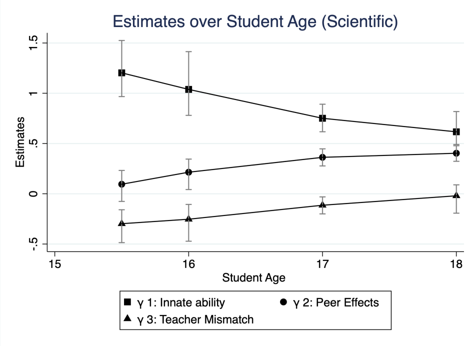
1. Two-dimensional skill production technology (γs) $+$ measurement equation (λs, μs)
    - CRTS assumption
    - "Guess-and-Verify" $+$ sequential IV regression
        - multiple measures of two-dimension skills over multiple periods
2. Tracking decision (w_1)
    - GMM
        - cross-sectional
        - tracking decision for student i
    - Clustered Bootstrap to get the standard error and confidence interval

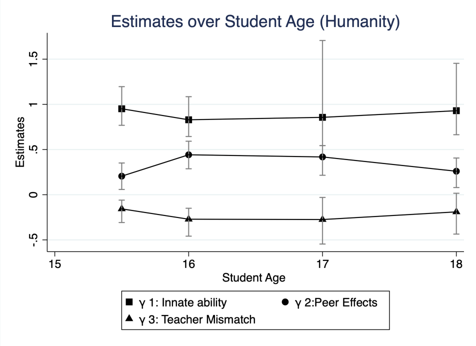
Results

Skill Formation Function Estimation

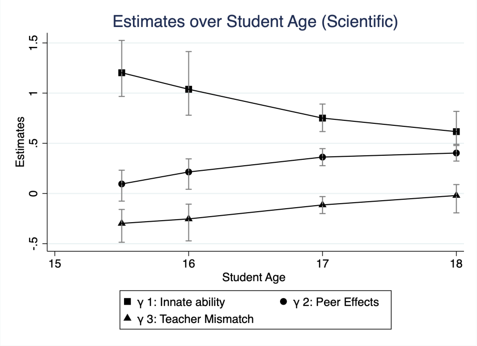


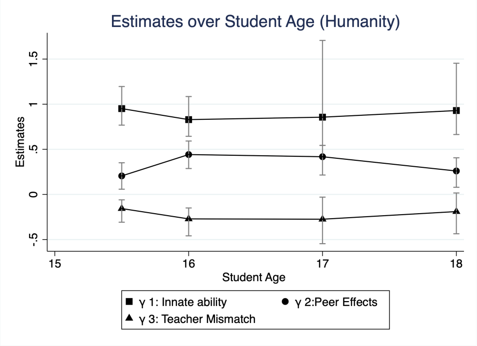
Skill Formation Function Estimation





Skill Formation Function Estimation





- Coefficients are changing over time.
- Scientific skill formation function is more malleable compared with Humanity.

Teacher's Decision Estimation

	Scientific Skill			
Model	Age 15.5	Age 16	Age 17	Age 18
Target Teaching	4.336	2.680	3.546	8.452
Low Track	(0.308)	(0.503)	(0.417)	(1.273)
	[3.79,5.01]	[2.14,3.95]	[3.01,4.76]	[7.39,9.58]
Target Teaching	12.336	21.786	24.439	31.076
High Track	(0.972)	(1.343)	(1.459)	(2.871)
	[10.76,14.38]	[19.25,25.94]	[21.02,27.34]	[27.18,37.05]
Weight	0	0.768	0.543	0.362
		(0.082)	(0.044)	(0.043)
		[0.65,0.87]	[0.43,0.62]	[0.26,0.41]

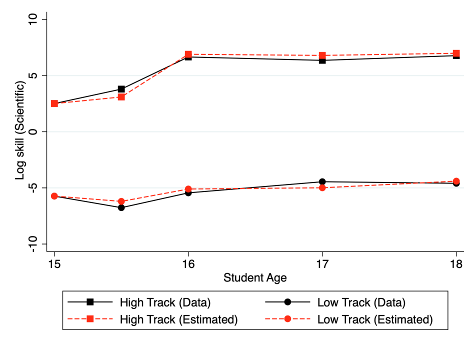
- Teachers adjust target teaching level based on the class composition at each period.
- Estimated target teaching level is higher than the average ability.
- Less weight put on teachers' belief of students' ability when tracking in later periods.

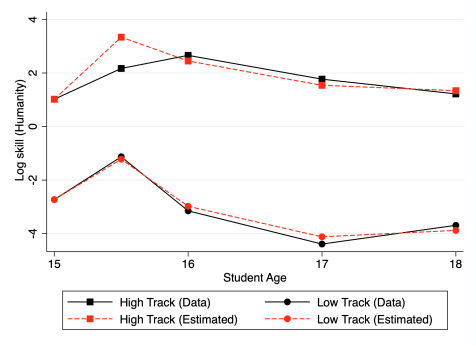
Measurement Equation Estimation

	Scientific Skill				
	Age 15	Age 15.5	Age 16	Age 17	Age 18
mu	90.663	78.752	79.979	74.482	67.558
		(1.188)	(1.350)	(0.829)	(1.618)
lambda	1	1.933	1.307	1.022	1.034
		(0.205)	(0.130)	(0.133)	(0.059)
	Humanity Skill				
mu	84.336	98.978	89.067	96.401	96.617
		(0.467)	(0.580)	(1.300)	(1.278)
lambda	1	1.239	1.155	0.982	0.955
		(0.136)	(0.125)	(0.195)	(0.078)

- μ of scientific skill is smaller in later periods.
- Humanity skill measurement is more stable.

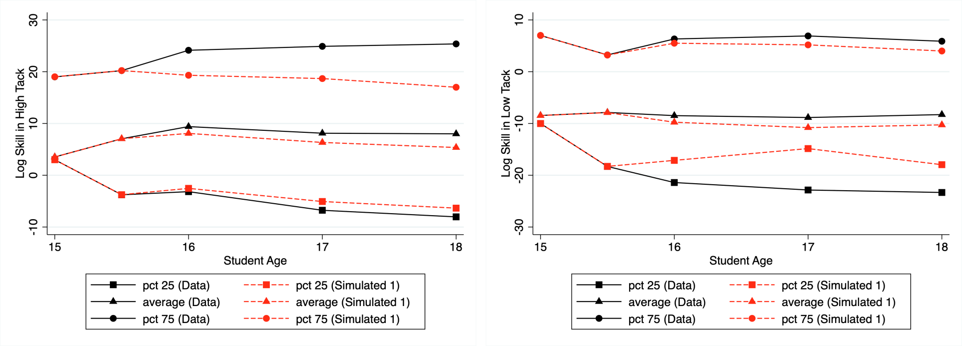
Estimated Skill Path





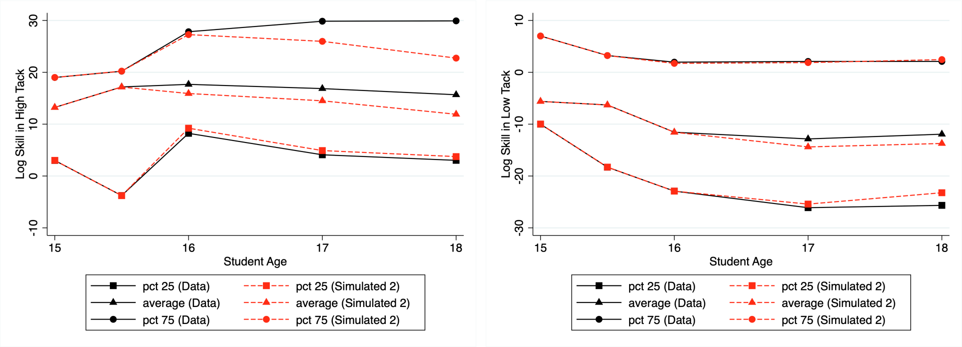
- Skill increased a lot in the first two periods.
- The Humanity skill didn't change much, even decreased.

One-Time Tracking



- Percentile 75 students performed worse, while percentile 25 students did better, and average students did worse in both tracks.
- Inequality decreased, but both track students did worse.

One-Time Discretionary Tracking



Conclusion

Conclusion and Discussion

- Discretionary tracking decreased the one-time tracking mis-allocation error, increased students' average ability, especially high-ability students, while increased students' inequality.
- Students' skill formation function is changing. Innate ability, peer effects, teachers mismatch play different roles over time.
- Teachers adjust target teaching level based on class composition and their payoff function.
- Scientific skill formation is more malleable than humanity skill.

Conclusion and Discussion

- Discretionary tracking decreased the one-time tracking mis-allocation error, increased students' average ability, especially high-ability students, while increased students' inequality.
- Students' skill formation function is changing. Innate ability, peer effects, teachers mismatch play different roles over time.
- Teachers adjust target teaching level based on class composition and their payoff function.
- Scientific skill formation is more malleable than humanity skill.

- Exclude the students'/teachers' effort input due to data limitation.
- Constant return to scale assumption of skill formation function.

Thank you

Appendix

Motivating Example

	college	tier1	tier2	tier3
H***	0.9	0.09	0.58	0.22
L***	0.86	0.06	0.5	0.29
X%	**5%**	**50%**	**16%**	**-24%**

	college	tier1	tier2	tier3
LLLL	**0.81**	**0**	**0.39**	**0.42**
LLHH/HHLL	0.89	0.02	0.58	0.28
LHHH/HLLL	0.95	0.08	0.76	0.11
HHHH	**0.96**	**0.27**	**0.66**	**0.02**

return

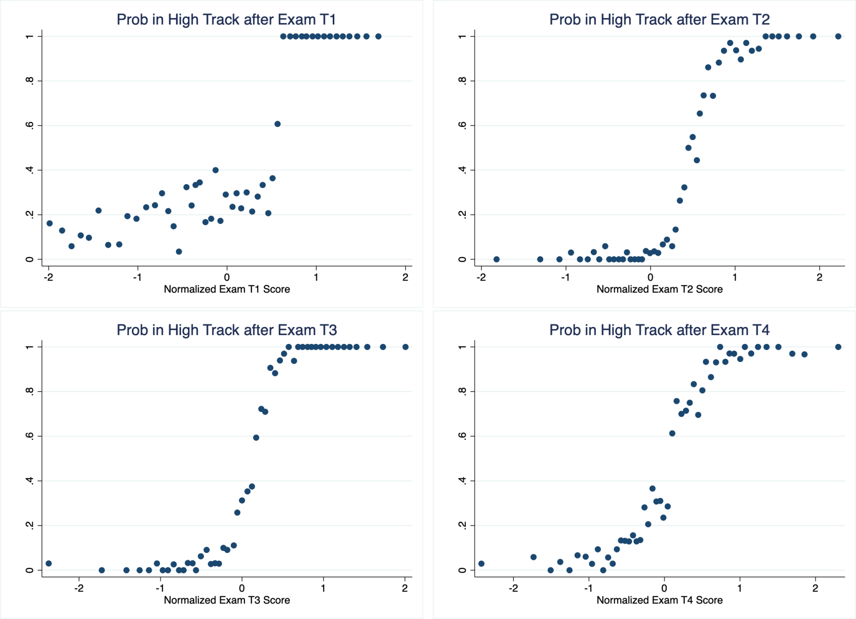
Transition Matrix

		period 3	
		L	H
period 2	L	920	198(17.7%)
	H	38(5.4%)	660

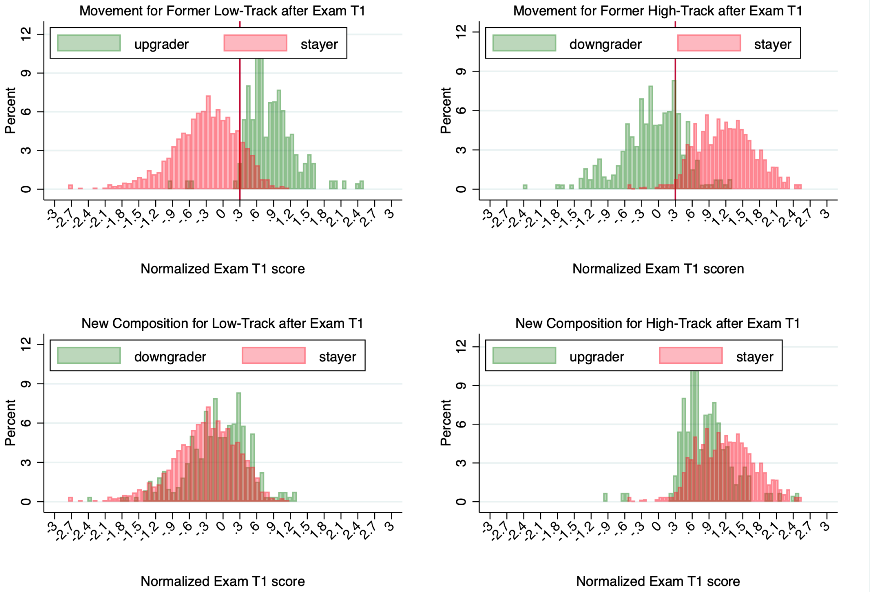
		period 4	
		L	H
period 3	L	865	91(9.5%)
	H	105(12%)	770

◂ return

Tracking Rule



Class Composition before/after the Second Tracking



First Tracking Criteria Check

$$log(\frac{prob_{i,H}}{1-prob_{i,H}}) = \beta_0 + \beta_1 s_{0ij} + \beta_2 I(urban_i = 1) + \beta_3 I(female_i = 1) + \epsilon_{i,t}$$

VARIABLES	(1) all students	(2) students in the overlap	$\beta_{all} - \beta_{overlap}$ z-score
s_{0ij}	0.494***	1.714***	-10.590
	(0.0887)	(0.0735)	
$I(urban_i = 1)$	0.625***	0.553***	0.388
	(0.141)	(0.121)	
$I(female_i = 1)$	-0.185	-0.125	-3.223
	(0.126)	(0.111)	
Constant	-1.478***	-0.863***	
	(0.145)	(0.124)	
Observations	1,613	2,207	

Standard errors in parentheses,*** $p<0.01$, ** $p<0.05$, * $p<0.1$

336/1277 are in high track with lower test score (smaller than cutoff 0.7507), 1044/2493 are in high track, 1449/2493 are in low track. return

Second Tracking Robustness Check

	students in the overlap					
	Assigned to low track			Assigned to high track		
	N	mean	sd	N	mean	sd
Urban ratio	637	0.642	0.480	557	0.704	0.457
Female ratio	637	0.540	0.499	557	0.490	0.500
Average standardized baseline score (mean 0, SD 1)	637	-0.407	0.433	557	0.425	0.437
Average standardized endline score (mean 0, SD 1)	627	-0.361	0.618	542	0.520	0.626
Previous section (high track =1)	637	0.133	0.340	557	0.871	0.336

return

RD Results

	(1)	(2)	(3)	(4)
VARIABLES	First Tracking	Second Tracking	Third Tracking	Fourth Tracking
RD_Estimate	-0.0600	0.496***	0.448***	0.388***
	(0.0631)	(0.0456)	(0.0397)	(0.0377)
Observations	4,396	9,761	11,176	17,933

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

RD estimates

return

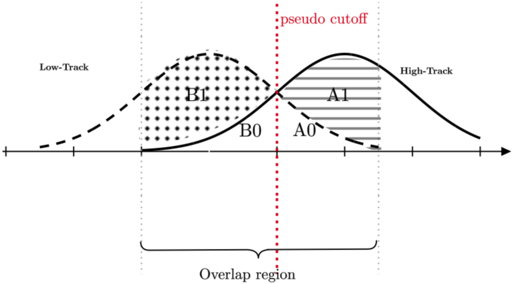
Summary of More Evidence

- As time goes on, tracking exam scores become less important, while prior class tracks become more influential in making tracking decision.
- Discretionary tracking in a corrective way.
  - predicted residue to show the unexplained parts
  - upgraders (higher), downgraders (lower).
- Teacher tailor instruction level – High track students performs much better in harder exams.

return

Treatment Effects Measured from Different Groups

$$s_{ijt} = \beta_0 + \beta_1 I(s_i^b > s_t^{cutoff}) + \beta_2 I(High_{ijt} = 1) + \beta_3 I(s_i^b > s_t^{cutoff}) \times I(High_{ijt} = 1) + controls + \epsilon_{ijt} \tag{7}$$



- $A0 : hL, A1 : hH, B0 : lH, B1 : lL$
- $\beta_0 : lL, (\beta_0 + \beta_1) : hL, (\beta_0 + \beta_2) : lH, (\beta_0 + \beta_1 + \beta_2 + \beta_3) : hH$
- controls : s^0, $I(urban_i = 1)$, $I(female_i = 1)$ when no student FE

Treatment Effects Measured from Different Groups

VARIABLES	(1) all Periods	(2) all Periods	(3) all Periods	(4) all Periods
$I(s_i^b > s_t^{cutoff})$	0.133***	0.140***	0.0373**	0.0434**
	(0.0274)	(0.0273)	(0.0184)	(0.0184)
$I(High_{ijt} = 1)$	0.0694*	0.111***	0.165***	0.163***
	(0.0379)	(0.0261)	(0.0182)	(0.0184)
$I(s_i^b > s_t^{cutoff}) \times I(High_{ijt} = 1)$	0.113***	0.106***	-0.0758***	-0.0702***
	(0.0429)	(0.0397)	(0.0244)	(0.0241)
Constant	-0.119***	-0.00865	-0.0948***	-0.136***
	(0.0140)	(0.0217)	(0.0111)	(0.0498)
$\beta_0(lL)$	-0.119	-0.00865	-0.0948	-0.136
$\beta_0 + \beta_1(hL)$	0.014	0.13135	-0.0575	-0.0926
$\beta_0 + \beta_2(lH)$	-0.0496	0.1024	0.0702	0.027
$\beta_0 + \beta_1 + \beta_2 + \beta_3(hH)$	0.1964	0.3484	0.0317	0.0004
Observations	28,223	28,223	28,216	28,216
R-squared	0.521	0.538	0.680	0.680
Student FE	NO	NO	YES	YES
Period FE	NO	YES	NO	YES

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

- $lH > hH > hL > lL$ after including student FE

Target Teaching Level

$$s_{ijt} = \beta_0 + \sum_{h=2}^{5} \beta_h I(hardlevel = h) I(High = 1) + ExamFE + controls + \epsilon_{ijt} \quad (8)$$

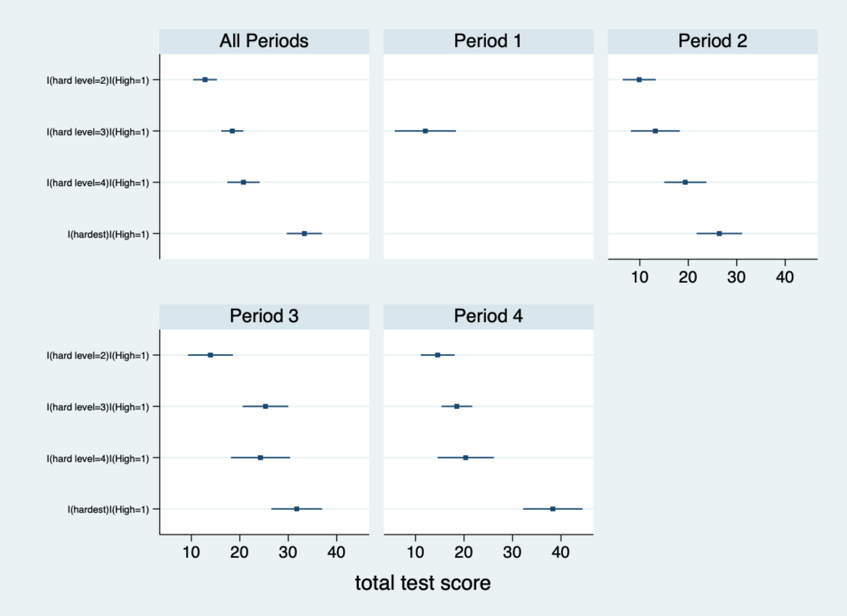
- Difficult level: the standard deviation of the raw test score

- level 5 is the hardest

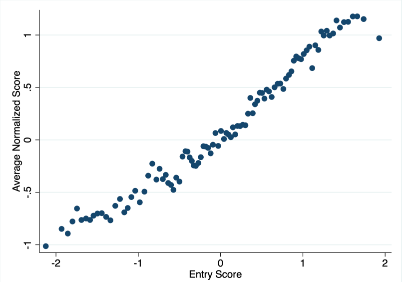
- Students in High track perform better in harder exams.

- controls: raw entry score, $I(urban_i = 1)$, $I(female_i = 1)$

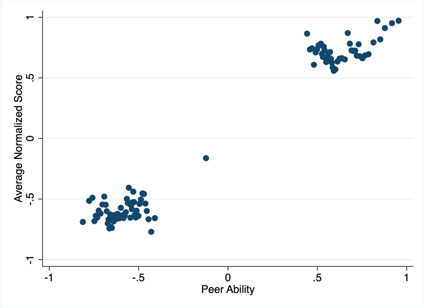
Target Teaching Level (β_h)

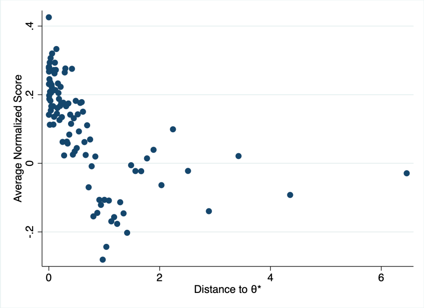


Correlation of Test Score vs Peers, Teachers, Initial Ability



1. higher entry score, higher exam score
2. higher peers ability, higher exam score
3. longer distance to θ^*, lower exam score





return

Prob of being in High Measured from Observable

$$Pr(High_{i,p} = 1) = F(\beta_0 + \beta_1 I(High_{i,p-1}) + \beta_2 s_{ij,p-1} + \beta_3 I(urban_i = 1))$$

VARIABLES	(1) 2nd tracking	(2) 3rd tracking	(3) 4th tracking
$s_{ij,p-1}$	6.482***	5.911***	2.664***
	(1.181)	(1.365)	(0.512)
$I(High_{ij,p-1} = 1)$	1.373***	1.691***	2.229***
	(0.288)	(0.380)	(0.236)
$I(urban_i = 1)$	0.519*	-0.221	0.433*
	(0.302)	(0.318)	(0.247)
Constant	-4.131***	-1.054***	-1.723***
	(0.668)	(0.299)	(0.242)
Observations	282	226	484

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

where F is the cdf of Logit distribution.

- Tracking exam score is becoming **less** important in later tracking decision.
- Previous class track is becoming **more** important in later tracking decision.

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

[ ]Apperson, Jarod, Carycruz Bueno, and Tim R Sass. 2016. **?**Do the cheated ever prosper? The long-run effects of test-score manipulation by teachers on student outcomes.**?** *National Center for Analysis of Longitudinal Data in Education Research (CALDER) Working Paper* 155.

[ ]Betts, Julian R. 2011. **?**The economics of tracking in education.**?** In *Handbook of the Economics of Education*, vol. 3. Elsevier, 341–381.

[ ]Cunha, Flavio and James Heckman. 2007. **?**The technology of skill formation.**?** *American economic review* 97 (2):31–47.

[ ]Cunha, Flavio and James J Heckman. 2008. **?**Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation.**?** *Journal of human resources* 43 (4):738–782.

[ ]Cunha, Flavio, James J Heckman, and Susanne M Schennach. 2010. **?**Estimating the technology of cognitive and noncognitive skill formation.**?** *Econometrica* 78 (3):883–931.

[ ]Diamond, Rebecca and Petra Persson. 2016. ?The Long-term Consequences of Teacher Discretion in Grading of High-stakes Tests.? URL http://www.nber.org/papers/w22207.pdf.

[ ]Dieterle, Steven, Cassandra M Guarino, Mark D Reckase, and Jeffrey M Wooldridge. 2015. ?How do principals assign students to teachers? Finding evidence in administrative data and the implications for value added.? *Journal of Policy Analysis and Management* 34 (1):32–58.

[ ]Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2011. ?Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya.? 101 (5):1739–1774. URL https://pubs.aeaweb.org/doi/10.1257/aer.101.5.1739.

[ ]Dustmann, Christian, Patrick A Puhani, and Uta Schönberg. 2017. ?The long-term effects of early track choice.? *The Economic Journal* 127 (603):1348–1380.

[ ]Epple, Dennis and Richard E Romano. 2011. ?Peer effects in education: A survey of the theory and evidence.? In *Handbook of social economics*, vol. 1. Elsevier, 1053–1163.

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

[ ]Fu, Chao and Nirav Mehta. 2018. ?Ability tracking, school and parental effort, and student achievement: A structural model and estimation.? *Journal of Labor Economics* 36 (4):923–979.

[ ]Glewwe, P. and K. Muralidharan. 2016. ?Chapter 10 - Improving Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications.? In *Handbook of the Economics of Education*, vol. 5, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann. Elsevier, 653–743. URL https://www.sciencedirect.com/science/article/pii/B9780444634597000105.

[ ]Heckman, James J and Stefano Mosso. 2014. ?The economics of human development and social mobility.? *Annu. Rev. Econ.* 6 (1):689–733.

[]Lavy, Victor and Edith Sand. 2018. ?On the origins of gender gaps in human capital: Short-and long-term consequences of teachers' biases.? *Journal of Public Economics* 167:263–279.

[ ]Meghir, Costas and Mårten Palme. 2005. ?Educational reform, ability, and family background.? *American Economic Review* 95 (1):414–424.

[ ]Piopiunik, Marc. 2014. ?The effects of early tracking on student performance: Evidence from a school reform in Bavaria.? *Economics of Education Review* 42:12–33.

[ ]Sacerdote, Bruce. 2011. ?Peer effects in education: How might they work, how big are they and how much do we know thus far?? In *Handbook of the Economics of Education*, vol. 3. Elsevier, 249–277.