Impact of Ability-tracking on Student's Academic and Non-Academic Outcomes: Empirical Evidence from Junior High Schools in China

by Shriyam Gupta, Chengfang Liu, Shaoping Li, Fang Chang, Yaojiang Shi

Title Page

Title: Impact of Ability-tracking on Student's Academic and Non-Academic Outcomes: Empirical Evidence from Junior High Schools in China

Abstract: Ability-tracking, a practice of grouping students into different classes based on their test scores, is prevalent in schools around the world. However, there is a lack of consensus on the impact of ability-tracking on students' learning outcomes. Drawing on a panel dataset of 9,170 students across 119 rural junior high schools from 23 counties in two provinces of China, this paper examines how ability-tracking affects student's math score, math academic self-concept and math anxiety. A school was classified as practicing ability-tracking if there was significant difference in test scores between any of its classes when they entered junior high school. Results show that ability-tracking has no impact on an average student's math score, self-concept, or anxiety score. Neither does it have any impact on math score or math self-concept of high-ability and low-ability class students. However, results from heterogeneity analysis show that ability tracking helps reduce the math anxiety of high-ability class students by 0.103 SD ($p<0.05$). Furthermore, the study also finds that ability-tracking reduces the math score of low-ability boarding students by 0.165SD ($p<0.05$), as compared to low-ability non-boarding students.

Keywords: Ability-tracking; Academic outcomes; Non-academic outcomes; Rural China; tracking; peer effects

Authors: Shriyam Gupta*[a], Chengfang Liu[b], Shaoping Li[b], Fang Chang[c], Yaojiang Shi[c]

*Presenting author

[a]Busara Center for Behavioral Economics, Innov8 Building, Saket, New Delhi, India - 110017

[b]China Center for Agricultural Policy, Peking University, Haidian District, Beijing - 100871

[c] Center for Experimental Economics in Education, Shaanxi Normal University, Xi'an, Shaanxi Province, China - 710119

Impact of Ability-tracking on Student's Academic and Non-Academic Outcomes: Empirical Evidence from Junior High Schools in China

Introduction

Ability-tracking/grouping, or sorting, refers to the process of purposely sorting students into different groups based on their 'ability' (Belfi et al., 2012; Betts, 2011). The 'ability' here usually refers to academic achievement, though students have been assigned to different groups based on combinations of their achievement, IQ, and even teacher judgements (Hattie, 2002).[1] Ability tracking is commonly practiced across the world across all education levels including primary, secondary and even college (Cheung and Rudowicz, 2003; Hanushek and WeoBmann, 2006; Betts, 2011; Duflo et al. 2011; Carman and Zhang, 2012; Steenbergen-Hu et al. 2016; Li et al., 2018). While there are different kinds of ability tracking (Steenbergen-Hu et al., 2016), this paper focuses on between-class ability ability-tracking, i.e. students are separated into different classes based on their starting academic performance.

Ability-tracking has recently become one of the most controversial subjects in educational policy literature. Proponents of ability tracking usually list the following three reasons. In the first place, they argue that it is easier for instructors to adjust their curriculum by teaching a homogeneous group than a heterogeneous one (Lou et al. 2000). Secondly, they assert that ability tracking has economic benefits, as the school can direct, invest and match its resources to a given type of students, and teach them what they need to know (Oakes & Guiton, 1995; Betts, 2011). Finally, they think ability tracking allows students to make progress in proportion to their ability, and thus maintain interest and motivation. In other words, students are less likely to be overshadowed by high ability students and suffer negative self-concept, which is called the "big-fish-little-pond-effect" (Loyalka et al. 2018), or bogged down by slower low-ability counterparts, thus creating an ideal learning environment within the classroom (Hattie, 2002).

[1] Scholars have noted that there is a difference between 'tracking' and 'ability grouping', and the former is considered less flexible than the latter (Tieso, 2003). Irrespective, there is a consensus that the practice involves assigning students to different groups based on prior academic achievement (Loveless, 2013).

Opponents of ability-tracking also have two arguments. Firstly, they argue that the use of standardized scores is unreliable to make student assignments (Betts, 2011), and thus, they magnify the already existing differences between low and high ability students to reproduce class differences and suppress mobility and opportunity in the class structure (Boaler, 1997; Kerckhoff & Glennie, 1999). Secondly, they point out that the so-called "adjusted" curriculum for low-ability students may be less stimulating, directed at behavior management than learning, be taught slowly and cause lesser content coverage and be less analytical in instructional discourse (Hong et al. 2012). For these reasons, they are concerned that ability-tracking is likely to demoralize low-ability students and make them prone to "to delinquency, absenteeism, dropout, and other social problems'' (Slavin, 1990, pp. 473).

Literature review

Empirical work on the impact of ability-tracking on students' academic performance is mixed. A meta-analytic review by Kulik and Kulik (1992) found that students in both high ability and low ability groups made academic gains when grouped homogeneously. More recently, results from a random controlled trial found that primary school students in tracking schools in Kenya had higher scores than those in non-tracking schools, and the effects persisted even a year after the program (Duflo et al., 2011). Similar positive results for ability-tracking were reported among pupils (Matthews et al., 2013) and undergraduate students (Booij et al., 2016). On the other hand, some studies have shown negative impacts of ability-tracking on both high-ability (Slavin, 1993) as well as low-ability students (Gamoran, 1992). Still, other studies found no impact of ability-tracking on math score (Betts & Shkolnik, 2000; Figlio and Page, 2002). Furthermore, international cross-country comparison on ability-tracking's impact on math score has revealed that that ability-tracking had a significant negative effect in countries with low competitiveness, a positive effect in competitive countries, and no effect in a medium competitive country (Thiemann, 2017). Lastly, Steenbergen-Hu et al. (2016) in their meta-analyses found that between class ability-grouping had no statistically significant impact on overall student's academic achievement, though within-class ability-tracking had a significantly positive impact on students' academic achievements.

In addition to its impact on academic outcomes, the impact of ability tracking on non-academic outcomes have gained the attention of scholars in recent years. This increased attention has been driven by the observation that students with lower levels of academic self-concept are more at risk of depressive reactions, lower grades, school dropout, and delinquency (Van Houtte, 2005). Research has shown that non-academic outcomes are directly correlated with student learning outcomes, for example - academic self-concept is positively correlated with academic achievement (Marsh et al., 2008). Thus, it becomes especially essential to study the effect of ability-tracking on non-academic outcomes separately.

Some studies have evaluated the impact of ability-tracking on non-academic outcomes such as academic self-concept, school well-being, leisure hours, future aspirations, and even anxiety, with mixed results. Liu et al. (2005) show that in the long-term, while self-concept declined for both higher and lower-ability stream students, lower-ability students reported higher self-concept than their higher-ability counterparts. Evidence from South Korean middle-schools has shown that ability sorting decreased the likelihood of students feeling anxious or worried about their grades (Wang, 2015). Others too have documented non-academic benefits of ability- tracking, especially for high-ability students, who gain confidence to pursue highly-competitive careers in STEM subjects (Faitar and Faitar, 2012). In the long term, Mulkey et al. (2005) found that high ability students who were tracked in middle school are likely to suffer a decline in self-concept which subsequently negatively impacts their achievements.

Ability tracking in China

China adopted the practice of ability-tracking in its early years, though introduced changes in later decades. At first, the key school policy was a "fast lane to cultivate talented students who had limited resources" (Yu et al., 2014, pp.81). However, concerns around inequity, combined with demand for quality education (Wang, 2009) and unease around rising concentration of high socio-economic profile students in key schools (Yang, 2005) facilitated many new changes. This included equalization of public expenditures & teacher salaries within municipalities, teachers from key schools being encouraged (or required) to teach, for a period, at low-performing schools, and most importantly, getting key schools to admit low-performing students (Yu et al., 2014).

Whilst China formally prohibited ability-tracking in junior-high schools in 2006, the practice still continued. Junior high-schools in China typically follow an S-shape allotment[2], however some schools continued to track students. They assigned students with the highest entrance test scores or special talents or awards to a few specially selected classes before assigning the rest of the students into different classes using the S-shape division rule. Some junior high schools have also been reported to divide students into two groups based on primary school test scores, and higher-performing students being assisted to select classes, while the lower-performing students into another set (Lai, 2007).

Why is ability-tracking still practiced in some schools despite the prohibition? According to Li. et al. (2018), the incentive system for junior-high-school principals and teachers, especially in rural schools, promote the use of ability-tracking. Economic (salary incentives) and social (reputation) benefits for teachers and principals are measured on high school admission rates (Wang et al., 2011) or ability to gain admission into prestigious high schools and universities (Tsang, 2000). Securing such admissions, especially in poor or counties is an exceedingly difficult task (Loyalka et al., 2017). Given the slim chances of success, teachers and principals disproportionately invest their time and effort in favor of the best students through ability tracking (Li et al., 2018).

Empirical work has assessed the impact of ability-tracking on both academic and non-academic outcomes in China with mixed results. Yu et al., (2014), in their study of ability-tracking in high-schools, find that high-ability classes ("key classes") do not benefit students in first-tier schools (as compared to non-key classes in the same school). However, in second-tier schools, high-ability classes benefit significantly due to ability-tracking and the result is consistent across Math, English and Chinese scores (as compared to non-key classes in the same school). With regards to non-academic outcomes, Li et. al (2018) show that fast-tracked students have higher confidence in all public institutions (schooling media, financial institutions, and government) than slow-tracked

[2] Carman and Zhang (2012) describe the process as in four distinct steps: first, "starting with the top three students, the 1st student is assigned to class one, the 2nd to class two, and the 3rd to class three. With the next three students, the order of class assignment is reversed: the 4th student is assigned to class three, the 5th to class two, and the 6th to class one. [Second], with the next three students, the order of class assignment starts with the second class and proceeds sequentially: the 7th student is assigned to class two, the 8th to class three, and the 9th to class one. With the next three students, the order is reversed: the 10th student is assigned to class one, the 11th to class three, and the 12th to class two. [Third], with the next three students, the order of class assignment starts with the third class and proceeds sequentially: the 13th student is assigned to class three, the 14th to class one, and the 15th to class two. With the next three students, the order is reversed: the 16th student is assigned to class two, the 17th to class one, and the 18th to class three. [Lastly], steps 1 through three repeats until all students are assigned."

students. Cheung and Rudowicz (2003) analysis in Hong Kong reveals that ability-tracked classes had no impact on student's self-esteem, test anxiety or academic self-concept. Though they do find that the students in higher banding schools had significantly higher self-esteem and test anxiety.

Potential gaps in the literature

A close examination of the literature reveals at least three potential gaps. In the first place, most studies on ability-tracking have come mainly from the developed world, especially the United States, reflecting data limitations and methodological challenges in other places (Duflo et al., 2011; Betts, 2011; Sacerdote, 2011; Angrist, 2014). Such findings from other developed countries may not be applicable in the context of China for multiple reasons. On the one hand, there are significant institutional differences in the way ability-tracking is organized in Chinese schools. In China, between-class ability-tracking is not based on a single subject (like it is in the US) but is done on administrative class units where students are grouped and 'key' classes (high-ability classrooms) are provided with the most privileged educational resources, especially the best teachers, for all subjects. This form of ability-tracking is usually formal and is based on test scores (Yu, et al. 2014). It is unclear if different education institutions may yield different impacts of ability-tracking. On the other hand, the rigidity of China's fast-tracked educational system along with the high social value placed on academic achievement may be particularly likely to create different environments for high- ability and low-ability tracked students (Li et al., 2018).

Secondly, most of the limited literature from China are focused on peer effects (Ding and Lehrer, 2007; Carman and Zhang, 2012; Feng and Li, 2016; Lai, 2007), not explicitly on the impact of ability-tracking. Even when some studies have attempted to assess ability-tracking, the comparison has been made between high-ability students and low-ability peers, with the latter being used as the comparison group (Li et al. 2018). This means the findings, therefore, reflect the difference between high-ability and low-ability groups, rather than ability-tracking. Thus, the question of the impact of ability-tracking remains fairly unexplored.

Finally, literature on ability-tracking from China comes from urban centres or municipalities (Yu et al., 2014), rural areas continue to be overlooked. As 70% of school-aged children in China grow up in rural areas, and rising concerns about the inequity between urban-rural education in the country (Loyalka et al., 2017, Zhang et al., 2015), it becomes critical that greater attention is paid to studying educational outcomes in rural China.

Goal and objectives of the current study

The overall goal of the paper is to assess the impact of ability-tracking on academic and non-academic outcomes in rural schools in China. Specifically, it has three key objectives: first, it examines the impact of ability-tracking on average student's Math score, math self-concept, and math anxiety. Second, it examines the impact of ability-tracking on high-ability and low-ability class student's math score, math self-concept, and math anxiety. Third, it examines the heterogeneous impact of ability-tracking by student's ability (top, middle, and bottom one-third in the class), gender, boarding status, and socioeconomic status. In doing so, the paper contributes to a growing body of literature on ability-tracking in developing countries (Glewwe, 1997; McEwan, 2003; Wang, 2015).

To achieve the goals and stated objectives, we draw on a panel dataset from rural areas in Shaanxi and Gansu province in North-west China. The data were collected by the authors themselves in three rounds and covered students, teachers, and principals. The baseline survey was conducted in the fall of 2015 followed by a midline in 2016 January, and an endline in spring 2016. The final sample consists of over 13,000 seventh-grade students from 200 junior-high schools as well as their teacher and school characteristics along with their math test scores, self-concept and anxiety measure.

The study employs a quasi-experimental design to create treatment and comparison groups. To detect ability-tracking, we test for differences in scores between classrooms across each school. If the difference in scores between classrooms is significant, the school is categorized into the treatment group of ability tracking. If not, the schools are categorized into the comparison group.

Preview of results

The results show little impact of ability-tracking, with some benefits to high-ability class students. Specifically, we find that ability-tracking has no impact on the academic outcome, i.e. math scores of students. Furthermore, there is no impact of ability-tracking on a student's academic self-concept. However, the analysis does find that ability-tracking reduced the math-anxiety levels of high-ability class students by 0.103 SD ($p<0.05$) relative to that of students from non-ability tracking schools. Last but not least, we did find some heterogeneous effects of ability-tracking. Boarding students from low-ability classes suffer a significant decline in math score due to ability-tracking. Their standardized math scores were 0.165 SD ($p<0.05$) lower as compared to their non-boarding counterparts.

The rest of the paper is organised as follows. Section 2 describes the data. Section 3 describes the research design and identification strategy. Section 4 describes the outcome variables. Section 5 describes the model specifications. Section 6 describes the results, and Section 7 presents the discussion, limitations, and conclusion.

Data

For this study, we use a panel dataset collected by the authors themselves in rural areas of two provinces of north-west China - Shaanxi, and Gansu. It covers more than 13,000 seventh-grade students from 200 junior high schools. All the 7th-grade classes of those schools become our sample classes. Amongst the 200 schools, 81 schools (40.5%) had only one 7th-grade class, whereas the remaining schools had two or more 7th-grade classes. We conducted three rounds of surveys. The baseline survey was conducted in the fall of 2015 followed by a midline in 2016 January, and an endline in spring 2016. For the purpose of the analysis, we rely on the baseline and the endline surveys. The detailed description on the sampling strategy and data collection is available in Lu et al. (2017).

We focus on three outcome variables i.e. math score, math self-concept, and math anxiety. The math score was computed from a 30-minute standardized mathematics test based on the Chinese National Curriculum Framework (Ministry of Education, 2011). The tests were strictly proctored and graded by the survey team to ensure minimal cheating. Students responded to PISA 2012 instruments to indicate their math self-concept and math anxiety score. Similar to the PISA 2012 survey, the items were given with a 4-point Likert-type response, and the measurements in the surveys were consistent with those used in PISA 2012 survey (OECD, 2012). Details of the measures used are available in Appendix Table 1. Higher values on the math self-concept index suggest that a student reported higher math self-concept i.e. he/she has higher levels of confidence in his/her math ability. On the other hand, a positive math anxiety score indicates a higher level of math anxiety i.e. a high math anxiety score means that a student suffers from higher levels of stress when doing math problems. For analysis, all three indicators were independently standardized into z-scores relative to the comparison group. This was done by subtracting the mean score and dividing by the standard deviation (SD) of the comparison group at the relevant point of time.

Research design

While ability-tracking was made illegal in China's secondary education system in 2006, its practice is still pervasive (Ministry of Education, 2006). While schools are likely to be tracking students into higher ability and lower ability classes, they may hesitate to admit doing so publicly, which makes it unviable for researchers to establish whether ability-tracking is being carried out within schools (Li et al., 2018). However, data can be exploited to establish whether schools have been tracking their students into ability groups. Previous studies in China have attempted to identify ability-tracking, including high and low track placement, to study the effects on student's academic and social outcomes by data exploitation (Li et al., 2018).

We rely on the available data to detect if ability-tracking is practised in the sample schools. International evidence suggests that ability-tracking is carried out based on pre-existing test scores of students (Oakes, 1985; Slavin, 1993). In China, when students enter junior-high school (grade 7), their new grade allocation is made based on prior test scores (Xinhua, 2010). If there is a significant disparity between average scores amongst classes, it could be attributed to the practice of ability-tracking in schools. Based on this assumption, we use raw baseline test score to test if ability-tracking is practised in a school. The baseline mathematics tests were administered to students immediately after they were assigned to different classrooms in seventh-grade, and thus capture the ability of students based on which they were placed in their respective classes. If the students were tracked while making the seventh-grade class allocation, average class math test scores are likely to capture the effect. Therefore, schools practising ability-tracking would reflect a significant difference in scores between its classes in the baseline survey.

The process of assigning the treatment and comparison groups went through a three-step process (See Figure 1 for summary description of the process). To begin with, we dropped all the students with missing baseline data (n =404) and students from schools where only one class was surveyed (n = 2964). This left us with 9740 students with baseline information and from 119 schools with multiple classes, which constitutes the sample we used to identify treatment and comparison groups for the rest of the study. As shown in Panel A of Table 1, the numbers of schools with two, three or four sample classes are 96 (80.67%), 20 (16.81%) and 2 (2.5%), respectively.

[INSERT FIGURE 1 HERE]

[INSERT TABLE 1 HERE]

In the second step, we used linear regression to test for difference in baseline scores between classes within schools to detect ability-tracking. For each school, we ran the following model:

$$Y_i = \beta_o (class)_i + \in_i \tag{1}$$

where Y_i represents the raw math score at baseline survey for any student i. β_o represents the coefficients for a vector of class dummies. Finally, $\in$ is the error term. Additionally, for schools with three or four classes, we ran Model (1) by changing the base class to test for pair-wise difference in math scores across all possible combinations of the classes. This analysis produced mean differences in baseline math scores amongst classes in each school. If the coefficient for any class was significant at 10 percent level, the school was classified as treatment school i.e. those practising ability-tracking. Otherwise a school was classified as non-tracking and assigned to the comparison group. Regression results show that 25 schools (21%) practiced tracking, which are categorized as the treatment group. The rest 94 schools (79%) constitute the comparison group.

In the second step, within the treatment group (i.e. students in ability-tracking schools), we further grouped the classes with the highest average score as the high-ability class, and remaining classes as low-ability. Panel A of Appendix Table 2 details sample distribution of treatment groups (schools with ability tracking) across high-ability and low-ability classes.

In the last step, after determining treatment and comparison groups (including high-ability and low-ability classes), we dropped all students for whom there were missing values for outcome variables (math score, self-concept and anxiety score) (n=570). Appendix Table 3 details the number of students with missing outcome variables at baseline and endline. Appendix Table 4 presents the balance test between the final study sample (n=9170) and the dropped students (n=570) across all the key variables. There were two reasons to drop the students with missing variable outcomes after using them to compute the treatment and comparison groups. Firstly, having a greater sample of students can more precisely detect differences in test scores between classes and thus, provides more accurate evidence for whether or not ability-tracking was practiced in a school. Secondly, whether a school practiced tracking only required baseline score, and not the endline scores. Further, detecting if a school practiced tracking is unrelated to the overall effect of ability-grouping on student learning outcome. In other words, the former refers to assignment of treatment while the latter is concerned with its impact.

After removing observations with missing outcome variables, we were left with 9170 students. Table 1, Panel B, details distribution of sample across treatment and comparison groups, amongst schools with two, three and four classrooms, after dropping the students with missing variables. Table 2, Panel B, details the distribution of treatment students (schools with ability tracking) across high-ability and low-ability groups.

A sample of 9170 students was used for all further analyses. At this stage, all three dependent variables (math scores, self-concept and anxiety scores) were independently standardised to z-scores relative to the comparison group. Figure 2, Panel A, plots the distribution of standardized math scores across the sample for treatment and comparison groups, while Figure 2, Panel B, plots the difference between the highest-scoring and lowest scoring classes, within each school, across treatment and comparison groups. Appendix Figure 1 reports the grade variance within each classroom for treatment and comparison schools.

[INSERT FIGURE 2 HERE]

Empirical strategy

We conducted multivariate regression analyses to estimate the impact of ability-tracking on students' academic and non-academic outcomes. Following the literature (Liu et al., 2009), we specify empirical models as follows:

$$\begin{aligned}\triangle \boldsymbol{Y}_{\boldsymbol{i}} = a + \beta_0(tracking) + \ \beta_1(baseline\ score)_i \ + \ \beta_2(student\ characteristics)_i \\ + \ \beta_3\ (household\ characteristics)_i + \ \beta_4\ (teacher\ characteristics)_i \\ + \ \beta_5(school\ characteristics)_i + \ \beta_6\ (teacher\ incentive\ treatment)_i + \epsilon_i\end{aligned} \tag{2}$$

where, $\triangle \boldsymbol{Y}$ represents the change in standardized score (from baseline to endline) for a given education-related outcome (i.e. math score, self-concept score, and anxiety score) for student *i* (Lei et al., 2018; Liu et al. 2010). *tracking* is a dummy variable indicating a student is from a school that practices ability-tracking. *baseline score* represents the standardized baseline score for a given education-related outcome i.e. math score, math self-concept score, and math anxiety score.

We also control for student, household, teacher and school characteristics. Specifically, following the literature, we control for students' gender (1 = female) (Duflo et al., 2011), their boarding status (1= boarding in school) (Lei et al., 2018), and students' age (in years) (Booij et al., 2016). We include five variables for household characteristics: whether mother graduated junior high school (1=yes) (Betts & Shkolnik, 2000), mother absent from home for the school year (1 =yes) (Li et al., 2018), father-absent from home for the school year (1=yes) (Lei et al., 2018), household socioeconomic status proxied by possession of durable assets (Loyalka et al., 2019), and whether students discuss homework at home (1 = more than once a week). We include three variables for teacher characteristics: teachers' gender (1= female) (Liu et al., 2010), years of teaching experience (Ding and Lehrer, 2007), and teachers' qualification (teacher has middle school math teaching certificate, 1=yes) (Lei et al., 2018). For school characteristics, we include two variables: school assets (scored out of 5) (Loyalka et al., 2019), and student-teacher ratio (Wang, 2015).

Lastly, as the data are sourced from a random control trial on teacher incentive, we add a teacher incentive treatment as a dummy variable which indicates if the student was from a school that was part of the treatment for the teacher incentive study. And, $\in$ is the regression error term. We cluster the standard errors at the school level.

To further estimate the effect of ability-tracking on students in high-ability and low-ability classes (in schools that practice ability- tracking), we modify the *tracking* variable in Model (2) to get an empirical specification as follows:

$$\begin{aligned}\triangle \boldsymbol{Y}_{\boldsymbol{i}} = a + \beta_0(ability - tracking) + \ &\beta_1(baseline\ score)_i \\ &+ \beta_2(student\ characteristics)_i + \beta_3\ (household\ characteristics)_i \\ &+ \beta_4\ (teacher\ characteristics)_i + \beta_5(school\ characteristics)_i \\ &+ \beta_6\ (teacher\ incentive\ treatment)_i + \in_i\end{aligned} \tag{3}$$

where, *ability- tracking* is a vector of dummy variables that indicates if a student is in a high-ability class in tracking schools, low-ability class in tracking schools, or comparison groups. Other variables in the model remain the same as in Model (2).

To estimate the heterogeneous effect of ability-tracking on student-ability, gender, boarding status and socioeconomic-status, we added their interaction terms with tracking and ability tracking variables in Models (2) and (3), respectively. The interaction variables include student-ability, a vector of dummy variables that expresses a student's standing within his/her class (Top 1/3rd, Middle 1/3rd and Bottom 1/3rd in class). Gender is a dummy variable which indicates if a student is female. Boarding status is a dummy variable which indicates if the student is boarding in the school. Lastly, socioeconomic status, a dummy variable which indicates if the student is in the top-half socioeconomic level of the sample.

Results

Descriptive Analysis

Table 1 reports the distribution of students by treatment status. Our data show that 21% of the schools (n=119), and 24% of students (n=9170) are in ability-tracking schools. Further, within the treatment group, 41% of the students were in high-ability tracked classes whereas the rest 59% in low-ability tracked classes (n=2164) (Table 2, Panel B).

Table 3 reports student, household, teacher and school characteristics across the entire sample. Average standardized baseline scores for math, math self-concept and math anxiety are 0.057, 0.01, and -0.017, respectively. Almost half of the sample students are female. About 60% of the students are boarding in the school. The average age of a student is about 13 years. Across the sample, only 25% of student's mothers have graduated from junior high school. Mothers were not present in the household for the entire school years for 15% of the sample. However, 41% of fathers in the sample were absent from the household for the entire school year. The average household asset ownership score is 2.748 out of 7. 63% of the sample students discussed homework at home at least once a week or more. About 36% of the teachers in the sample are female. Average years of teaching experience amongst teachers in the study sample is close to 10 years. About three-quarters of the teachers (72.3%) have a junior high school math teaching certificate. Lastly, sample schools have an average asset ownership score of 4.685 out of 5 and an average student-teacher ratio of 8.753.

[INSERT TABLE 3 HERE]

Table 2 reports summary statistics across key variables in treatment and comparison groups. T-tests are used to compare differences between the means of the two groups. Results from balance test show that there is a significant difference between treatment and comparison groups for baseline math, self-concept and anxiety scores. With regards to student's characteristics, gender composition is equally distributed across treatment and comparison groups, however, there is a significant difference in student's boarding status and age. Besides two variables, all other household characteristics are significantly different for treatment and comparison groups. Similarly, all teacher and school characteristics variables are significantly different across treatment and comparison groups. These observed differences imply the necessity to control for these baseline characteristics in our multivariate analyses.

[INSERT TABLE 2 HERE]

Multivariate Analysis

1. *Impact of Ability-tracking on Average Student:*

Before we report results from multivariate analyses, it is necessary to note that we define average students as all students who were in the schools that practice ability-tracking. We also define high-ability and low-ability class students as only students in high and low ability classes, respectively, in schools that practice ability-grouping. Therefore, average student is composed of both high- and low-ability class students. Therefore, gains for 'average students' refers to the overall gains in schools that practice ability-tracking relative to schools that do not practice it.

Regression results show that controlling for baseline score, students, household, teacher, and school characteristics, ability-tracking has no significant impact on student's math score (Table 3,Column 6). Similarly, no significant impact is recorded for either math self-concept[3] (Table 4, Column 6) or math anxiety[4] (Table 5, Column 6).

[INSERT TABLE 3 HERE]
[INSERT TABLE 4 HERE]
[INSERT TABLE 5 HERE]

2. *Impact of Ability-tracking on High-ability and Low-ability Class Students:*

Regression results from Model (3) show that there is no significant impact of ability-tracking on math scores of high or low ability class students. Similarly, ability-tracking does not have a significant impact on math self-concept of high and low-ability classroom students (Table 6, Panels A and B, Column 6).

Unlike math score and self-concept, results show that ability-tracking has a significant impact on reducing the anxiety of high-ability class students. Controlling for baseline anxiety score, student, household, teacher and school characteristics, high-ability classrooms students experience a

[3] We use self-concept self/academic self-concept in the chapter to refer to math self-concept.
[4] We use anxiety in this chapter to refer to math anxiety.

significant decline in math anxiety by 0.103 SD (SE=0.051, $p<0.05$) due to ability-tracking. However, no such impact is recorded on math anxiety of low-ability class students (Table 6, Panel C, Column 6).

[INSERT TABLE 6 HERE]

3. *Heterogeneous Impact of Ability-tracking on Average Student:*

We assessed the heterogenous impact of ability-tracking on average student's academic and non-academic outcomes across four key characteristics. We found no heterogenous impact on test scores across ability, gender, boarding status or socio-economic status. Similar results were observed for both math self-concept and anxiety (Table 7).

[INSERT TABLE 7 HERE]

4. *Heterogeneous Impact of Ability-Tracking on High and Low-Ability Class Students*

We also assessed the heterogenous impact of ability-tracking on high and low-ability and class student's academic and non-academic outcomes across the same four key characteristics. With regards to academic outcomes, in high-ability and low-ability classes, math scores of middle 1/3rd and top 1/3rd students (within a class) are not significantly different as compared to bottom 1/3rd students' due to ability-tracking. Similar results are found for both classes across gender, and socio-economic status (Table 8, Panel A, Columns 1, 2 and 4).

In contrast, boarding students in low-ability classes suffer a significant negative impact of ability-tracking as compared to non-boarding students, and their standardized math scores declined by 0.165SD (SE=0.073, $p<0.05$). No such heterogenous impact is observed for boarding students in high-ability class students (Table 8, Panel A, Column 3).

For both high and low-ability class students, regression results show that there was no heterogenous impact of ability-tracking on academic self-concept across student-ability, gender, boarding status and socio-economic status. Similar results are found for math anxiety across all four heterogeneous characteristics (Table 8, Panels B and C).

[INSERT TABLE 8 HERE]

Discussion

Using panel data of 9170 students from 119 junior schools in rural China, the paper evaluated the impact of between-class ability-tracking on students' academic and non-academic outcomes. All schools where we found significant differences in scores between classrooms were categorized as treatment schools (i.e. schools practicing ability-tracking), whilst others were categorized as comparison groups. Results from our analyses revealed that ability-tracking has no significant impact on average student's math score, math academic self-concept or math anxiety as compared to students in non ability-tracking schools. There was no differential impact by student-ability, gender, boarding status of students, and socioeconomic-levels. Further, we did not find ability tracking had significant impact on high and low-ability students' math score or math academic self-concept as compared to the comparison group. However, our data did show that ability tracking significantly reduced math anxiety of high-ability classroom students by -0.103 SD ($p<0.05^{**}$). The study also finds that ability-tracking has significant heterogeneous impact on math scores of low-ability boarding students.

Findings of the study on academic outcomes contradict some literature on ability-tracking, while consistent with others. The results are inconsistent with findings of Duflo et al., (2011), Fuligni et al., (1995), Matthews et al., (2013) and Booij et al. (2016), all of whom have found some kind of positive impact on students' academic outcomes i.e. test scores. Similarly, the findings are also not in accordance with Slavin (1993) and Gamoran (1992) both of whom have found negative impact of ability-tracking. However, the findings are consistent with those of Betts & Shkolnik (2000) and Figlio and Page (2002) both of whom find no or minimal impact of ability-tracking on student's math test scores. Studies in Asia too have found similar results. Cheung and Rudowicz (2003) find no significant impact of ability-tracking on academic outcomes. Similarly, Yu. et al., 2014 in their study too found ability-tracking had mixed effects, and only limited positive effect.

Steenbergen-Hu et al. (2016) in their second-order meta-analysis found no statistically significant impact of ability-tracking on overall student's academic achievement.

The study also finds that ability-tracking has some heterogeneous impact on math score of low-ability boarding students, who experience a reduction in their score by 0.165 SD (SE=0.073, p<0.05**), relative to their non-boarding counterparts. Although, it is unclear why the interplay of two factors could potentially explain the results. Studies have shown that boarding students perform much worse than their non-boarding counterparts across, both, mental health and academic outcome indicators (Wang et al., 2016). This combined with the potential adverse effect of having low-ability classmates could possibly explain the negative impact of ability-tracking on math score of low-ability boarding students.

With regards to math academic self-concept, the study results are contrary to previous literature, but are consistent with findings of other studies. The study findings are inconsistent with Liu et al. (2005), Mulkey et al. (2005), and Chmielewski et al., (2013) all of whom found that ability-tracking significantly affected the academic self-concept of both high and low-ability students. However, findings are consistent with those of Ireson and Hallam (2009) who found that the policy did not have a significant impact on a student's subject-specific self-concepts in math, science, and English (though it did find a negative effect on students' general academic self-concept). Similarly, Cheung and Rudowicz (2003) study in Hong Kong found no significant impact of ability-tracking on academic self-concept.

Caution needs to be employed while concluding that ability-tracking is ineffective on China's rural-school student's math scores and math self-concept. The characteristics of the data may have made it difficult to capture the effect on test score. Firstly, the data reports findings over a period of only one year, which may not have been sufficient to show the impact of the policy. Secondly, and more importantly, there may be substantially less variation in ability amongst students in rural schools of China, as compared to students in US (Ding and Lehrer, 2007), which may have made it difficult to capture the differences in scores over a period of one year.

Despite these, the key findings of the study are that ability-tracking has a significant and large impact on reducing math anxiety amongst high-ability classroom students in comparison to comparison students (those in non ability-tracking schools). Very few studies have assessed the impact of ability-tracking on math anxiety of students. Previous studies have assessed the impact on grade anxiety and test anxiety, and report both similar and dissimilar results. Wang (2015) finds similar results to those of the paper, and shows that ability-tracking significantly reduced grade anxiety in South Korea. However, results from Hong Kong show that students in the high-band schools experience the highest test anxiety.

In China, high-ability students may experience less anxiety because of three reasons. Firstly, Students in high-ability classrooms are often allotted better resources and teachers. Studies have shown that teachers and principals disproportionately invest their time in high-ability classrooms. Evidence from around the world suggests that students in high-ability groups are often allotted higher quality teachers (Hallberg & Schaufeli, 2006). In contrast, low-ability classrooms have teachers who are less engaged in their work (Klusmann et al., 2008). Access to such resources means that students in high-ability classes learn better and thus, suffer from lesser anxiety. Secondly, not only do the high-ability students benefit with better resources, the general social well-being of low-ability class students in China lags behind those of high-ability students in that they are more likely to suffer from mental health problems and have a more likely to dropout (Mo et al., 2013; Shi et al., 2015; Wang et al., 2015; Yi et al., 2012). Moreover, high-ability classroom students in Chinese rural schools have higher interpersonal trust, greater confidence in the educational institutions, and more faith in financial and government systems that their low-ability counterparts (Li et al., 2018). These factors could also have a positive impact on their general anxiety levels. Lastly, being in high-ability classes means that students incur positive effects of having higher quality peers. Studies have shown that access to better quality peers in China significantly increases academic learning (Ding and Lehrer, 2007; Lai, 2007), and thus may reduce anxiety.

We acknowledge five limitations of the study. Firstly, the study employs a quasi-experimental design. This means that the findings cannot be considered causal, they are just evidence for correlation between ability-tracking and the student outcomes. Secondly, the study identified ability-tracking by comparing differences in scores between classes. It is possible that significant differences in classes could have accrued from factors other than ability-tracking. However, this is highly unlikely for two reasons: a.) the schools and students were sampled from a similar region and share common socio-economic backgrounds; and b.) given that ability-tracking is banned at the junior-high school level and the class composition is based on an S-shape policy, it is implausible that these differences in scores could be attributed to anything but ability-tracking. Thirdly, the treatment and comparison groups are unbalanced at the baseline in terms of some student, teacher and school characteristics. This could have biased the results. The analysis tried to minimize the potential bias by introducing these characteristics as control variables in the models. However, we recognize that it does not completely eliminate the endogeneity problem. Fourthly, while the study finds that ability-tracking reduced math-anxiety it could not test for underlying mechanisms that influence the outcome as it was beyond the scope of the study. Lastly, the study tracked academic and non-academic outcomes for only a year, which may not have been sufficient time to see the effects of ability-tracking.

Despite its limitations, the paper contributes to the growing literature around the contentious debate on ability-tracking. To address the paucity of empirical analysis on ability-tracking in developing context, the study employed a quasi-experimental design and showed that ability-tracking has a significant impact on student's math anxiety, but had no significant effect on their math scores or academic self-concept. As far as we know, this is the first English language literature that attempts to study the impact of ability-tracking on both academic and non-academic outcomes using a panel data set from rural China. Unlike other studies, it did so by comparing schools with ability-tracking to non-ability-tracking schools, and thus is able to assess the impact of the policy. In this way, the paper attempts to contribute to the growing body of literature on economics of education literature, and is also relevant to the larger body of work on the study of peer effects in education.

The study provides direction for future research on peer effects in education. The study adds to the debate on ability-tracking however, there is still no consensus whether it is effective in improving student learning. While the study did find that ability-tracking reduced anxiety of high-ability students, it is unclear if the results will persist over a longer period of time. Previous studies have shown that benefits of non-academic outcomes may disappear or switch from high-ability to low-ability students (Mulkey et al., 2005). Therefore, future studies may need to assess the impact of ability-tracking over a longer period of time. Secondly, the future work would need to be investigated why ability-tracking significantly reduced math anxiety of high-ability students. Further, it would need to investigate why ability-tracking reduced math anxiety, but did not have any significant impact on academic self-concept, and math score. While association between math academic self-concept and math scores are well documented (Marsh et al., 2008), work on cross interaction between anxiety, self-concept and test scores need to be investigated, especially in developing contexts. Lastly, ability-tracking is legally allowed in the high-schools of China and is also widely practiced. Studies would need to investigate the effects of the policy at the high-school level on academic and non-academic outcomes. Moreover, attempts must be made to disentangle whether there are long term associations between students being tracked in junior-high schools, and its effect on performance in high schools.

References

Angrist, J.D., 2014. The perils of peer effects. Labour Economics 30, 98–108. https://doi.org/10.1016/j.labeco.2014.05.008

Belfi, B., Goos, M., De Fraine, B., Van Damme, J., 2012. The effect of class composition by gender and ability on secondary school students' school well-being and academic self-concept: A literature review. Educational Research Review 7, 62–74. https://doi.org/10.1016/j.edurev.2011.09.002

Betts, J.R., 2011. The Economics of Tracking in Education. Handbook of the Economics of Education 3, 341–381. https://doi.org/10.1016/B978-0-444-53429-3.00007-7

Betts, J.R., Shkolnik, J.L., 2000. The effects of ability grouping on student achievement and resource allocation in secondary schools. Economics of Education Review 19, 1–15. https://doi.org/10.1016/s0272-7757(98)00044-2

Boaler, J., 1997. Setting, Social Class and Survival of the Quickest. British Educational Research Journal 23, 575–595. https://doi.org/10.1080/0141192970230503

Bong, M., Skaalvik, E.M., 2003. Academic Self-Concept and Self-Efficacy: How Different Are They Really? 15, 1–40. https://doi.org/10.1023/A:1021302408382

Booij, A.S., Leuven, E., Oosterbeek, H., 2016. Ability Peer Effects in University: Evidence from a Randomized Experiment. The Review of Economic Studies 84, 547–578. https://doi.org/10.1093/restud/rdw045

Carman, K.G., Zhang, L., 2012. Classroom peer effects and academic achievement: Evidence from a Chinese middle school. China Economic Review 23, 223–237. https://doi.org/10.1016/j.chieco.2011.10.004

Cheung, C.K., Rudowicz, E., 2003. Academic outcomes of ability grouping among junior high school students in hong kong. Journal of Educational Research 96, 241–254. https://doi.org/10.1080/00220670309598813

China National Bureau of Statistics, 2014. China Statistical Yearbook, 2014. . China Statistics Press, Beijing.

Chinese National Curriculum Framework, 2011. Ministry of Education, Government of China.

Chmielewski, A.K., Dumont, H., Trautwein, U., 2013. Tracking Effects Depend on Tracking Type: An International Comparison of Students' Mathematics Self-Concept. American Educational Research Journal 50, 925–957. https://doi.org/10.3102/0002831213489843

Compulsory education law of the People's Republic of China, 2006. Ministry of Education, Government of China.

Ding, W., Lehrer, S.F., 2007. Do peers affect student achievement in china's secondary schools? Review of Economics and Statistics 89, 300–312. https://doi.org/10.1162/rest.89.2.300

Duflo, E., Dupas, P., Kremer, M., 2011. Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. American Economic Review 101, 1739–1774. https://doi.org/10.1257/aer.101.5.1739

OECD. 2012. Education at a Glance, 2012. Highlights from Education at a Glance. OECD. https://doi.org/https://doi.org/10.1787/eag_highlights-2012-en

Faitar, G.M., Faitar, S.L., 2012. The Influence of Ability Tracking on the Performances of Minority Learners. Journal of Instructional Pedagogies 7, 1–9.

Feng, H., Li, J., 2016. Head teachers, peer effects, and student achievement. China Economic Review 41, 268–283. https://doi.org/10.1016/j.chieco.2016.10.009

Figlio, D.N., Page, M.E., 2002. School choice and the distributional effects of ability tracking: Does separation increase inequality? Journal of Urban Economics 51, 497–514. https://doi.org/10.1006/juec.2001.2255

Fuligni, A.J., Eccles, J.S., Barber, B.L., 1995. The Long-Term Effects of Seventh-Grade Ability Grouping in Mathematics. The Journal of Early Adolescence 15, 58–89. https://doi.org/10.1177/0272431695015001005

Gamoran, A., 1992. The Variable Effects of High School Tracking. American Sociological Review 57, 812–828. https://doi.org/10.2307/2096125

Gamoran, A., Nystrand, M., Berends, M., LePore, P.C., 1995. An Organizational Analysis of the Effects of Ability Grouping. American Educational Research Journal 32, 687–715. https://doi.org/10.3102/00028312032004687

Glewwe, P., 1997. Estimating the Impact of Peer Group Effects on Socioeconomic Outcomes: Does the Distribution of Peer Group Characteristics Matter? Economics of Education Review 16, 39–43. https://doi.org/10.1016/s0272-7757(96)00017-9

Hallberg, U.E., Schaufeli, W.B., 2006. “Same same” but different? Can work engagement be discriminated from job involvement and organizational commitment? European Psychologist 11, 119–127. https://doi.org/10.1027/1016-9040.11.2.119

Hanushek, E.A., Wößmann, L., 2006. Does Educational Tracking Affect Performance and Inequality? Differences- in-Differences Evidence Across Countries. The Economic Journal 116, C63–C76. https://doi.org/10.1111/j.1468-0297.2006.01076.x

Hattie, J.A.C., 2002. Classroom composition and peer effects. International Journal of Educational Research 37, 449–481. https://doi.org/10.1016/S0883-0355(03)00015-6

Hong, G., Corter, C., Hong, Y., Pelletier, J., 2012. Differential Effects of Literacy Instruction Time and Homogeneous Ability Grouping in Kindergarten Classrooms: Who Will Benefit? Who Will Suffer? Educational Evaluation and Policy Analysis 34, 69–88. https://doi.org/10.3102/0162373711424206

Ireson, J., Hallam, S., 2009. Academic self-concepts in adolescence: Relations with achievement and ability grouping in schools. Learning and Instruction 19, 201–213. https://doi.org/10.1016/j.learninstruc.2008.04.001

Kerckhoff, A.C., Glennie, E., 1999. The Matthew effect in American education. Research in Sociology of Education and Socialization 12, 35–66.

Klusmann, U., Kunter, M., Trautwein, U., Lüdtke, O., Baumert, J., 2008. Teachers’ Occupational Well-Being and Quality of Instruction: The Important Role of Self-Regulatory Patterns. Journal of Educational Psychology 100, 702–715. https://doi.org/10.1037/0022-0663.100.3.702

Kulik, J.A., Kulik, C.-L.C., 1992. Meta-analytic Findings on Grouping Programs. Gifted Child Quarterly 36, 73–77. https://doi.org/10.1177/001698629203600204

Lei, W., Li, M., Zhang, S., Sun, Y., Sylvia, S., Yang, E., Ma, G., Zhang, L., Mo, D., Rozelle, S., 2018. Contract teachers and student achievement in rural China : evidence from class fixed effects. Australian Journal of Agricultural and Resource Economics 299–322. https://doi.org/10.1111/1467-8489.12250

Lai, F., 2007. How do classroom peers affect student outcomes? Evidence from a natural experiment in Beijing's middle schools. Unpublished manuscript.

Li, F., Loyalka, P., Yi, H., Shi, Y., Johnson, N., Rozelle, S., 2018. Ability tracking and social trust in China's rural secondary school system. School Effectiveness and School Improvement 29, 545–572. https://doi.org/10.1080/09243453.2018.1480498

Liang, C., Zhang, H., Li, Z., 2013. Silent Revolution – A Study on the Social Resources of Students from Peking University and Suzhou University from 1949–2002. SDX Joint Publishing Company, China (in Chinese).

Liu, C., Zhang, L., Luo, R., Rozelle, S., Loyalka, P., 2010. The effect of primary school mergers on academic performance of students in rural China. International Journal of Educational Development 30, 570–585. https://doi.org/10.1016/j.ijedudev.2010.05.003

Liu, C., Zhang, L., Luo, R., Rozelle, S., Sharbono, B., Shi, Y., 2009. Development challenges, tuition barriers, and high school education in China. Asia Pacific Journal of Education 29, 503–520. https://doi.org/10.1080/02188790903312698

Liu, W.C., Wang, C.K.J., Parkins, E.J., 2005. A longitudinal study of students' academic self-concept in a streamed setting: The Singapore context. British Journal of Educational Psychology 75, 567–586. https://doi.org/10.1348/000709905X42239

Lou, Y., Abrami, P.C., Spence, J.C., 2000. Effects of within-class grouping on student achievement: An exploratory model. Journal of Educational Research 94, 101–112. https://doi.org/10.1080/00220670009598748

Loveless, T., 2013. How Well Are American Students Learning? The 2013 Brown Center Report on American Education,(3) 2. Washington, DC: Brookings Institution.

Loyalka, P., Chu, J., Wei, J., Johnson, N., Reniker, J., 2017. Inequalities in the Pathway to College in China: When Do Students from Poor Areas Fall Behind? China Quarterly 229, 172–194. https://doi.org/10.1017/S0305741016001594

Loyalka, P., Sylvia, S., Liu, C., Chu, J., Shi, Y., 2019. Pay by design: Teacher performance pay design and the distribution of student achievement. Journal of Labor Economics 37, 621–662. https://doi.org/10.1086/702625

Loyalka, P., Zakharov, A., Kuzmina, Y., 2018. Catching the big fish in the little pond effect: Evidence from 33 countries and regions. Comparative Education Review 62, 542–564. https://doi.org/10.1086/699672

Lu, M., Shi, Y., Chang, F., Kenny, K., Rozelle, S., 2017. The Gender gap in math performance, self-concept, and anxiety: rural and urban China in an international context (No. Working Paper 312).

Marsh, H.W., Seaton, M., Trautwein, U., Lüdtke, O., Hau, K.T., O'Mara, A.J., Craven, R.G., 2008. The big-fish-little-pond-effect stands up to critical scrutiny: Implications for theory, methodology, and future research. Educational Psychology Review 20, 319–350. https://doi.org/10.1007/s10648-008-9075-6

Matthews, M.S., Ritchotte, J.A., McBee, M.T., 2013. Effects of schoolwide cluster grouping and within-class ability grouping on elementary school students' academic achievement growth. High Ability Studies 24, 81–97. https://doi.org/10.1080/13598139.2013.846251

McEwan, P.J., 2003. Peer effects on student achievement: Evidence from Chile. Economics of Education Review 22, 131–141. https://doi.org/10.1016/S0272-7757(02)00005-5

Mo, D., Zhang, L., Yi, H., Luo, R., Rozelle, S., Brinton, C., 2013. School Dropouts and Conditional Cash Transfers: Evidence from a Randomised Controlled Trial in Rural China's Junior High Schools. Journal of Development Studies 49, 190–207. https://doi.org/10.1080/00220388.2012.724166

Mulkey, L.M., Catsambis, S., Steelman, L.C., Crain, R.L., 2005. The long-term effects of ability grouping in mathematics: A national investigation. Social Psychology of Education 8, 137–177. https://doi.org/10.1007/s11218-005-4014-6

Oakes, J., Guiton, G., 1995. Matchmaking: The Dynamics of High School Tracking Decisions. American Educational Research Journal 32, 3–33. https://doi.org/10.3102/00028312032001003

Oakes, J., 1985. Keeping track : how schools structure inequality. Yale University Press.

Sacerdote, B., 2011. Peer Effects in Education: How might they work, how big are they and how much do we know Thus Far?, 1st ed, Handbook of the Economics of Education. Elsevier B.V. https://doi.org/10.1016/B978-0-444-53429-3.00004-1

Shi, Y., Zhang, L., Ma, Y., Yi, H., Liu, C., Johnson, N., Chu, J., Loyalka, P., Rozelle, S., 2015. Dropping out of Rural China's Secondary Schools: A Mixed-methods Analysis. China Quarterly 224, 1048–1069. https://doi.org/10.1017/S0305741015001277

Singer, E. 1981. Reference groups and social evaluations. In M. Rosen- berg & R. H. Turner (Eds.), Social psychology: Sociological perspective (pp. 66-93). New York: Basic Books.

Slavin, R.E., 1990. Achievement Effects of Ability Grouping in Secondary Schools: A Best-Evidence Synthesis. Review of Educational Research 60, 471–499. https://doi.org/10.3102/00346543060003471

Slavin, R.E., 1993. Ability Grouping in the Middle Grades: Achievement Effects and Alternatives. The Elementary School Journal 93, 535–552. https://doi.org/10.1086/461739

Steenbergen-Hu, S., Makel, M.C., Olszewski-Kubilius, P., 2016. What One Hundred Years of Research Says About the Effects of Ability Grouping and Acceleration on K–12 Students' Academic Achievement: Findings of Two Second-Order Meta-Analyses, Review of Educational Research. https://doi.org/10.3102/0034654316675417

Thiemann, K., 2017. Does the Impact of Ability Grouping vary with the Culture of Competitiveness ? - Evidence from PISA 1–46.

Tieso, C.L., 2003. Ability grouping is not just tracking anymore. Roeper Review 26, 29–36. https://doi.org/10.1080/02783190309554236

Tsang, M.C., 2000. Education and National Development in China since 1949: Oscillating Policies and Enduring Dilemmas. China Review 579–618.

Van Houtte, M., 2005. Global self-esteem in technical/vocational versus general secondary school tracks: A matter of gender? Sex Roles 53, 753–761. https://doi.org/10.1007/s11199-005-7739-y

Wang, S., 2009. Key school policy hurts education equity. China Education Daily, March 8th, 2009. http://www.jyb.cn/xwzx/jcjy/sxkd/t20070308_68693htm (in Chinese).

Wang, H., Yang, C., He, F., Shi, Y., Qu, Q., Rozelle, S., Chu, J., 2015. Mental health and dropout behavior: A cross-sectional study of junior high students in northwest rural China. International Journal of Educational Development 41, 1–12. https://doi.org/10.1016/j.ijedudev.2014.12.005

Wang, L.C., 2015. All work and no play? The effects of ability sorting on students' non-school inputs, time use, and grade anxiety. Economics of Education Review 44, 29–41. https://doi.org/10.1016/j.econedurev.2014.10.008

Wang, A., Medina, A., Luo, R., Shi, Y., Yue, A., 2016. To Board or Not to Board: Evidence from Nutrition, Health and Education Outcomes of Students in Rural China. China and World Economy 24, 52–66. https://doi.org/10.1111/cwe.12158

Wills, T. A. 1987. Downward comparison as a coping mechanism. In Coping with negative life events (pp. 243-268). Springer, Boston, MA.

Wang, X., Chengfang, L., Linxiu, Z., Renfu, L., Thomas, G., Yaojiang, S., Scott, R., Brian, S., 2011. What is keeping the poor out of college? Enrollment rates, educational barriers and college matriculation in China. China Agricultural Economic Review 3, 131–149. https://doi.org/10.1108/17561371111131281

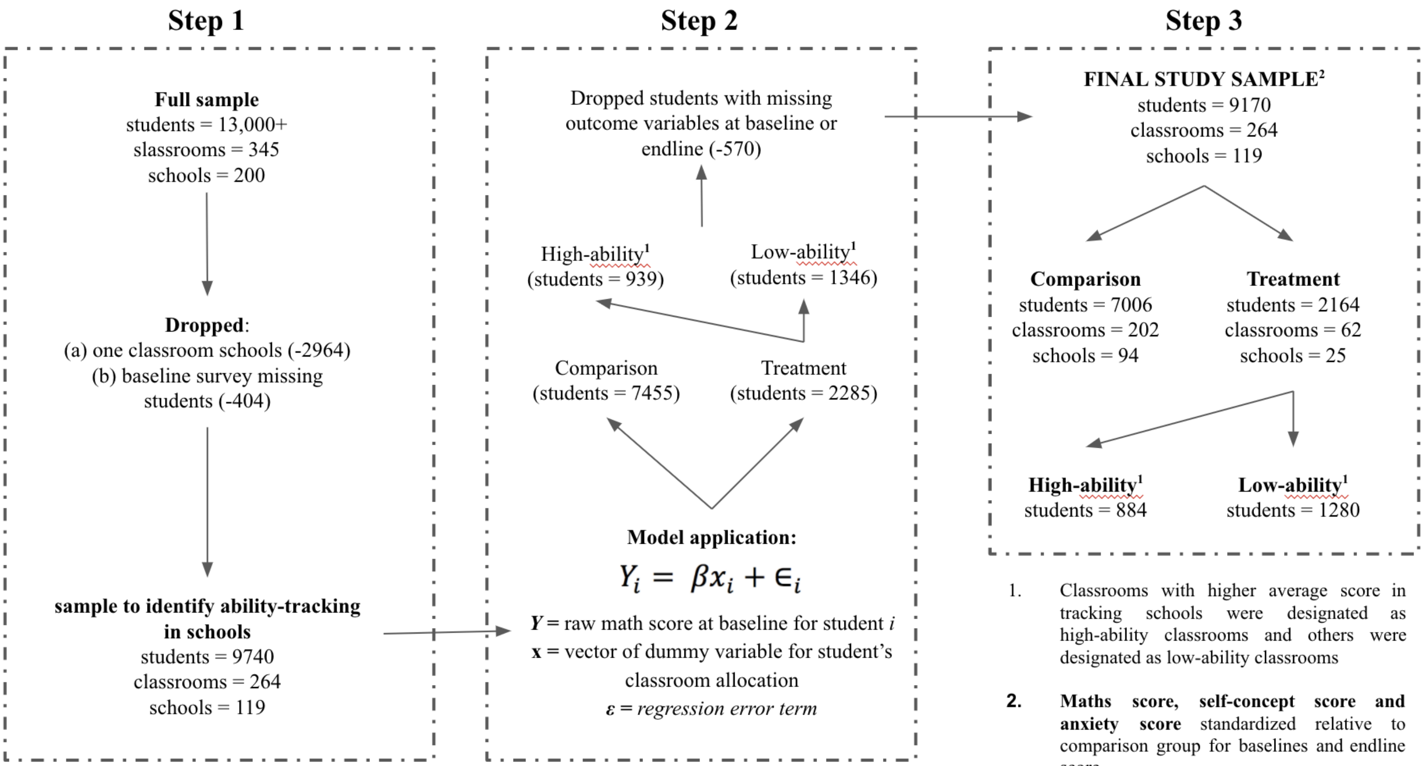
Xinhua, 2010. Setting up the "key schools", exams for tracking placement [in Chinese: Qiao Li Ming Mu She "Zhong Dian Ban", Fen Ban Kao Shi Pai Man Kai Xue Ji].

Yang, D., 2005. A Research into the Senior High School Students' Social-class- delamination and Education Acquisition. Research on Education Tsinghua University 26 (3), 52–59. (in Chinese)

Yi, H., Zhang, L., Luo, R., Shi, Y., Mo, D., Chen, X., Brinton, C., Rozelle, S., 2012. Dropping out: Why are students leaving junior high in China's poor rural areas? International Journal of Educational Development 32, 555–563. https://doi.org/10.1016/j.ijedudev.2011.09.002

Yu, Z., Dongsheng, C., Wen, W., 2014. The heterogeneous effects of ability grouping on national college entrance exam performance - evidence from a large city in China. International Journal of Educational Development 39, 80–91. https://doi.org/10.1016/j.ijedudev.2014.08.012

Figure 1: Creating treatment and control group



1. Classrooms with higher average score in tracking schools were designated as high-ability classrooms and others were designated as low-ability classrooms

2. **Maths score, self-concept score and anxiety score** standardized relative to comparison group for baselines and endline score.

Figure 2: Standardized math scores across treatment and comparison groups for students without missing values for outcome variables

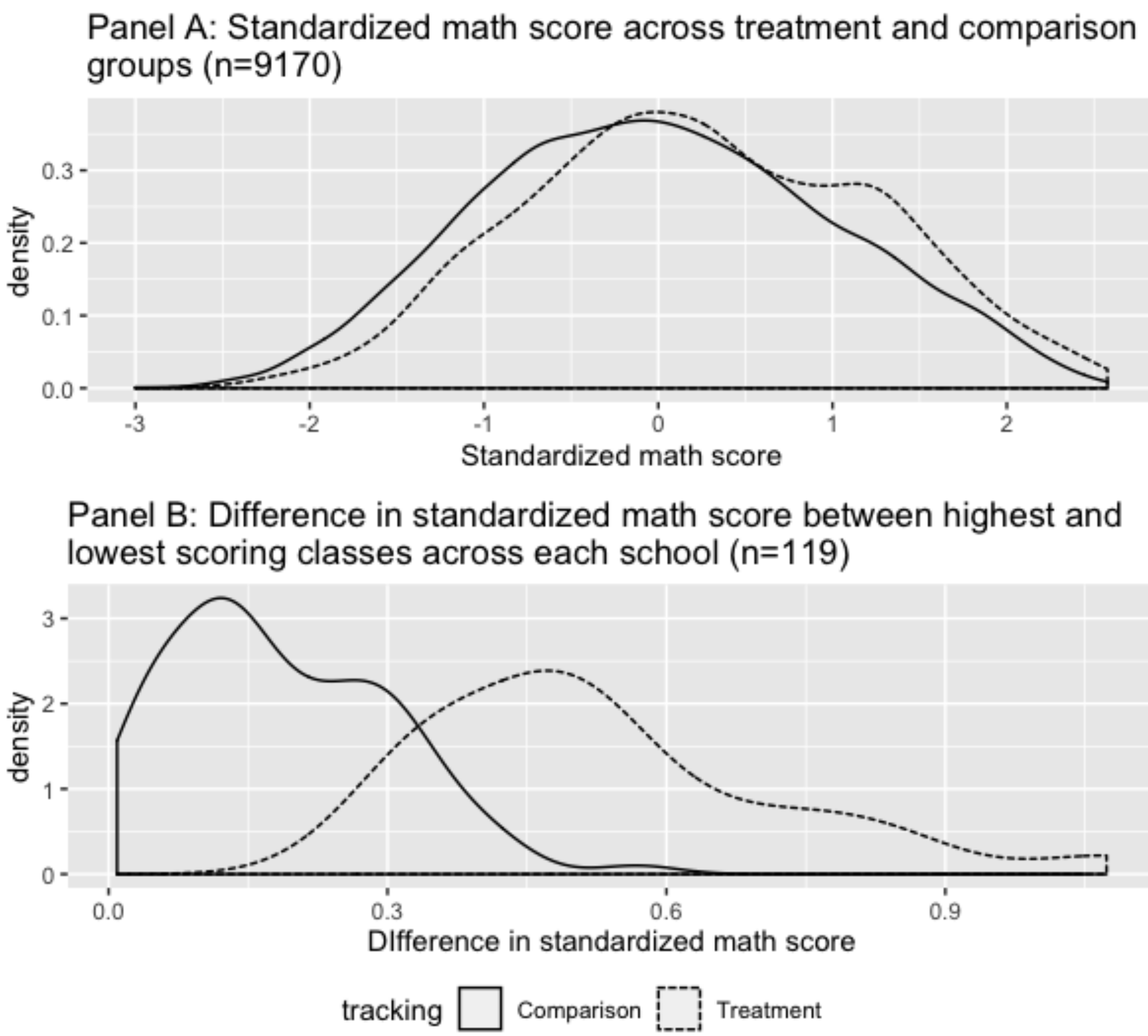


Table 1: Distribution of sample across treatment and comparison groups

No. of classrooms within each school	Total		Treatment		Comparison	
	Schools	Students	Schools	Students	Schools	Students
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Distribution of students across treatment and comparison groups (with students' that are missing outcome variable values)						
2	96	7215	15	1171	81	6044
3	20	2170	8	864	12	1306
4	3	355	2	250	1	105
Total	119	9740	25	2285	94	7455
Panel B: Final sample of students across treatment and comparison groups (without students' that are missing outcome variable values)						
2	96	6773	15	1101	81	5672
3	20	2057	8	823	12	1234
4	3	340	2	240	1	100
Total	119	9170	25	2164	94	7006
Proportion	100%	100%	21%	23.60%	79%	76.40%

Note:

1. We dealt with missing students outcome variables (those for whom either baseline or endline was missing for one of the three dependent variables (i.e. math score, self-concept score, or anxiety;) as follows. These students (n= 570) were kept in the data set while modelling for tracking (see model 1). However, these students were removed from the data set for all subsequent analysis. As discussed in the text, there were two reasons to drop these students after using them to compute the treatment and comparison group. Firstly, having a greater sample of students can more precisely detect differences in test scores between classrooms and thus, provides a more accurate evidence for whether or not tracking was prevalent in a school. Secondly, detecting incidence of tracking in a school is unrelated to the overall effect of ability-grouping on student learning outcomes.

Table 2: Balance test between treatment and comparison groups across key variables

	Overall	Treatment	Comparison	p-value
	(n=9170)	(n=2164)	(n=7006)	(T vs C)
	(1)	(2)	(3)	(4)
OUTCOME VARIABLES				
Baseline standardized math score	0.057 (1.00)	0.242 (0.990)	0.00 (1.00)	<0.01***
Baseline standardized self-concept score	0.010 (1.00)	0.043 (1.01)	0.00 (1.00)	<0.1*
Baseline standardized anxiety score	-0.017 (1.01)	-0.072 (1.03)	0.00 (1.00)	<0.01***
STUDENT CHARACTERISTICS				
Female student (1=yes)	0.499 (0.500)	0.496 (0.500)	0.500 (0.500)	0.735
Student boarding at school (1=yes)	0.597 (0.491)	0.654 (0.476)	0.579 (0.494)	<0.01***
Age (in years)	13.0 (0.940)	13.0 (0.976)	13.0 (0.929)	<0.1*
HOUSEHOLD CHARACTERISTICS				
Mother's graduated junior high school (1=yes)	0.251 (0.434)	0.276 (0.447)	0.243 (0.429)	<0.01***
Mother absent for both semesters (1=yes)	0.149 (0.356)	0.132 (0.338)	0.155 (0.362)	<0.01***
Father absent for both semesters (1=yes)	0.413 (0.492)	0.446 (0.497)	0.403 (0.491)	<0.01***
Household assets (score 0-7)	2.75 (1.70)	2.75 (1.66)	2.75 (1.71)	0.999
Students discuss homework with parents - more than once a week (1=yes)	0.633 (0.482)	0.633 (0.482)	0.633 (0.482)	0.997
MATH TEACHER CHARACTERISTICS				
Female teacher (1 =yes)	0.358 (0.479)	0.341 (0.474)	0.363 (0.481)	<0.1*
Years of teaching experiences	9.61 (7.07)	9.19 (6.65)	9.74 (7.19)	<0.01***
Teacher has middle school math teaching certificate (1=yes)	0.723 (0.447)	0.821 (0.384)	0.693 (0.461)	<0.01***
SCHOOL CHARACTERISTICS				
School assets (score 0-5)	4.69 (0.620)	4.66 (0.563)	4.69 (0.636)	<0.1*
Student/teacher ratio	8.75 (2.87)	7.83 (2.39)	9.04 (2.95)	<0.01***

Note

1. N = number of observations
2. *p<0.1; **p<0.05; ***p<0.01

Table 3: Impact of ability-tracking on average student's math score

	Value Added: Math Score					
	(1)	(2)	(3)	(4)	(5)	(6)
Tracking	-0.171***	-0.080	-0.075	-0.074	-0.070	-0.075
	(0.061)	(0.070)	(0.068)	(0.067)	(0.066)	(0.066)
Baseline math score		-0.363***	-0.395***	-0.400***	-0.401***	-0.408***
		(0.015)	(0.014)	(0.013)	(0.013)	(0.013)
Teacher incentive treatment (1=yes)	0.016	-0.005	0.009	0.009	0.007	0.010
	(0.049)	(0.056)	(0.054)	(0.053)	(0.054)	(0.052)
STUDENT CHARACTERISTICS						
Female student (1=yes)			-0.014	-0.014	-0.014	-0.013
			(0.018)	(0.018)	(0.018)	(0.018)
Student boarding in school (1=yes)			0.085***	0.087***	0.078***	0.064**
			(0.030)	(0.030)	(0.029)	(0.030)
Age of student (in years)			-0.134***	-0.132***	-0.132***	-0.133***
			(0.013)	(0.013)	(0.013)	(0.013)
HOUSEHOLD CHARACTERISTICS						
Mother graduated junior high school (1=yes)				0.004	0.016	0.009
				(0.025)	(0.025)	(0.023)
Mother absent for both semesters (1=yes)				-0.034	-0.030	-0.025
				(0.026)	(0.026)	(0.026)
Father absent for both semesters (1=yes)				-0.001	-0.002	-0.003
				(0.019)	(0.019)	(0.019)
Household assets (score 0-7)				0.020**	0.018**	0.013*
				(0.008)	(0.008)	(0.007)
Students discuss homework with parents - more than once a week (1=yes)				0.009	0.007	0.008
				(0.022)	(0.021)	(0.021)
TEACHER CHARACTERISTICS						
Female teacher (1=yes)					0.005	0.010
					(0.054)	(0.053)
Years of teaching experience of teachers					-0.005	-0.005
					(0.004)	(0.005)
Teacher has junior high school math teaching certificate (1=yes)					-0.066	-0.066
					(0.050)	(0.049)
SCHOOL CHARACTERISTICS						
School assets (score 0-5)						0.064*
						(0.036)
Student/teacher ratio						-0.008
						(0.010)
Constant	-0.008	0.003	1.699***	1.604***	1.716***	1.516***
	(0.032)	(0.038)	(0.170)	(0.166)	(0.183)	(0.288)
Observations	9,170	9,170	9,168	9,129	9,062	9,062
R^2	0.007	0.186	0.207	0.209	0.212	0.215

Note:

1. SEs clustered at school level
2. *p<0.1; **p<0.05; ***p<0.01

Table 4: Impact of ability-tracking on average student's math self-concept

	Value Added: Math self-concept score					
	(1)	(2)	(3)	(4)	(5)	(6)
Tracking	0.008	0.027	0.025	0.024	0.023	0.039
	(0.050)	(0.054)	(0.054)	(0.053)	(0.051)	(0.049)
Baseline self-concept score		-0.404***	-0.422***	-0.424***	-0.428***	-0.429***
		(0.011)	(0.011)	(0.011)	(0.011)	(0.010)
Teacher incentive treatment (1=yes)	-0.034	-0.041	-0.034	-0.031	-0.024	-0.015
	(0.039)	(0.041)	(0.040)	(0.040)	(0.040)	(0.039)
STUDENT CHARACTERISTICS						
Female student (1=yes)			-0.143***	-0.143***	-0.143***	-0.143***
			(0.020)	(0.020)	(0.020)	(0.020)
Student boarding in school (1=yes)			0.040	0.037	0.029	0.033
			(0.026)	(0.026)	(0.026)	(0.025)
Age of student (in years)			-0.057***	-0.057***	-0.060***	-0.062***
			(0.014)	(0.014)	(0.014)	(0.013)
HOUSEHOLD CHARACTERISTICS						
Mother graduated junior high school (1=yes)				-0.034	-0.025	-0.019
				(0.023)	(0.022)	(0.022)
Mother absent for both semesters (1=yes)				0.010	0.015	0.011
				(0.024)	(0.024)	(0.024)
Father absent for both semesters (1=yes)				-0.004	-0.005	-0.011
				(0.020)	(0.020)	(0.020)
Household assets (score 0-7)				0.008	0.005	0.005
				(0.007)	(0.007)	(0.007)
Students discuss homework with parents - more than once a week (1=yes)				0.044**	0.042*	0.040*
				(0.022)	(0.022)	(0.021)
TEACHER CHARACTERISTICS						
Female teacher (1=yes)					0.044	0.042
					(0.046)	(0.044)
Years of teaching experience of teachers					-0.005*	-0.004
					(0.003)	(0.003)
Teacher has junior high school math teaching certificate (1=yes)					-0.017	-0.015
					(0.039)	(0.038)
SCHOOL CHARACTERISTICS						
School assets (score 0-5)						0.050**
						(0.025)
Student/teacher ratio						0.013**
						(0.006)
Constant	0.017	0.021	0.807***	0.763***	0.859***	0.516**
	(0.033)	(0.034)	(0.182)	(0.185)	(0.190)	(0.231)
Observations	9,170	9,170	9,168	9,129	9,062	9,062
R^2	0.0003	0.197	0.207	0.208	0.213	0.215

Note:
1. SEs clustered at school level
2. *p<0.1; **p<0.05; ***p<0.01

Table 5: Impact of ability-tracking on average student's math anxiety score

	Valued added: Math anxiety score

	(1)	(2)	(3)	(4)	(5)	(6)
Tracking	-0.025	-0.061	-0.059	-0.057	-0.055	-0.058
	(0.037)	(0.042)	(0.040)	(0.040)	(0.038)	(0.038)
Baseline anxiety score		-0.494***	-0.506***	-0.508***	-0.512***	-0.513***
		(0.011)	(0.011)	(0.011)	(0.012)	(0.012)
Teacher incentive treatment (1=yes)	0.040	0.049	0.039	0.040	0.026	0.023
	(0.033)	(0.037)	(0.035)	(0.034)	(0.034)	(0.034)
STUDENT CHARACTERISTICS						
Female student (1=yes)			0.139***	0.136***	0.135***	0.135***
			(0.018)	(0.018)	(0.018)	(0.018)
Student boarding in school (1=yes)			-0.061**	-0.062**	-0.054**	-0.053**
			(0.025)	(0.025)	(0.025)	(0.025)
Age of student (in years)			0.069***	0.064***	0.067***	0.068***
			(0.012)	(0.012)	(0.012)	(0.012)
HOUSEHOLD CHARACTERISTICS						
Mother graduated junior high school (1=yes)				-0.008	-0.020	-0.020
				(0.028)	(0.027)	(0.028)
Mother absent for both semesters (1=yes)				0.009	0.005	0.005
				(0.025)	(0.025)	(0.025)
Father absent for both semesters (1=yes)				0.018	0.017	0.018
				(0.022)	(0.022)	(0.022)
Household assets (score 0-7)				-0.020***	-0.017**	-0.016**
				(0.007)	(0.007)	(0.007)
Students discuss homework with parents - more than once a week (1=yes)				-0.027	-0.025	-0.024
				(0.020)	(0.020)	(0.020)
TEACHER CHARACTERISTICS						
Female teacher (1=yes)					-0.062*	-0.063*
					(0.036)	(0.035)
Years of teaching experience of teachers					0.005*	0.005*
					(0.003)	(0.003)
Teacher has junior high school math teaching certificate (1=yes)					0.009	0.009
					(0.035)	(0.035)
SCHOOL CHARACTERISTICS						
School assets (score 0-5)						-0.023
						(0.033)
Student/teacher ratio						-0.002
						(0.005)
Constant	-0.020	-0.024	-0.949***	-0.818***	-0.893***	-0.770***
	(0.026)	(0.027)	(0.162)	(0.164)	(0.180)	(0.259)
Observations	9,170	9,170	9,168	9,129	9,062	9,062
R^2	0.0005	0.249	0.258	0.259	0.263	0.264

Note:

1. SEs clustered at school level
2. *p<0.1; **p<0.05; ***p<0.01

Table 6: Impact of ability-tracking on high-ability and low-ability class student's math, self-concept and anxiety score

Panel A:	*Value added: Math score*					
	(1)	(2)	(3)	(4)	(5)	(6)
Tracking (High-ability)	-0.300***	-0.120	-0.104	-0.104	-0.097	-0.100

	(1)	(2)	(3)	(4)	(5)	(6)
	(0.074)	(0.076)	(0.074)	(0.074)	(0.073)	(0.075)
Tracking (Low-ability)	-0.082	-0.053	-0.055	-0.053	-0.052	-0.058
	(0.059)	(0.070)	(0.069)	(0.067)	(0.066)	(0.065)
Constant	-0.007	0.003	1.694***	1.600***	1.712***	1.512***
	(0.033)	(0.038)	(0.169)	(0.164)	(0.183)	(0.287)
Observations	9,170	9,170	9,168	9,129	9,062	9,062
R^2	0.011	0.186	0.208	0.210	0.212	0.215
Panel B:	*Value added: Math self-concept score*					
Tracking (High-ability)	-0.022	0.005	0.004	0.005	0.011	0.025
	(0.057)	(0.054)	(0.053)	(0.053)	(0.051)	(0.050)
Tracking (Low-ability)	0.029	0.042	0.040	0.037	0.032	0.049
	(0.054)	(0.060)	(0.061)	(0.061)	(0.059)	(0.056)
Constant	0.017	0.021	0.807***	0.763***	0.859***	0.515**
	(0.033)	(0.034)	(0.182)	(0.185)	(0.190)	(0.231)
Observations	9,170	9,170	9,168	9,129	9,062	9,062
R^2	0.001	0.198	0.207	0.208	0.213	0.215
Panel C:	*Value added: Math anxiety score*					
Tracking (High-ability)	-0.040	-0.095*	-0.094*	-0.094*	-0.100**	-0.103**
	(0.050)	(0.055)	(0.052)	(0.052)	(0.050)	(0.051)
Tracking (Low-ability)	-0.014	-0.038	-0.035	-0.032	-0.024	-0.028
	(0.044)	(0.049)	(0.049)	(0.049)	(0.047)	(0.046)
Constant	-0.020	-0.024	-0.949***	-0.818***	-0.894***	-0.773***
	(0.026)	(0.027)	(0.162)	(0.165)	(0.180)	(0.260)
Observations	9,170	9,170	9,168	9,129	9,062	9,062
R^2	0.0005	0.249	0.258	0.259	0.264	0.264
Baseline score		YES	YES	YES	YES	YES
Student characteristics			YES	YES	YES	YES
Household characteristics				YES	YES	YES
Teacher characteristics					YES	YES
School characteristics						YES

Note:

1. Student characteristics: student's gender (1=female); boarding status (1=yes); and, student's age (in years).
2. Household characteristics: mother graduated junior-high school (1=yes); mother absent for both semester (1=yes); father absent for both semester (1=yes); household assets (out of 7); and, students discuss homework with parents - more than once a week (1=yes).
3. Teacher characteristics: teacher's gender (1=female); years of teaching experience; and, teacher has junior high school math teaching certificate (1=yes).
4. School characteristics: school asset (out of 5); and, student teacher ratio.
5. SEs clustered at school level
6. *p<0.1; **p<0.05; ***p<0.01

Table 7: Heterogeneous impact of ability-tracking tracking (average students)

PANEL A:	*Value added: Math score*			
	(1)	(2)	(3)	(4)
Tracking	-0.081	-0.068	-0.007	-0.096
	(0.061)	(0.072)	(0.075)	(0.069)
Tracking*ability_by_class(Middle1/3)	-0.047			
	(0.045)			
Tracking*ability_by_class(Top1/3)	-0.004			
	(0.059)			
Tracking*Female		-0.014		
		(0.041)		
Tracking*Boarding student			-0.107	
			(0.069)	
Tracking*Top half socioeconomic status				0.041
				(0.065)
Constant	1.767***	1.515***	1.492***	1.524***
	(0.280)	(0.288)	(0.289)	(0.289)
Observations	9,062	9,062	9,062	9,062
R^2	0.221	0.215	0.215	0.215

PANEL B:	*Value added: Math self-concept score*			
Tracking	0.052	0.009	0.067	0.013
	(0.046)	(0.060)	(0.057)	(0.051)
Tracking*ability_by_class(Middle1/3)	-0.020			
	(0.050)			
Tracking*ability_by_class(Top1/3)	-0.016			
	(0.061)			
Tracking*Female		0.061		
		(0.049)		
Tracking*Boarding student			-0.043	
			(0.060)	
Tracking*Top half socioeconomic status				0.051
				(0.054)
Constant	0.005	0.523**	0.506**	0.526**
	(0.237)	(0.232)	(0.235)	(0.230)
Observations	9,062	9,062	9,062	9,062
R^2	0.236	0.215	0.215	0.215

PANEL C:	*Value added: Math anxiety score*			
Tracking	-0.063	-0.063	-0.010	-0.092**
	(0.039)	(0.042)	(0.049)	(0.036)
Tracking*ability_by_class(Middle1/3)	-0.006			
	(0.056)			
Tracking*ability_by_class(Top1/3)	0.017			
	(0.065)			
Tracking*Female		0.010		
		(0.044)		
Tracking*Boarding student			-0.076	
			(0.053)	
Tracking*Top half socioeconomic status				0.065

				(0.054)
Constant	-0.338	-0.769***	-0.787***	-0.756***
	(0.260)	(0.260)	(0.264)	(0.261)
Observations	9,062	9,062	9,062	9,062
R^2	0.275	0.264	0.264	0.264
Baseline score	YES	YES	YES	YES
Student characteristics	YES	YES	YES	YES
Household characteristics	YES	YES	YES	YES
Teacher characteristics	YES	YES	YES	YES
School characteristics	YES	YES	YES	YES

Note:

1. Student characteristics: student's gender (1=female); boarding status (1=yes); and, student's age (in years).
2. Household characteristics: mother graduated junior-high school (1=yes); mother absent for both semester (1=yes); father absent for both semester (1=yes); household assets (out of 7); and, students discuss homework with parents - more than once a week (1=yes).
3. Teacher characteristics: teacher's gender (1=female); years of teaching experience; and, teacher has junior high school math teaching certificate (1=yes).
4. School characteristics: school asset (out of 5); and, student teacher ratio.
5. SEs clustered at school level
6. *p<0.1; **p<0.05; ***p<0.01

Table 8: Heterogeneous impact of ability-tracking (high and low ability class students)

PANEL A:	*Value added: Math score*			
	(1)	(2)	(3)	(4)
Tracking (High-ability)	-0.082	-0.110	-0.083	-0.122
	(0.083)	(0.081)	(0.084)	(0.092)
Tracking (Low-ability)	-0.083	-0.039	0.050	-0.078
	(0.058)	(0.071)	(0.077)	(0.060)
Tracking (High-ability)*ability-by-class(Middle1/3)	-0.090			
	(0.070)			
Tracking (Low-ability)* ability-by-class(Middle1/3)	-0.016			
	(0.048)			
Tracking (High-ability)*ability-by-class(Top1/3)	-0.121			
	(0.095)			
Tracking (Low-ability)*ability-by-class (Top1/3)	0.078			
	(0.061)			
Tracking (High-ability)*Female		0.020		
		(0.070)		
Tracking (Low-ability)*Female		-0.038		
		(0.046)		
Tracking (High-ability)*Boarding student			-0.029	
			(0.082)	
Tracking (Low-ability)*Boarding student			-0.165**	
			(0.073)	
Tracking (High-ability)* Top half socioeconomic status				0.042
				(0.089)
Tracking (Low-ability)*Top half socioeconomic status				0.039
				(0.075)
Constant	1.767***	1.512***	1.484***	1.520***
	(0.279)	(0.287)	(0.288)	(0.288)
Observations	9,062	9,062	9,062	9,062
R^2	0.222	0.215	0.216	0.215

PANEL B:	*Value added: Self-concept score*			
Tracking (High-ability)	0.033	-0.007	0.040	-0.009
	(0.051)	(0.068)	(0.073)	(0.067)
Tracking (Low-ability)	0.064	0.021	0.087	0.029
	(0.057)	(0.066)	(0.070)	(0.054)
Tracking (High-ability)*ability-by-class(Middle1/3)	-0.025			
	(0.054)			
Tracking (Low-ability)* ability-by-class(Middle1/3)	-0.016			
	(0.071)			
Tracking (High-ability)*ability-by-class(Top1/3)	0.008			
	(0.076)			
Tracking (Low-ability)*ability-by-class (Top1/3)	-0.031			
	(0.066)			
Tracking (High-ability)*Female		0.066		
		(0.060)		
Tracking (Low-ability)*Female		0.056		
		(0.059)		
Tracking (High-ability)*Boarding student			-0.024	
			(0.090)	

Tracking (Low-ability)*Boarding student			-0.059	
			(0.064)	
Tracking (High-ability)* Top half socioeconomic status				0.067
				(0.069)
Tracking (Low-ability)*Top half socioeconomic status				0.039
				(0.072)
Constant	0.006	0.523**	0.505**	0.525**
	(0.238)	(0.232)	(0.234)	(0.230)
Observations	9,062	9,062	9,062	9,062
R^2	0.236	0.215	0.215	0.215
PANEL C:	*Value added: Anxiety score*			
Tracking (High-ability)	-0.130**	-0.135**	-0.070	-0.136**
	(0.059)	(0.064)	(0.065)	(0.058)
Tracking (Low-ability)	-0.018	-0.012	0.035	-0.060
	(0.048)	(0.053)	(0.060)	(0.044)
Tracking (High-ability)*ability-by-class(Middle1/3)	0.027			
	(0.083)			
Tracking (Low-ability)* ability-by-class(Middle1/3)	-0.026			
	(0.053)			
Tracking (High-ability)*ability-by-class(Top1/3)	0.058			
	(0.098)			
Tracking (High-Low)*ability-by-class (Top1/3)	-0.009			
	(0.060)			
Tracking (High-ability)*Female		0.066		
		(0.063)		
Tracking (Low-ability)*Female		-0.031		
		(0.060)		
Tracking (High-ability)*Boarding student			-0.053	
			(0.062)	
Tracking (Low-ability)*Boarding student			-0.096	
			(0.072)	
Tracking (High-ability)* Top half socioeconomic status				0.066
				(0.063)
Tracking (Low-ability)*Top half socioeconomic status				0.063
				(0.078)
Constant	-0.341	-0.769***	-0.790***	-0.759***
	(0.260)	(0.261)	(0.265)	(0.261)
Observations	9,062	9,062	9,062	9,062
R^2	0.276	0.264	0.264	0.264
Baseline score	YES	YES	YES	YES
Student characteristics	YES	YES	YES	YES
Household characteristics	YES	YES	YES	YES
Teacher characteristics	YES	YES	YES	YES
School characteristics	YES	YES	YES	YES

Note:

1. Student characteristics: student's gender (1=female); boarding status (1=yes); and, student's age (in years).
2. Household characteristics: mother graduated junior-high school (1=yes); mother absent for both semester (1=yes); father absent for both semester (1=yes); household assets (out of 7); and, students discuss homework with parents - more than once a week (1=yes).

3. Teacher characteristics: teacher’s gender (1=female); years of teaching experience; and, teacher has junior high school math teaching certificate (1=yes).
4. School characteristics: school asset (out of 5); and, student teacher ratio.
5. SEs clustered at school level
6. *p<0.1; **p<0.05; ***p<0.01

Appendix

Figure A1: Variance in math scores across each classroom using standing math scores in each classroom

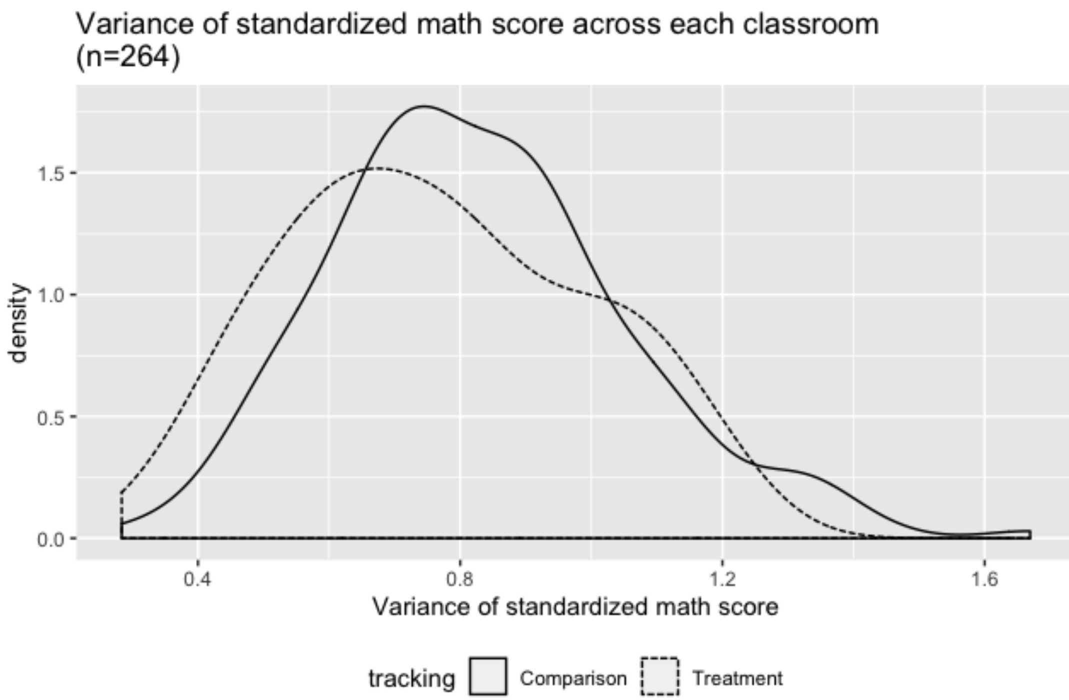


Table A1: Survey questions to measure math self-concept and anxiety as described in PISA (2012)[1]

Math Self-concept	1. I am just not good at math 2. I get good grades in math 3. I learn math quickly 4. I have always believed that math is one of my best subjects 5. In my math class, I understand even the most difficult work
Math anxiety	1. I often worry that it will be difficult for me in math classes 2. I get very tense when I have to do math homework 3. I get very nervous doing math problems 4. I feel helpless when doing a math problem 5. I worry that I will get poor grades in math

1. Students answer on a 4-point Likert of strongly agree, agree, disagree, and strongly disagree

Table A2: Distribution of treatment group across high-ability and low-ability students

	Treatment	High-ability	Low-ability
	(1)	(2)	(3)
Panel A: Distribution of students across high-ability and low-ability groups (with students' that are missing outcome variable values)			
Total students	2285	939	1346
Panel B: Final sample of treatment group across high-ability and low-ability groups (without students' that are missing outcome variable values)			
Total students	2164	884	1280
Proportion	100%	40.85%	59.15%

Table A3: Number of students with missing values for outcome variables

	Math score	Self-concept	Anxiety
Baseline score missing	0	34	38
Endline score missing	468	488	490
Either baseline or endline score missing	468	518	524

Note:

1. Students with either math, self-concept or anxiety score missing at the baseline or endline = 570. These were dropped from further analysis.

Table A4: Balance test between students dropped from analysis and the sample

	Overall (n=9740)	Dropped (D) (n=570)	Sample (S) (n=9170)	p-value (D v/s S)
	(1)	(2)	(3)	(4)
OUTCOME VARIABLES				
Baseline standardized math score	0.0583 (1.00)	-0.235 (1.03)	0.0765 (0.998)	p<0.01
Baseline standardized self-concept score	0.0106 (1.00)	-0.177 (1.03)	0.0216 (0.997)	p<0.01
Baseline standardized anxiety score	-0.0189 (1.01)	0.107 (1.03)	-0.0262 (1.01)	p<0.01
STUDENT CHARACTERISTICS				
Female student (1=yes)	0.494 (0.500)	0.408 (0.492)	0.499 (0.500)	p<0.01
Student boarding at school (1=yes)	0.598 (0.490)	0.613 (0.487)	0.597 (0.491)	0.428
Age (in years)	13.0 (0.964)	13.6 (1.16)	13.0 (0.940)	p<0.01
HOUSEHOLD CHARACTERISTICS				
Mother's graduated junior high school (1=yes)	0.250 (0.433)	0.231 (0.422)	0.251 (0.434)	0.276
Mother absent for both semesters (1=yes)	0.150 (0.357)	0.167 (0.375)	0.149 (0.356)	0.644
Father absent for both semesters (1=yes)	0.413 (0.492)	0.402 (0.493)	0.413 (0.492)	0.818
Household assets (score 0-7)	2.74 (1.70)	2.67 (1.75)	2.75 (1.70)	0.274
Students discuss homework with parents - more than once a week (1=yes)	0.630 (0.483)	0.589 (0.492)	0.633 (0.482)	0.039
MATH TEACHER CHARACTERISTICS				
Female teacher (1 =yes)	0.359 (0.480)	0.374 (0.484)	0.358 (0.479)	0.424
Years of teaching experiences	9.67 (7.07)	10.7 (7.05)	9.61 (7.07)	p<0.01
Teacher has junior high school math teaching certificate (1=yes)	0.725 (0.447)	0.744 (0.437)	0.723 (0.447)	0.272
SCHOOL CHARACTERISTICS				
School assets (score 0-5)	4.69 (0.622)	4.72 (0.652)	4.69 (0.620)	0.250
Student/teacher ratio	8.76 (2.90)	8.89 (3.29)	8.75 (2.87)	0.348

Note:

1. Students with either baseline or endline missing value for outcome variables were dropped from analysis
2. N = number of observations
3. p<0.1; p<0.05; p<0.01