Early between-school tracking and gender gaps in achievement

Name: Junying Zhao

Student Number: 612803

Institution: Erasmus University Rotterdam

Master Program: Social Inequalities

Supervisor: Sjaak Braster

Abstract

The between-school tracking system is responsible for assigning students to academic or applied tracks based on their educational achievement. However, the time that the first between-school selection happened may have different influence on gender gap in achievement. This study assumes that early between-school tracking will widen the gender gap in overall achievement from primary school to secondary school. Besides, I also hypothesize that female will expand advantages in reaching achievement in secondary school by early between-school tracking, whereas early tracking make male outperform female more in secondary school. I utilized the data from PIRLS, TIMSS and PISA, and matched the waves based on birth cohort of respondents and the year that surveys carried out. This sample contains 44 countries (15 countries with early tracking and 29 countries with late tracking), 22 matches and 2699532 observations. Difference-in-differences model and meta-analysis are used in data analysis. As I expected, the results show significant gender gap changes because of early between-school tracking in reading achievements and mathematics achievement. But empirical evidence did not show significance in science achievement and overall achievement.

Key words: Tracking; Gender gap; Difference-in-differences; educational achievement; meta-analysis

Early between-school tracking and gender gaps in achievement

Introduction

Education inequalities, both structural and systematic, have long been a central focus of scholarly research globally. Among these inequalities, gender has emerged as a critical variable of interest, with numerous studies examining gender gaps in various educational aspects, such as STEM fields, teachers' attitudes, and academic outcomes (e.g., Organization for Economic Co-operation and Development [OECD], 2015; Charles, 2017; Reinhold et al., 2018; Mann et al., 2015). Studies, such as those conducted by the OECD in 2011 and 2015, have indicated that boys tend to underachieve relative to girls across almost all OECD countries. The degree of this gender difference, however, varies due to institutional reasons, including the age at which students are selected for applied or academic educational tracks. (Contini & Cugnata, 2020; Terrin & Triventi, 2022; Van Houtte, 2004). One of the most prominent institutional factors influencing educational achievement is the implementation of a tracking system, which divides students into different types of schools (academic tracking and vocational tracking) based on their aptitude, preferences, and, in some countries (especially early-tracking countries), teachers' recommendations (Nürnberger et al., 2016). This paper aims to contribute to this topic by investigating the effect of early between-school tracking on gender gaps in student achievement.

While all educational systems use between-school tracking at some stage of students' academic career, there are variations in timing. Some systems implement early tracking (e.g., at 4th grade), while others employ late tracking (e.g., tracking occurs at age 18 in the United States). Proponents of early tracking argue that it facilitates the individual development of students, matching them with suitable educational environments based on their preferences

(Lassibile & Gomez, 2010). Conversely, supporters of late tracking contend that starting tracking at a young age, when students have not fully developed cognitive abilities, can result in unequal distribution of educational resources and expectations, affecting educational attainment and labor market opportunities (Moller & Stearns, 2012; Scheeren et al., 2018; Scheeren & Bol, 2022; Schindler et al., 2023). Another significant concern related to the tracking system, which forms the primary focus of this research, is that the assignment of students to different tracks depends not only on their quantified exam scores but also on factors such as socio-economic status (SES) and gender. This potential bias in the tracking process can further amplify educational inequalities between different groups (;van de Werfhorst, 2018; Batruch et al., 2019; Strello et al., 2021). Furthermore, differences in the content and quality of different tracks may contribute to larger achievement gaps between groups in early-tracking countries compared to late-tracking countries (Strello et al., 2021;Scheeren & Bol, 2022).

Unlike existing research that often emphasizes the influence of SES, this study aims to explore the impact of early between-school tracking on students' achievements, with a particular focus on gender gaps in achievements. Early-tracking students may exhibit gender differences in the choice of schools during the initial selection process (OECD, 2015). While previous literature has highlighted the advantages of girls in cognitive and noncognitive abilities during elementary school period, leading them to choose higher tracks and accumulate advantages in subsequent academic careers (Jürges & Schneider, 2011; Alam, 2022), it remains unclear how these gender differences manifest in specific subject categories, such as reading, mathematics, and science achievement.

Although some studies have demonstrated that girls would have higher educational achievement in early tracking(Strello et al., 2021), as well as studies demonstrating that girls' reading achievement is generally stronger than that of men(Herd et al., 2019), the clear results about gender gaps in specific subjects still lacking. This research seeks to address these gaps by investigating whether early between-school tracking further amplifies existing gender differences in academic achievement, particularly in reading, mathematics, and science. Unlike most cross-sectional research designs commonly employed in tracking studies, this study will utilize the difference-in-differences design, allowing for better control of uncontrollable context variables.

Theoretical Framework

I will illustrate the theoretical frame work in three aspects: the benefits and risks of early tracking, gender gaps in early tracking, and self-fulfilling prophecy.

The benefits and risks of early tracking

Proponents of the tracking system argue that early between-school tracking can positively impact overall learning levels and school performance (Domina et al., 2019). In a tracking system, students are placed in a school environment tailored to their needs, receiving instruction appropriate to their level of ability. This personalized approach maximizes the potential of each student, enhancing their academic achievements (Lassibile & Gomez, 2010). The underlying logic of tracking is to ensure that students' learning environment matches their academic potential. From an institutional perspective, a well-designed tracking system should prompt students to perform at higher academic levels compared to non-tracking systems when they are properly assigned to tracks that suit their ability levels (Betts, 2001).

Additionally, early research on educational systems suggests that a more homogeneous learning environment can improve learning efficiency. This conclusion, however, assumes an idealized tracking system, where the process of classifying students is solely based on their academic ability (Gamoran & Mare, 1989). Under such assumptions, a well-developed tracking system not only positively affects the efficiency of the education system but also structurally compensates for achievement gaps between low-performing and high-performing students. After track assignment, inequality is reduced if students in lower tracks improve their grades more than those in higher tracks (Terrin & Triventi, 2022).

Despite its advantages, the tracking system can also lead to certain inequalities. In particular, early tracking may result in problems with track assignment, with earlier tracking ages being associated with a higher likelihood of assignment issues (Betts, 2001; Brunello &Checchi, 2007). Furthermore, studies have pointed out that the negative consequences of the tracking system, especially early tracking, extend beyond achievement dispersion and gaps and can lead to inequality of educational opportunities (van de Werfhorst & Mijs, 2010). Since education is inevitably affected by students' socioeconomic status and intergenerational capital, students from disadvantaged socio-economic backgrounds may suffer from unfair tracking assignments (van de Werfhorst, 2018), particularly in systems where assignments depend on teacher recommendations rather than students' independent choices (e.g., the first tracking selection in the Netherlands), creating structural inequalities that are deeply embedded in the social system and challenging to address (Contini & Scagni, 2011).

Even in self-selection systems, students' academic expectations can be influenced by social context and stereotypes (Alam, 2022). For instance, gender differences in STEM aspirations have been observed, with men having higher STEM aspirations due to traditional gender stereotypes associating them with higher achievements in mathematics and science (Batruch et al., 2019; Nürnberger et al., 2016). The topic of gender differences will be further explored in the next section.

In summary, tracking system, especially early tracking can intensify gaps among some groups of students in different aspects, but can also improve overall educational achievements.

Gender gap in early tracking.

According to the PIRLS report, girls tend to outperform boys in reading performance across most countries with available data (Mullis et al., 2017). Moreover, numerous studies have consistently shown that girls exhibit a more substantial advantage in language skills compared to mathematical sciences (O'Dea et al., 2018; Voyer & Voyer, 2014). This pattern is further supported by findings from PISA, where, in mathematics, most countries either reported no significant gap or demonstrated better results for girls on average (OECD, 2015). These findings suggest that girls generally perform better in standardized tests, which may be attributed to biological and developmental differences between the sexes, leading to earlier development of cognitive and non-cognitive abilities beneficial for high academic achievements (Geven et al., 2017).

In addition to outperforming boys in school achievement, girls also receive more school recommendations than boys in several European countries (Jürges & Schneider, 2011). For instance, in the Netherlands, teachers tend to have lower academic expectations of boys in primary school (Timmermans et al., 2015), and in Austria, more girls than boys are assigned to higher tracks at the end of primary school (Bacher, 2009).However, as mentioned in the introduction section, most studies examining the relationship between early between-school tracking and gender gaps in educational achievements utilize a cross-sectional design(Steinmann et al., 2023). This design has limitations in fully controlling for the complex and diverse factors affecting educational achievements, leading to potential errors in estimating the true relationships. To address this issue, some researchers have adopted a longitudinal approach

(Pekkarinen, 2008) to explore the impact of tracking on gender gaps in educational achievements. For instance, Scheeren and Bol (2022) investigate the association of standardized achievement scores, gender, tracking status, and two-way and three-way interactions among these variables. They use separate models for reading and math achievement and find that early tracking may contribute to achievement gaps in girls' strengths in reading, but the effect on mathematics is not statistically significant. However, this study utilizes multiple survey datasets, including PIRLS, TIMSS, PISA, PIAAC, and ESS, which vary in scale and measurement, making it challenging to quantify the data to the same dimension fully.

In summary, early between-school tracking may have a positive influence on the educational achievements of girls. However, gender gaps in achievements are still significant consideration. Therefore, I hypothesize early between-school tracking can make girls gain more advantages and widen the gender gap in achievements than non-tracking (late tracking) (hypothesis 1).

Self-fulfilling prophecy

Merton (1968) introduced the concept of self-fulfilling prophecy to illustrate the potential influence of groups on individuals. Self-fulfilling prophecies occur when people’s or groups' beliefs or expectations about specific outcomes lead those outcomes to come true. Thus, individuals' ideas can significantly influence their actions. From this perspective, social groups at various levels, from broad gender stereotypes at the structural level to middle schools at the institutional level, including teachers and parents, may hold different expectations for men's and women's achievements in various disciplines. Research has consistently shown that female students often have lower aspirations in mathematics and science, which may be related to the low expectations others hold for them. Schools, being small societies, expose students to potential influences from teachers and classmates, which can be significant yet challenging to detect. When the majority of individuals around a student expect that women excel in reading and men excel in mathematics and science, students of different genders may unconsciously align their behaviors with the expectations of the group.

Similarly, the theory of self-concept assimilation posits that individuals continuously monitor their relative distance from the heart of the groups to which they subjectively identify as belonging. Maintaining a considerable distance from the heart of a group can lead to negative emotions, and individuals may seek ways to reduce this distance to avoid emotional discomfort. Consequently, individuals may actively adjust their behaviors or opinions to align with the prevailing beliefs of the groups to which they perceive they belong. Applying this theory to the subject of this paper, we anticipate that gender differences in self-concept will align with common stereotypes. For instance, it is expected that women may perceive themselves as better in subjects with emotional solid orientation, such as reading, while men may see themselves as more assertive in subjects requiring logical solid thinking, such as mathematics and science. In the context of tracking, gender-segregated groups may form, intensifying the gender-based hints on performance.

In summary, the theoretical framework suggests that self-fulfilling prophecies and self-concept assimilation may contribute to gender gaps in educational achievements in the context of early between-school tracking. The expectations and beliefs held by social groups and individuals within schools may influence students' academic performance and aspirations, potentially leading to the widening of gender gaps in specific subjects. I hypothesize that early tracking system will make female advantage in reading achievements further widen in secondary level(Hypothesis 2a). In contrary, male will keep advantage in mathematics achievement (hypothesis 2b) and science achievement (hypothesis 2c) even increase the gender gap (male outperform female) in secondary school because of early between-school tracking.

Methods and Analytical Approach

Data

I combined 3 large-scale cross-national surveys including the Progress in International Reading Literacy Study (PIRLS), the Trends in International Mathematics and Science Study (TIMSS), and the Program for International Student Assessment (PISA). I chose the waves from 2006 to 2019 and matched them together. Specifically, the Primary level data are derived from PIRLS and TIMSS. Then the Secondary level data is from PISA, the participants of which are all 15 years old.

The International Association for the Evaluation of Educational Achievement (IEA) conducted two assessments I used in this article. One is PIRLS, a large-scale survey measures 4th-grade students' reading achievement in primary school. The other one is TIMSS, measures student achievement in mathematics in the fourth and eighth grades. PISA, developed by the Organization for Economic Coordination and Development (OECD), tests cognitive skills in mathematics and reading of 15-year-old students (TIMSS & PIRLS International Study Center, 2021a, 2021b; OECD, 2021).

I followed Hanushek & Oßmann (2006) and Hanushek & Wößmann (2006) articles and matched the data evaluated by primary and secondary schools respectively as 22 matches (7 matches for reading, 8 matches for mathematics and 7 matches for science). The matching followed two standards. Firstly, I match the survey waves which were carried out at the same or similar calendar year, which meant that I could consider the possible periodic effects (Blanchard et al., 1977). Secondly, I considered of birth cohort to match the waves that the respondents of them birth in approximately same calendar year. Since PISA can only provide data for reading and mathematics achievements, I have matched three research combinations, namely PIRLS (primary) with PISA (secondary), TIMSS (primary) with Pisa (secondary) and TIMSS (primary 4^{th} grade) with TIMSS (secondary 8^{th} grade) 22 methods to match primary level and secondary level school data (see Table2 for details). In each set of data, we select 44 countries, 15 early-tracking countries and 29 late-tracking countries (see Table1 for details).

Variables

Independent variables

Gender: The key variable to measure gender gap in a DiD model. I assign girls=1 and boys=0 in dataset.

Educational stage: The time variable in DiD models. I assign secondary school=1 (after early-tracking happened) and primary school =0 (before early-tracking happened).

Tracking system: The variable about if a country has an early-tracking system (early-tracking country=1; late-tracking country=0). I define the early-tracking system as the treatment group and the late-tracking system as the control group. According to the information on tracking age from the OECD, a country's tracking system is defined as early tracking when students are tracked before the age of 15. So, I select 15 as the dividing of age.

Dependent variables

Reading achievement: data from PIRLS PISA. The ability about reading and writing texts in school and their lives. Mathematics achievement: data from TIMSS PISA. The ability about

Table 1 *Overview of countries and their tracking age*

EARLY TRACKING		LATE TRACKING	
COUNTRY	TRACKING AGE	COUNTRY	TRACKING AGE
Argentina	12	Albania	15

Austria	10	Australia	16
Belgium	12	Botswana	16
Bulgaria	14	Canada	18
Czech	11	Chile	16
Germany	10	Croatia	15
Hungary	10	Cyprus	15
Ireland	12	Denmark	16
Italy	14	Finland	16
Luxembourg	12	France	16
Netherlands	12	Greece	15
Romania	14	Hongkong	16
Singapore	12	Iceland	16
Slovak Rep.	10	Indonesia	16
Turkey	14	Israel	15
		Japan	15
		Morocco	15
		Norway	16
		Philippines	16
		Poland	15
		Portugal	15
		Slovenia	15
		South Africa	15
		Spain	15
		Sweden	16
		Taiwan	15
		Ukraine	15
		U. K	16
		U. S	18

Note: Tracking age reflects the typical age in the tracking grade. Early tracking countries are countries with a tracking grade below Grade 8 for matches with TIMSS data and tracking age below 15 for matches with PISA data.

Table2 *Overview of survey waves matching*

	Primary level	Secondary level
	PIRLS—PISA matches (Reading achievement test)	

Number	Year	Grade	Year	Age
1	2006	4th grade	2006	15
2	2006	4th grade	2009	15
3	2006	4th grade	2012	15
4	2011	4th grade	2012	15
5	2011	4th grade	2015	15
6	2016	4th grade	2015	15
7	2016	4th grade	2018	15
	TIMSS—PISA matches (Mathematics achievement test)			
	Year	Grade	Year	Age
8	2007	4th grade	2006	15
9	2007	4th grade	2009	15
10	2007	4th grade	2012	15
11	2011	4th grade	2012	15
12	2011	4th grade	2015	15
13	2015	4th grade	2015	15
14	2015	4th grade	2018	15
15	2019	4th grade	2018	15
	TIMSS—TIMSS matches (Science achievement test)			
	Year	Grade	Year	Grade
16	2007	4th grade	2007	8th grade
17	2007	4th grade	2011	8th grade
18	2011	4th grade	2011	8th grade
19	2011	4th grade	2015	8th grade
20	2015	4th grade	2015	8th grade
21	2015	4th grade	2019	8th grade
22	2019	4th grade	2019	8th grade

calculating and interpreting mathematical information and solving mathematical problems in school and their lives. Science achievement: data from TIMSS. The ability about understanding and using scientific knowledge in school and their lives.

It should be noted that the raw test scores of the three large-scale assessments are not available in dataset. I therefore use multiple plausible values which are estimations based on students' responses to the tests and their background information. PISA, PIRLS and TIMSS datasets all have five plausible values representing reading mathematics and science overall achievements respectively. I also applied senate weights (sampling weight) to make the results more representative and generalized.

A problem needs to be solved is all these dependent variables will get from three different assessments, and they have different achievement scales to them. So, the step of standardization is mandatory. In this research, I use z-score technique to standardize the achievements that calculated from plausible values to make the data of different level comparable. So that the achievements data fit a distribution with a mean of 0 and a variance of 1.

Analytical approach

Gender gap in student achievement

Since gender gap is not a given value in the datasets, I need to calculate the gender gap first. Firstly, I defined the gender gap in achievement as the level female outperform male. So, the formula is:

$$G = \frac{R_f - R_m}{0.5\left(SD_f + SD_m\right)}$$

R_f and R_m represent the average achievement score of reading mathematics and science for female and male students. SD_f and SD_m imply the standard deviation for female and male students. So, in the data analysis process, when *G* have a value of 0 means that girls and boys have equal scores and a value > 0 means girl outperformed boys. Conversely, when *G* have a negative value means that boys outperformed girls. In Stata it can also be interpreted by effect size, which is Cohen’s D used in this research. Cohen's D is a parameter for comparing mean differences. It defines the effect of the effect through the ratio of the average difference and the ratio of the sample standard.

Difference-in-Differences design

As mentioned earlier, a large number of studies have employed cross-sectional design, which is difficult to control covariant and endogeneity problem. Therefore, in order to avoid the effects of various social contexts and ideology, which are difficult to quantify and control, this article partly followed Ruhose & Schwerdt (2016) and Scheeren & Bol (2022) to employ 'Difference-in-Differences' approach.

Difference-in-differences approach has been used for many years to assess policy effects, the basic interpretation of it is by comparing the difference between the control group and the treatment group before and after the implementation of the policy, a DiD statistic reflecting the effect of the policy is constructed. In this article, I compare the gender gap in reading mathematics and science achievements of primary-school students with secondary-school students in countries that implement different tracking policies. In this comparison, two differences appeared. One is the difference between primary and secondary school students, another one is the difference of early-tracking and late-tracking countries. The basic DiD with a continuous time variable is based on a parallel assumption, assume the two groups that compare without policy intervention will develop in the same trend. In this article, when tracking does not happen, students in early-tracking countries as a treatment group and students in late-tracking countries as a control group will change in the same trend. Three sets of DID regression have been performed, which are the DID analysis of reading, mathematics, and science achievements. I checked the tracking effects through changes in the gender gap between two time slots T0 (before tracking occur) and T1 (after tracking occur) by comparing the changes in the gap between the gender gaps between the late-tracking countries and early-tracking countries. By controlling the gender gap between T0, most of the non-observed cross-national or endogeneity changes can be ignored. Since there are no countries began tracking before grade 4, PIRLS and TIMSS data (4th grade) is used as T0, PISA data and TIMSS (8th grade) data is used as T1. In other words, this method can compare the gender gap between the grade of grade 4 and the gender gap between the 15 -year -old age group. The general equation is:

$$Y_{ics} = \beta_0 + \beta_1 TI_s + \beta_2 G_i + \beta_3 TR_c + \beta_4 TI_s G_i + \beta_5 TI_s TR_c + \beta_6 G_i TR_c + \beta_7 TI_s TR_c G_i + \epsilon_{\text{cns}}$$

Y_{ics} reflects standardized achievement score of three subjects (reading mathematics and science) of students (i) in multiple countries (c) in educational stage (s). G_i (0=male, 1=female) describes gender achievement in different timeslot and tracking system. TR_c represents different tracking system (0 = late tracking, 1 = early tracking) of countries c. TI_s describes the educational stage (0= primary school, 1=secondary school). β_{0-7} mean the main effects of single variable and interactions of the three key variable. The equation includes a three-way interaction '$TI_s TR_c G_i$', which means β_7 is the key DiD coefficient I want to measure. ϵ_{cns} means the fixed effects including country and matches in this article. As a result of absorbing fixed-effects of country, the single variable TR_c become meaningless in the regression because of collinearity. To analyze gender gaps' transition in reading, mathematics, and science achievements, and also test if the single match would show same effects, I will do the n = 7 PIRLS → PISA matches regression analysis. The regression will be replicated to know about the gender gaps in mathematics and science achievements for the n = 8 TIMSS → PISA matches and the n = 7 TIMSS → TIMSS matches. To estimate the overall gender gap differences, the DiD model will be applied on all data with fixed effect country and matches. This result can be mutually confirmed with synthesized result of each match which will be explained in next section.

Meta-analysis to synthesize matches per outcome

As mentioned in DiD model, I run DiD equation for all matches. So, the regression of each match produces a key coefficient β_7 . To synthesize the key coefficients in order to estimate the overall gender gap differences affected by early tracking from primary school to secondary school, I would use meta-analysis with precomputed coefficients and standard errors for each outcome, which is also the basic meta-analysis (Fisher et al., 2022). For each outcome, I need to create weight each coefficient because the number of observations of each match is different as well as some other factors could affect the result. Since matches already combined data from two survey, the error and difference are more likely from single studies rather than between studies. Thus, I choose fixed effect model and the weight can be understood as the inverse of standard error.

$$W_{\beta n} = 1/{SE_{\beta n}}^2$$

In this equation, $W_{\beta n}$ indicate coefficient (β) weight of each match (n), SE stands for standard error. Then the mean of β and standard error of each outcome can be calculated by employing the equations:

$$\bar{\beta} = \frac{\sum(W_{\beta n} * \beta)}{\sum W_{\beta n}} \qquad \bar{SE} = \sqrt{\frac{1}{\sum W_{\beta n}}}$$

Thus, I can use the synthesized results of $\beta$and SE to examine if the mean key effect is significantly different from zero. This analysis would be replicated for all three outcomes.

Results

Descriptive results

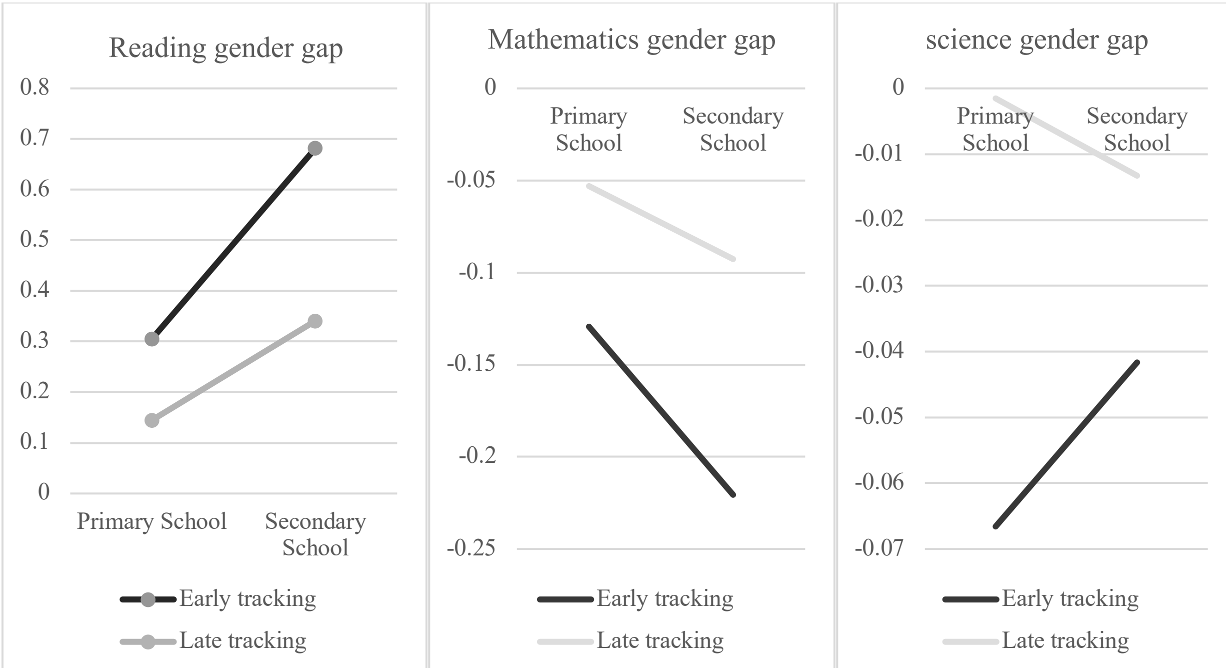
Table3 *Descriptive statistics: reading mathematics and science achievements and effect size by gender, educational stage and tracking system*

	tracking	Female	Male

educational stage	system	Mean	Std,dev.	N	Mean	Std.dev.	N	Cohen's D
reading								
Primary school	early	.6014	.677	69,668	.4907	.699	71,685	.159***
Secondary school	late	.1031	.927	423187	-.2262	1.002	423563	.144***
Primary school	late	.2231	1.024	145,896	.0708	1.077	149,140	.332***
Secondary school	early	.0341	.959	244076	-.3067	1.038	248929	.338***
math								
Primary school	early	.3273	.850	90,480	.3931	.873	93,345	-.072***
Secondary school	late	-.1140	.940	425958	-.0239	1.004	426403	-.057***
Primary school	late	.0718	1.013	174,333	.1276	1.087	179,964	-.125***
Secondary school	early	-.1531	.988	244076	-.0229	1.044	248929	-.086***
science								
Primary school	early	.2395	.767	90,480	.2906	.801	93,345	-.065***
Secondary school	late	-.1961	.995	145121	-.1825	1.056	145461	-.057***
Primary school	late	-.0475	1.021	174,333	-.4591	1.091	179,964	-.032***
Secondary school	early	.2451	.867	. 43171	.2707	.930	44962	-.011***

Noted: This descriptive statistics are summarized from PIRLS, TIMSS (both 4^{th} grade and 8^{th} grade) and PISA dataset, detailed introduction of the variables can be found in 'Data' section. The Cohen's D column is the result of effect size which represented the gender gap with female achievement minus male achievement.

Figure1 *Overview of the gender gap by tracking system and educational stage*



Noted: gender gap is the estimated value from effect size and the flow chart show the change happened in gender gap between primary school and secondary school in early-tracking and late-tracking countries. The specific value can be found in the last column of Table3.

Table2 demonstrated an overview of the data including average achievement scores, standard deviations and the number of observations. What should be noticed is the results of effect size. Effect size based on mean comparison, which is the achievement scores of girls minus the achievement scores of boys. Thus, positive value of it means girls outperform boys, negative value captures boys outperform girls.

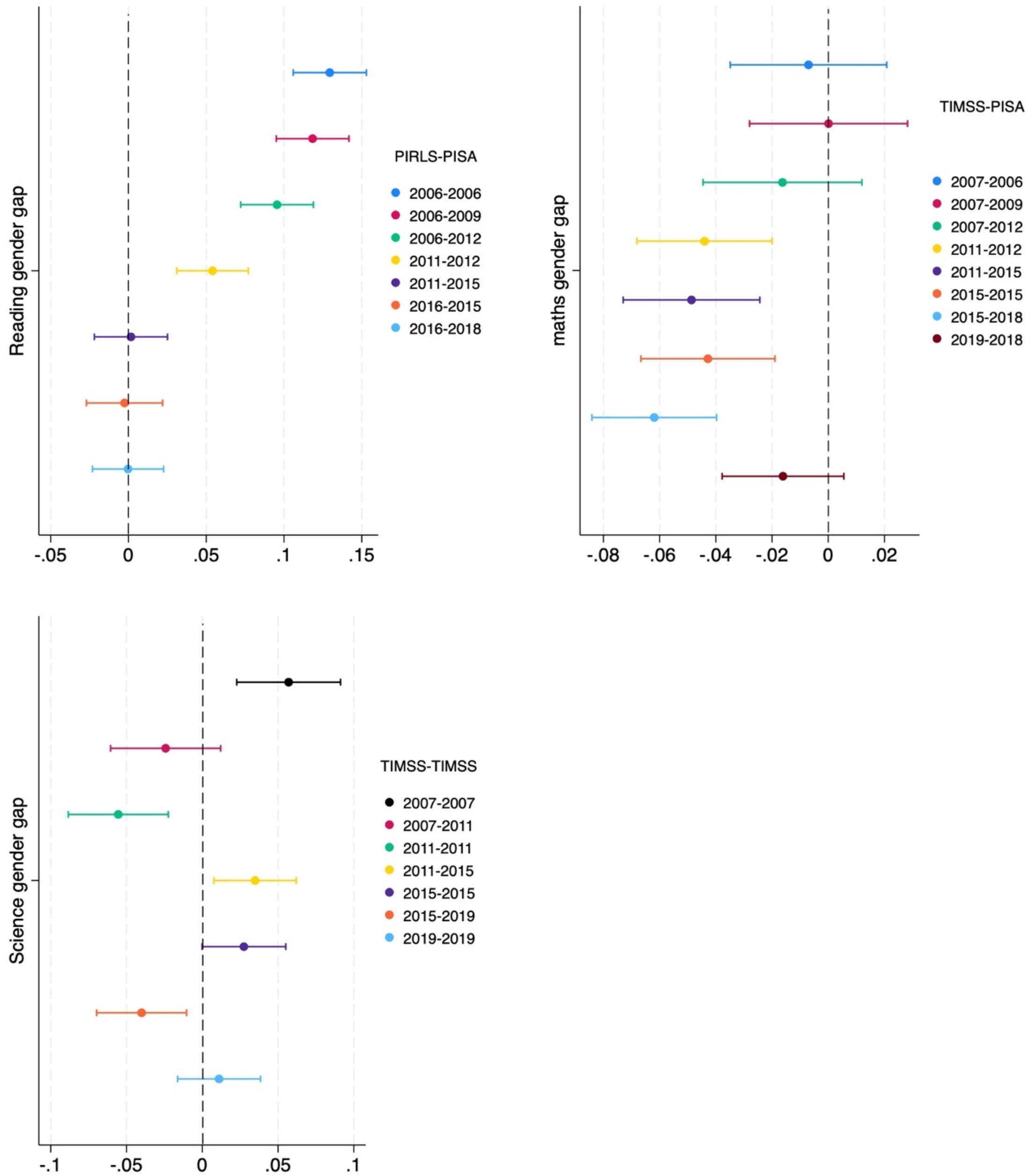
In order to show the result of effect size and its changing trend more clearly, I made a flow chart (figure1). It can be seen from figure 1 left panel that the gender gap changes in early-tracking countries are more obvious than those in late-tracking countries, which means that the girls followed early tracking system have advantage in reading and this advantage continue to expand at secondary level compared with girls in late-tracking countries. The gender gap of mathematics achievements showed a different trend. The overall math performance of girls in the primary school period is lower than that of boys, and the early tracking system further increases the disadvantage of girls in mathematics. However, the gender gap in science achievements showed a bigger difference. Compared with the widening gender gap in late-tracking countries at secondary level, the early tracking system narrows the gender gap, the disadvantage of girls in science performance is narrowed. Although this may be because the gender gap in science achievement is quite small combining numeric value.

Difference-in-differences model results

Results per match

I employ the difference-in-differences model in 'method and analytical approach' section. I created a dummy variable named 'post' to represent binary educational stage indicator (0=primary school, 1=secondary school) and a dummy variable named 'treatment' stands for binary tracking system indicator (0=late tracking, 1=early tracking). The regression command use the three-way interactions of two dummy variables as well as gender variable as the independent variable and standardized reading, mathematics and science achievement scores as outcome variables by (n=22) matches. I replicated the regression for all study matches and three different outcomes. The coefficients of the three-way interaction (female#early tracking#secondary school) show the gap in achievements of different outcome. Figure2 displays the estimated gender gap effected by tracking system and the confidence intervals. In reading achievement gender gaps panel, three lines of confidence interval touched the dash line.

Figure2 *Overview of DiD main effect per match and outcome variable*



Note: PRILS is the abbreviation of Progress in International Reading Literacy Study, TIMSS is Trends in International Mathematics and Science Study; PISA is Programme for International Student Assessment. The value on x-axis represents regression coefficients of each match and the capped lines imply confidence intervals. These coefficients reflect the main DiD effects, as they displayed the difference that achievement gender gaps were affected by early tracking in secondary school compared to late tracking (which was not assigned when the assessments happened) after making the gender gap in primary school controlled. Positive value implies increasing achievement gender gaps.

perpendicular to x-axis, which means there are no significant difference of gender gap in the three matches. However, there are still four matches are positive parameter different from zero.

In mathematics achievement gender gap panel, half of parameters show negative difference from zero and half not. The results of science achievement gender gaps are different from the first two panels with both positive and negative parameters without including zero, which imply girls in two matches (TIMSS 2007-TIMSS2007 and TIMSS2011-TIMSS2015) had advantage of science achievement in secondary level by early tracking.

Results across all matches

Table4 *Results of overall DiD model by outcome*

VARIABLES	Reading achievement	Math achievement	Science achievement
Female	0.159***	-0.0572***	-0.0113***
	(0.00213)	(0.00220)	(0.00207)
Female # Early tracking	-0.0491***	-0.00680*	-0.0421***
	(0.00370)	(0.00369)	(0.00349)
Secondary school	-0.466***	-0.247***	0.0727***
	(0.00198)	(0.00197)	(0.00225)
Female # Secondary school	0.168***	-0.0214***	0.0127***
	(0.00267)	(0.00266)	(0.00302)
Early tracking # Secondary school	-0.230***	-0.00163	-0.104***
	(0.00335)	(0.00327)	(0.00434)
Female # Early tracking # Secondary school	0.0572***	-0.0358***	0.00798
	(0.00454)	(0.00444)	(0.00582)
Constant	0.228***	0.212***	-0.0124***
	(0.00127)	(0.00128)	(0.00121)
Observations	2,849,449	3,183,597	1,589,030

Note: This table indicates the main DiD effect coefficient β_7 with statistically significance and standard errors. Country and match (n=22) were absorbed as fixed effects. '#' indicates the interaction variable. Standard errors in parenthesis *** p<0.01, ** p<0.05, * p<0.1

I employ DiD model to test if early tracking system affect the differences in the reading, mathematics and science achievement gender gap from primary school to secondary school (before late tracking happened). As mentioned in equation, I absorbed country as space fixed effects and matches as time fixed effect, so that the results can exclude the influence of these two factors. The regression results are displayed in Table4. Although the three-way interactions of gender, tracking and educational stage is the key variable in this research, the other variables in the table still need to be interpreted. All the two-way interactions are the differences that have been existed. For instance, the interaction between gender and secondary school means the difference of achievements in primary school and secondary school in late tracking countries. And the interactions between gender and tracking means the difference in achievement gender gaps between early-tracking countries and late-tracking countries in primary school.

The three-way interactions are the key variable to indicates the topic of this research, that we can see from the table, have statistically significance and are different from zero in reading achievements. The coefficient in reading achievement is positive, although the effect value is 0.0572, which is quite small. The hypothesis 2a which support that female advantage in reading achievements would widen in secondary level because of early tracking thus can be corroborated.

In mathematics achievement, the coefficient is negative, represents boys outperform more than girls and the gender gap to the advantage of boys increased. Like the coefficient effect value of reading achievement, that of mathematics achievement is -0.0358, also small. Hence, the hypothesis 2b that assume early tracking system makes boys increase their advantage on mathematics achievement in secondary school can be accepted.

However, unlike reading and mathematics achievement, the key three-way interactions variable does not have the significant difference from zero in science achievement. Which suggests that gender gap in science achievement has no statistically significant difference from primary school to secondary school because of early tracking. Therefore, I can reject the hypothesis 2c that assume gender gap in science achievement would have the same pattern as that in mathematics achievement.

Considering the results per match and this results across all matches, we found that the two can be mutually confirmed to some extent. For example, science achievement is the only outcome that has both positive and negative coefficients, which probably makes the overall result insignificant.

The synthesized result of DiD regressions

As the result displayed in Figure2, I conducted DiD model on each match. However, they can only represent the result of their match. Thus, I employed basic meta-analysis to synthesize the results to make the final mean coefficient of the key variable of interest can represent all individuals that participated in large-scale surveys. The results are shown in Table5. To illustrate, the three-way interaction coefficient β_7 is the same as that in overall DiD model, indicates the gender gap differences affected by early tracking system from primary to secondary level.

Compared with the results of overall DiD regression, we can find out that the results are quite similar. The reading achievement in both results are positively and significantly different from zero. The only difference is the effect, 0.0572 for overall DiD regression and 0.5689 for synthesized result. As well as the difference in mathematics, -0.0358 for overall DiD regression and -0.0321 for synthesized result. The difference between these two results maybe because meta-analysis has better tolerance to random errors.

In summary, differences in reading and math achievement gender gaps have significance and that in science have not. However, early tracking show significant effect on gender gaps from primary to secondary level, the coefficient of math shows the exact opposite. Although.

Even if I do not consider science outcome as it has no significance, it is still very hard to estimate the difference of overall achievement gender gaps. Since reading and math, which subject weights more when measuring the overall achievement is uncertain. Besides, the coefficient that represent difference-in-differences is not the natural number that can be simply calculated. Thus, I reject the hypothesis 1 which suggests early tracking can make girls gain more advantages and widen the gender gap in achievements than non-tracking (late tracking).

Table5 *Synthesized Difference-in-Differences result per outcome*

VARIABLES	Reading achievement	Mathematics achievement	Science achievement
Three-way interaction coefficient β_7	0.0568871*** (0.0045185)	-0.0321681** (0.0044504)	0.0041627 (0.0058278)
Number of matches	7	8	7
Number of obervations	2849449	3183597	1589970

Note: Standard error in parenthesis *** p<0.01, ** p<0.05, * p<0.1

Robustness check

To make the results more convincing and reliable, Robustness check is essential. Even though most of the literature employed DiD model use placebo test to check the robustness, this research cannot use this because of the lack of reading achievement data in other grades. Meanwhile, as I also use meta-analysis in this article, the heterogeneity becomes important. Therefore, I will use winsorization to limit the extreme values in the dataset and ‘reduce the possible outliers’ (Dixon, 1960).

Table6 displays the results after a 98% winsorization, which means the Stata system automatically replaced the number less than 1% of the percentile with the value of 1% of the percentile and replaces the number greater than 99% of the percentile with the value of 99% of the percentile. I use this command on the achievement scores that estimated by plausible values, and then standardized the achievement scores again. The subsequent steps are the same as what have done. Focused on the key variable of interest, gender gap differences have significance on all outcomes except science achievement, which is the same as previous results. However, the effect value of reading achievement differs much. Although the gender gap difference in reading achievement affected by early tracking in secondary school level is still significant, the value decreased. This may indicate that there are more extreme values in the reading achievement in the previous data。

Table6 *Results of overall DiD model by outcome with a 98% winsorization*

VARIABLES	Reading achievement	Mathematics achievement	Science achievement
Female	0.132***	-0.0720***	-0.0260***
	(0.00341)	(0.00311)	(0.00278)
Secondary school	0.623*	0.650**	0.931***
	(0.366)	(0.284)	(0.0164)
Female # Secondary school	0.193***	-0.00942**	0.00959**
	(0.00394)	(0.00368)	(0.00419)
Female # Early tracking	-0.0240***	0.00631	-0.0286***
	(0.00592)	(0.00529)	(0.00471)
Secondary school # Early tracking	-0.193***	-0.00944**	-0.0930***
	(0.00500)	(0.00456)	(0.00593)
Female # Secondary school #Early tracking	0.0333***	-0.0379***	0.00530
	(0.00675)	(0.00621)	(0.00795)
Constant	-0.559**	-0.421**	-0.030***
	(0.278)	(0.205)	(0.00669)
Observations	2,849,449	3,183,597	1,589,030

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

Discussion

Discussion about gender gaps and further research

In this research, I hypothesize that early tracking can provide female more advantages overall in secondary school than late tracking and hypothesize early tracking would make female keep advantage in reading but underperform boys in math and science. The results have shown that the gender gap differences affected by early tracking system are significant in reading achievements and mathematics achievement from primary school to secondary school, whereas that of science achievement did not have obvious difference.

What I want to focus on is the gender gap changes in reading and math affected by early tracking. As mentioned in the self-fulfilling prophecy in the theoretical framework, boys and girls will receive feedback from other people whether in school or other environments during the growth process, and this feedback gradually makes boys and girls produce a solidified gender identity. This process appears obvious in school. Shu (2004) found some primary school teachers in China were affected by traditional gender views from Confucianism having a stereotype on which gender advantages on which subjects. The ideology and the traditional gender view which embedded in history and rooted in social structures, became a ‘hidden curriculum’ in everyone’s education. Boys and girls would be judged by their teachers and parents based on the common traditional standard, principled by the cultural restriction. It is not a unique phenomenon that men and women were depicted differently in China, the Japanese textbooks also presented women as the ones who should take domestic responsibilities, the corresponding female vignettes are either absent or even operate to exaggerate bias (Clark, 2016). These suggestions form part of the self-fulfilling prophecy of boys and girls. Considered of the results of this research, what we can discuss is, why early tracking widens the gender gap in reading and mathematics achievements in secondary level. The research result of Steinmann et al. (2023) can provide a possible reason. They test the difference of school gender segregation in primary level and secondary level, the result shows higher degree of school gender segregation after early tracking happened. Which means early between-school tracking is more likely to assign students with same gender to same school. Thus, we can assume that this is one of the reasons make early tracking have significant effect on achievement gender gap, since homogeneous clusters are always groups that make members feel belonging but deepen the stereotype at the same time.

Besides, tracking system can be associated with meritocracy (Oakes, 1986). The author believes “tracking is deeply rooted in assumptions about the meritocratic nature of schooling”. From this perspective, early tracking is more meritocratic than late tracking. Hence, I can speculate that the gender gap changes because early tracking force students to think about their future and face the social compete at a relatively young age, which will make students work harder in the subject they are good at potentially.

Furthermore, the results or this article inspired me about what further research can focus. First, due to early tracking reinforce female advantage in reading but not in math, the relationship between reading achievement and math achievement of girls can be tested. Breda &Napp (2019) brought up that reading scores and math scores are inversely correlated. However, this research lack of a detailed research design with a suitable method, so this topic can be investigated in further analysis. Moreover, achievements as key factors that influence the educational career, the gender gaps of them can produce some far-reaching effect. Although this research only shows the result on secondary level, some other research measures the relationship between tracking system and educational attainment (Bodovski & Munoz, 2020; Scheeren & Bol, 2022). Meanwhile, the study of female underrepresented in STEM can be associated with tracking system in further research. That is, if early tracking system can affect the female disadvantages in STEM labor market by affecting mathematic achievement gender gap.

This study shows a clear and detailed result about whether tracking will affect achievements gender gap, and the respective impact directions regarding to reading, math and science achievement. It makes the relationship between tracking system and gender gap in specific subject more explicit.

Discussion of research design

Compared with previous research on tracking system and gender gaps, there is difference in the results. Some research (Contini & Cugnata, 2020; Steinmann et al., 2023) shows insignificant gender gap changes in mathematics achievement from primary to secondary school. The possible reason of causing the difference is the difference selection of dataset and the different research design, this is also what makes this research different from some research. I mainly employed difference-in-differences model to examine the hypotheses and utilize meta-analysis to synthesize results per match as supplement. DiD model certainly has its advantages especially compared to the mostly conducted cross-sectional study, as it avoid many social context covariates that have effect on educational achievement gender gap but hard to qualified. However, this research topic, about whether early tracking increase the gender gap in achievements than late tracking from primary school to secondary school, is not a typical panel data suitable for DiD model. Most of the scholars applied DiD model on testing the effect of policy implementation, the data of which always includes a regional variable and a continuous and unique time variable (always a continuous calendar year). In this case, the data can be easily transformed to panel data. Although DiD model can be used in cross-sectional data as well, errors may occur.

The time-varying DiD that most literature employed have enough data that happened in different time slot to test parallel assumption and a placebo test. The parallel assumption is quite important in DiD analysis with assume the control group and treatment group will have the similar developing trends without the policy implementation. To test this assumption, researchers have to collect data with at least three time slots, with two time slots before and one after. However, in this research, it is hard to find the relatively earlier data than PIRLS and TIMSS (4^{th} grade), as no large-scale cross-national survey for students younger than 4^{th} grade. In term of this, it is more suitable to collect longitudinal data with employing of DiD model. However, it is hard and complicated to carry out a longitudinal survey that across multiple countries with a specific standard that apply to all countries. For now, the possible solution of this problem may be restructuring the data and bringing individual-level data to national-level data. In this research, I tried to solve this by matching the data based on birth cohort and the year that survey waves carried out. So, the data used for DiD regression is not a simply individual-level data, this may have wider applicability.

In this research, I also tried to use meta-analysis to synthesize the results of matches as an estimation to a more general result, which aim to increase the statistical accuracy and reliability. With the utilizing of meta-analysis, the generalization and validity of the research can be improved. However, a problem is I choose fixed effect model as what DiD does to do the meta-analysis, which infer the random error exist within studies rather than between studies. It makes sense because single match includes data from two surveys except science matches, more error may exist because of e.g. different questions in assessment. Nevertheless, the between-studies difference may also exist because the questions are not always the same from first wave to last wave. This may influence the validity of this study.

Finally, this research design did not include within school tracking as a variable. The within school tracking is also happened in different ages, some earlier than between-school tracking and some later. Thus, it is hard to control within school tracking in this research design and I cannot tell if the gender gap changes in secondary school partly because of within school tracking.

Practical implications

Based on the results of this research, some points need to be focused when considering the interventions or policy. First, I found not significantly gender gap difference of overall achievement by early tracking, but it does not mean early tracking is equal to girls and boys. Since early tracking makes girls gain more advantage in reading and boys gain more advantage in math, how to make girls perform better in math and boys perform better in reading is important. For example, the determination of tracking is always partly based on qualified scores and partly based on teachers' recommendation. To narrow the gender gap in reading and mathematics achievement, schools can adjust the weights of reading scores and math scores to girls and boys respectively to push them work hard on the subject that they do not skilled in.

From the institutional level, employ more female mathematics teachers is also useful. More female working as mathematics teachers is a kind of implication to female students that girls can also perform very well in mathematics.

In summary, policy makers and leaders of educational institutions must consider the gender gaps in reading and math achievement affected by tracking system and try to narrow it by interventions.

Conclusion

Although scholars have conducted various research about tracking systems and relative inequalities, the achievement gender gaps in difference subjects still unclear. According to the review of self-fulfilling prophecy and relative literature, I put forward 4 hypotheses. This research utilize different research designs based on previous literature. I combined data from three different international assessments (PIRLS, TIMSS and PISA) and matched the waves based on birth cohort of respondents and the year that surveys carried out. The sample contains 44 countries (15 countries with early tracking and 29 countries with late tracking), 22 matches and 2699532 observations. To analyze data with a more reliable representative and valid way, I mainly employed difference-in-differences model to examine the changes of achievement gender gaps from primary school to secondary school overall and per match. To synthesize the 22 regression results per match, meta-analysis with fixed-effect model was used in this research as well. Results can be seen in tables and figures.

The results show: (a) early tracking make female have more advantages in reading achievement compared to late tracking. (b) girls underperform boys in mathematics achievement and the gap widened of being affected by early tracking. (c) no significant changes in science achievement gender gap by early tracking. (d) because the weight of reading achievement and math achievement cannot be confirmed, the gender gap changes of overall achievement by early tracking is uncertain. Thus, I accept hypothesis 2a and 2b, while rejected hypothesis 1 and 2c.

This study shows a clear and detailed result about whether tracking will affect achievements gender gap, and the respective impact directions regarding to reading, math and science achievement. This study provides additional details about gender gap in different subject, which should be considered by educational policy makers.

References

Alam, A. (2022). *Psychological, Sociocultural, and Biological Elucidations for Gender Gap in STEM Education: A Call for Translation of Research into Evidence-Based Interventions*.

Batruch, A., Autin, F., Bataillard, F., & Butera, F. (2019). School Selection and the Social Class Divide: How Tracking Contributes to the Reproduction of Inequalities. *Personality and Social Psychology Bulletin*, *45*(3), 477–490. https://doi.org/10.1177/0146167218791804

Bodovski, K., & Munoz, I. G. (2020). Do education system characteristics moderate the socioeconomic, gender and immigrant gaps in math and science achievement? *International Journal of Sociology of Education*, *9*(2), 122–154. https://doi.org/10.17583/rise.2020.4807

Breda, T., & Napp, C. (2019). Girls' comparative advantage in reading can largely explain the gender gap in math-related fields. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(31), 15435–15440. https://doi.org/10.1073/pnas.1905779116

Charles, M. (2017). Venus, Mars, and Math: Gender, Societal Affluence, and Eighth Graders' Aspirations for STEM. *Socius*, *3*. https://doi.org/10.1177/2378023117697179

Contini, D., & Cugnata, F. (2020). Does early tracking affect learning inequalities? Revisiting difference-in-differences modeling strategies with international assessments. *Large-Scale Assessments in Education*, *8*(1). https://doi.org/10.1186/s40536-020-00094-x

Domina, T., McEachin, A., Hanselman, P., Agarwal, P., Hwang, N. Y., & Lewis, R. W. (2019). Beyond Tracking and Detracking: The Dimensions of Organizational Differentiation in Schools. *Sociology of Education*, *92*(3), 293–322. https://doi.org/10.1177/0038040719851879

Fisher, D. J., Zwahlen, M., Egger, M., & Higgins, J. P. T. (2022). Meta-Analysis in Stata. In *Systematic Reviews in Health Research: Meta-Analysis in Context: Third Edition* (pp. 483–509). wiley. https://doi.org/10.1002/9781119099369.ch25

Hanushek, E. A., & Oßmann, L. W. (2006). *Does Educational Tracking Affect Performance And Inequality? Differences-in-Differences Evidence Across Countries*.

Hanushek, E. A., & Wößmann, L. (2006). Does educational tracking affect performance and inequality? differences-in-differences evidence across countries. *Economic Journal*, *116*(510). https://doi.org/10.1111/j.1468-0297.2006.01076.x

Herd, P., Freese, J., Sicinski, K., Domingue, B. W., Mullan Harris, K., Wei, C., & Hauser, R. M. (2019). Genes, Gender Inequality, and Educational Attainment. *American Sociological Review*, *84*(6), 1069–1098. https://doi.org/10.1177/0003122419886550

Mann, A., Legewie, J., & DiPrete, T. A. (2015). The role of school performance in narrowing gender gaps in the formation of STEM aspirations: A cross-national study. *Frontiers in Psychology*, *6*(FEB). https://doi.org/10.3389/fpsyg.2015.00171

Moller, S., & Stearns, E. (2012). Tracking Success: High School Curricula and Labor Market Outcomes by Race and Gender. *Urban Education*, *47*(6), 1025–1054. https://doi.org/10.1177/0042085912454440

Nürnberger, M., Nerb, J., Schmitz, F., Keller, J., & Sütterlin, S. (2016). Implicit Gender Stereotypes and Essentialist Beliefs Predict Preservice Teachers Tracking Recommendations. *Journal of Experimental Education*, *84*(1), 152–174. https://doi.org/10.1080/00220973.2015.1027807

Pekkarinen, T. (2008). Gender differences in educational attainment: Evidence on the role of tracking from a finnish quasi-experiment. *Scandinavian Journal of Economics*, *110*(4), 807–825. https://doi.org/10.1111/j.1467-9442.2008.00562.x

Reinhold, S., Holzberger, D., & Seidel, T. (2018). Encouraging a career in science: a research review of secondary schools' effects on students' STEM orientation. *Studies in Science Education*, *54*(1), 69–103. https://doi.org/10.1080/03057267.2018.1442900

Ruhose, J., & Schwerdt, G. (2016). Does early educational tracking increase migrant-native achievement gaps? Differences-in-differences evidence across countries. *Economics of Education Review*, *52*, 134–154. https://doi.org/10.1016/j.econedurev.2016.02.004

Scheeren, L., & Bol, T. (2022). Gender inequality in educational performance over the school career: The role of tracking. *Research in Social Stratification and Mobility*, *77*. https://doi.org/10.1016/j.rssm.2021.100661

Scheeren, L., van de Werfhorst, H. G., & Bol, T. (2018). The gender revolution in context: How later tracking in education benefits girls. *Social Forces*, *97*(1), 193–220. https://doi.org/10.1093/sf/soy025

Schindler, S., Bar-Haim, E., Barone, C., Fels Birkelund, J., Boliver, V., Capsada-Munsech, Q., Erola, J., Facchini, M., Feniger, Y., Heiskala, L., Herbaut, E., Ichou, M., Karlson, K. B., Kleinert, C., Reimer, D., Traini, C., Triventi, M., & Vallet, L.-A. (2023). Educational tracking and social inequalities in long-term labor market outcomes: Six countries in comparison. *International Journal of Comparative Sociology*, 002071522311513. https://doi.org/10.1177/00207152231151390

Steinmann, I., Strello, A., & Strietholt, R. (2023). The effects of early between-school tracking on gender segregation and gender gaps in achievement: a differences-in-differences study.

School Effectiveness and School Improvement. https://doi.org/10.1080/09243453.2023.2165510

Strello, A., Strietholt, R., Steinmann, I., & Siepmann, C. (2021). Early tracking and different types of inequalities in achievement: difference-in-differences evidence from 20 years of large-scale assessments. *Educational Assessment, Evaluation and Accountability*, *33*(1), 139–167. https://doi.org/10.1007/s11092-020-09346-4

Terrin, É., & Triventi, M. (2022). The Effect of School Tracking on Student Achievement and Inequality: A Meta-Analysis. *Review of Educational Research*. https://doi.org/10.3102/00346543221100850

van de Werfhorst, H. G. (2018). Early tracking and socioeconomic inequality in academic achievement: Studying reforms in nine countries. *Research in Social Stratification and Mobility*, *58*, 22–32. https://doi.org/10.1016/j.rssm.2018.09.002

Van Houtte, M. (2004). Tracking effects on school achievement: A quantitative explanation in terms of the academic culture of school staff. *American Journal of Education*, *110*(4), 354–388. https://doi.org/10.1086/422790

Acknowledgements

Thanks to my supervisor Professor Sjaak Braster with the suggestions in finishing this thesis. I used the data from PIRLS TIMSS and PISA, thanks to all the volunteers and the staffs who working for the international surveys' institution.

Also thanks to my parents, they comforted me a lot when I was having trouble writing my dissertation and feeling depressed.

ERASMUS UNIVERSITEIT ROTTERDAM

CHECKLIST ETHICAL AND PRIVACY ASPECTS OF RESEARCH

INSTRUCTION

This checklist should be completed for every research study that is conducted at the Department of Public Administration and Sociology (DPAS). This checklist should be completed *before* commencing with data collection or approaching participants. Students can complete this checklist with help of their supervisor.

This checklist is a mandatory part of the empirical master's thesis and has to be uploaded along with the research proposal.

The guideline for ethical aspects of research of the Dutch Sociological Association (NSV) can be found on their website (http://www.nsv-sociologie.nl/?page_id=17). If you have doubts about ethical or privacy aspects of your research study, discuss and resolve the matter with your EUR supervisor. If needed and if advised to do so by your supervisor, you can also consult Dr. Bonnie French, coordinator of the Sociology Master's Thesis program.

PART I: GENERAL INFORMATION

Project title: **Early between-school tracking and gender gaps in achievement**

Name, email of student:Junying Zhao 612803jz@eur.nl

Name, email of supervisor: Sjaak Braster sjaak.braster@gmail.com

Start date and duration: 20/12/2022 6 month

Is the research study conducted within DPAS **YES** - NO

If 'NO': at or for what institute or organization will the study be conducted?
(e.g. internship organization)

PART II: HUMAN SUBJECTS

1. Does your research involve human participants. YES - **NO**

If 'NO': skip to part V.

If 'YES': does the study involve medical or physical research? YES - NO

Research that falls under the Medical Research Involving Human Subjects Act (WMO) must first be submitted to an accredited medical research ethics committee or the Central Committee on Research Involving Human Subjects (CCMO).

2. Does your research involve field observations without manipulations that will not involve identification of participants. YES - NO

If 'YES': skip to part IV.

3. Research involving completely anonymous data files (secondary data that has been anonymized by someone else). YES - NO

If 'YES': skip to part IV.

PART III: PARTICIPANTS

1. Will information about the nature of the study and about what participants can expect during the study be withheld from them? YES - NO
2. Will any of the participants not be asked for verbal or written 'informed consent,' whereby they agree to participate in the study? YES - NO

3. Will information about the possibility to discontinue the participation at any time be withheld from participants? YES - NO

4. Will the study involve actively deceiving the participants? YES - NO
Note: almost all research studies involve some kind of deception of participants. Try to think about what types of deception are ethical or non-ethical (e.g. purpose of the study is not told, coercion is exerted on participants, giving participants the feeling that they harm other people by making certain decisions, etc.).

5. Does the study involve the risk of causing psychological stress or negative emotions beyond those normally encountered by participants? ` YES - NO

6. Will information be collected about special categories of data, as defined by the GDPR (e.g. racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data for the purpose of uniquely identifying a person, data concerning mental or physical health, data concerning a person's sex life or sexual orientation)?

   YES - NO

7. Will the study involve the participation of minors (<18 years old) or other groups that cannot give consent? YES - NO

8. Is the health and/or safety of participants at risk during the study? YES - NO

9. Can participants be identified by the study results or can the confidentiality of the participants' identity not be ensured? YES - NO

10. Are there any other possible ethical issues with regard to this study? YES - NO

If you have answered 'YES' to any of the previous questions, please indicate below why this issue is unavoidable in this study.

____

What safeguards are taken to relieve possible adverse consequences of these issues (e.g., informing participants about the study afterwards, extra safety regulations, etc.).

____

Are there any unintended circumstances in the study that can cause harm or have negative (emotional) consequences to the participants? Indicate what possible circumstances this could be.

____

Please attach your informed consent form in Appendix I, if applicable.

Continue to part IV.

PART IV: SAMPLE

Where will you collect or obtain your data?

____

Note: indicate for separate data sources.

What is the (anticipated) size of your sample?

____

Note: indicate for separate data sources.

What is the size of the population from which you will sample?

____

Note: indicate for separate data sources.

Continue to part V.

Part V: Data storage and backup

Where and when will you store your data in the short term, after acquisition?

_I'll get my data from the online database, specifically PISA, PIRLS and TMISS. Then store them in my laptop.

Note: indicate for separate data sources, for instance for paper-and pencil test data, and for digital data files.

Who is responsible for the immediate day-to-day management, storage and backup of the data arising from your research?

_Myself.____

How (frequently) will you back-up your research data for short-term data security?

Once a month.

In case of collecting personal data how will you anonymise the data?

_As far as I planed, there is no personal data in this research.

Note: It is advisable to keep directly identifying personal details separated from the rest of the data. Personal details are then replaced by a key/ code. Only the code is part of the database with data and the list of respondents/research subjects is kept separate.

PART VI: SIGNATURE

Please note that it is your responsibility to follow the ethical guidelines in the conduct of your study. This includes providing information to participants about the study and ensuring confidentiality in storage and use of personal data. Treat participants respectfully, be on time at appointments, call participants when they have signed up for your study and fulfil promises made to participants.

Furthermore, it is your responsibility that data are authentic, of high quality and properly stored. The principle is always that the supervisor (or strictly speaking the Erasmus University Rotterdam) remains owner of the data, and that the student should therefore hand over all data to the supervisor.

Hereby I declare that the study will be conducted in accordance with the ethical guidelines of the Department of Public Administration and Sociology at Erasmus University Rotterdam. I have answered the questions truthfully.

Name student:  Name (EUR) supervisor: Sjaak Braster

Date: 20/03/2023 Date: 20/3/2023