Utrecht University

School of Economics

Master Thesis U.S.E

Economic Policy

Academic year: 2023/2024

Name: Lucas Agterberg

Mail: l.agterberg@students.uu.nl

Student number: 2773509

1st Supervisor: Sönke Matthewes

2nd Supervisor: Kattia Moreno

Wordcount: 10,859

JEL: I21, I24, I28,

Keywords: secondary education, ability tracking, inequality, fixed effects

The relationship between mixed classes, inequality and average educational performance in Dutch secondary education.[1]

Abstract

In this thesis the relationship between mixed classes, inequality and average educational performance in Dutch secondary education is examined. Using cross-sectional, cohort fixed effects and TWFE models a careful empirical examination of these relationships is constructed. The findings suggest that the net effect of mixed classes on average educational performance is positive. This suggests that more comprehensive schooling by the use of mixed classes is in the interest of Dutch policymakers focused on the net effect of tracking, given that similar effects are also found in the literature. The expected relationship between inequality and mixed classes is not found in these models. This is due to the reason that the heterogeneous effects of ability tracking as suggested in existing literature were not found in this study. This does not mean that such relationship does not exist. Although the use of a TWFE model significantly restricts the risk of endogeneity, there is still endogeneity possible. This is also indicated by the robustness tests done. The study adds therefore to the evaluation of mixed classes and average educational performance but leaves room for further research on possible heterogenous effects of mixed classes related to inequality in Dutch secondary education.

Table of Contents

1. Introduction

For the past couple of years, average student performance has been decreasing across OECD countries. Between 2014 and 2018 average student performance on mathematics, science and reading had already been decreasing. However, the scores from 2018 onwards decreased even more. For example, the average decrease of mathematics skills was four times higher than ever recorded before. Currently, a fourth of the students in OECD countries is a low performer in mathematics, reading and science (OECD, 2023).

The decrease of average educational performance among OECD countries is concerning. According to the European Commission, good quality education increases the match between workers and the labor market. When a mismatch occurs between individual's education and the labor market this decreases one's employment opportunities, but also contributes to weakening economic growth and social cohesion (European Commission, n.d.).

Whereas policymakers and society should therefore be looking for improvements in educational systems, evaluating what aspects of educational systems are efficient in terms of equity and performance is often difficult. This is due to the reason that educational systems can be seen as a complex web of different aspects that are often interrelated (Blossfeld et al., 2016). Whereas several of these aspects have already been studied into detail and obtained a consensus, it remains unclear whether ability tracking can be considered to contribute to the efficiency of educational systems.

Ability tracking refers to a system in which students are sorted based on their educational ability. Tracking is applied in various ways at various levels. Different applications range from grouping students into different classes based on ability, to offering different curricula or instructional methods. (Betts, 2011).

Whether ability tracking is effective is still an ongoing debate, although it has been discussed frequently in existing literature. Nevertheless, it remains part of the interests of politics and society as explained by the previous developments. Advocates argue that ability tracking increases efficiency, offering schools the opportunity to meet the needs of all students from different levels.

For example, students heading towards vocational education can be offered skills that they need in their future, while students heading towards university can be prepared for academics. Others argue that there is also a significant loss because ability tracking would increase already existing inequality. (Betts, 2011).

Furthermore, existing literature seems to suggest that the effects of ability tracking are possibly heterogenous among individuals e.g. among students from different socioeconomic backgrounds. However, the discussion of such potential heterogenous effects is rather limited and generally focused on specific socioeconomic characteristics such as gender or family background (Scheeren & Bol; Brunello & Checchi 2007).

This study will exploit the within country variation of between-school tracking resulting from the use of mixed classes in Dutch secondary education. Around 60% of the Dutch secondary schools offer so-called 'brugklassen' in which students of two tracks attend classes together. This contrasts with their peers attending homogenous classes. These peers are forwarded into one track. Tracking is thus partially delayed for one or two years for those attending these mixed classes. For others, tracking is not delayed. The differences between these students therefore offer the opportunity to compare different tracking applications within a between-school tracking system. This is of particular interest, given that previous literature has been mostly focusing on within-school tracking.

The Dutch system of mixed classes has been studied already by Borghans et al. (2012) in a cross-sectional model. Although Borghans et al. (2012) argue that their instrument is exogenous, I argue that this cannot be said with complete certainty. By using several cross-sectional models, a cohort fixed effects and a two-way fixed effects model, I extent the research done by Borghans et al. (2012) on the effect of mixed classes on average educational performance keeping in mind the advantages and disadvantages of these models.

By doing so, I will isolate the net effect of mixed classes on average educational performance. Furthermore, this study will also examine possible heterogenous effects of mixed classes by using socioeconomic variation among schools and municipalities. As a result, it will add to the understanding of the net effects of ability tracking on average educational performance and the discussion of underlying mechanisms behind potential heterogenous effects of ability tracking. By doing so, I serve both perspectives on the debate around ability tracking. First, those policymakers, academics and individuals interested in the net effect of tracking. Secondly, those interested in the effects of ability tracking on individuals with different socioeconomic backgrounds.

I find that the effects of offering mixed classes in cross-sectional models on both municipality and school level are insignificant in the cross-sectional and cohort fixed effects models, but become significant after adding two-way fixed effects. I suggest that the effect of mixed classes on average educational performance is strongly positive. This is in line with et al. (2012) found that early tracking (homogeneous classes) negatively impacts average educational performance.

Furthermore, I use the same methodology to isolate possible heterogenous effects of tracking by using SES-scores on both municipal and school level. In contradiction to the existing literature suggesting heterogenous effects of tracking, I conclude that the effects of mixed classes in the Dutch educational system are not differential across different SES groups. It can be questioned to what extent these findings are valid, given that these were not found to be robust.

The paper is structured as follows: In section 2 in-depth information about the concept of ability tracking and the Dutch educational system are explained next to a review of the existing literature on ability tracking, average educational performance and inequality. This eventually leads to my research question and hypothesis. In section 3 the data sources, samples and strategy are discussed. In section 4 I discuss the results. In section 5 I apply a robustness check. Afterwards, the discussion and conclusions are added in section 6.

2. Literature review and theoretical framework.

2.1 Ability tracking and educational systems.

Nowadays, education is viewed as one of the drivers behind economic development. It is therefore not surprising that education is one the permanent topics of debate in society and notably marked in the agenda of policymakers, often aiming for a system that is efficient in terms of attainment and performance. Besides that, policymakers often demand that the system is as equal as possible whereas differences in performance and attainment should be due to differences in ability rather than be related to socioeconomic factors (Blossfeld et al., 2016).

Analyzing educational systems is not straightforward. Educational systems can be seen as a complex web of different aspects that are often interrelated and eventually effect educational attainment and performance all together. Blossfeld et al. (2016) explain these various single aspects using three umbrella terms. The first one is input which describes how much time and financial support educational institutions obtain and require. The second one is organization which describes how students are instructed and evaluated e.g. by which type of exam. The final one is tracking which explains how much segregation based on ability is applied to the system. All together these umbrella terms are a general description of the building blocks of educational systems, affecting educational outcomes.

Before ability tracking can be studied empirically in this study, a clear explanation of ability tracking is necessary. Generally, ability tracking refers to a system in which students are sorted based on their believed educational ability. Tracking is applied in various ways at various levels and as a result, the concept of ability tracking itself is quiet flexible. Different applications range from grouping students into different classes, to offering different curricula or instructional methods. Although countries differ in the way if and how they implement tracking into their educational systems, three types of countries can be identified. First, there are countries that forward students into separate tracks after primary education of which the Netherlands is an example. This is also known as between-school tracking. Furthermore, there are countries that only offer tracking within schools, the national system is therefore comprehensive but in schools tracking is applied. A well-known example of this is the United States where many schools design classes based on ability following different course levels. A third of countries do not apply tracking which offer a complete comprehensive system. Sweden is a well-known example of this (Betts, 2011; OECD & PISA 2020).

Proponents of ability tracking argue that ability tracking increases efficiency. Due to ability tracking, a school could meet the needs of all students from different levels. For example, students heading towards vocational education can be offered skills that they need in their future, while students heading towards university can be prepared for academics. Others argue that there is also a significant loss because ability tracking would increase already existing inequality. Ability tracking would lock students from lower socioeconomic backgrounds into lower education and as a result into lower paid jobs. Furthermore, the effects of tracking may differ among individuals from different socioeconomic backgrounds (Betts, 2011).

2.2 Dutch educational system in a global context.

As explained previously, countries have different educational systems with or without different types of tracking. On average, OECD countries track students around the age of 15. The Netherlands tracks relatively early but also offers more tracks than the average OECD countries. Around the age of 12 / grade 6 at primary school, all students take part in a test which tests the performance of students and ranks them accordingly to the tracks that are offered in secondary school. This test outcome together with the recommendation of the teacher decides in which track a student starts in secondary school (Rijksoverheid, n.d.-a, OECD & PISA, 2020).

Students can be placed into three different tracks that prepare student for their post-secondary education. Students ranked with lower abilities are placed into VMBO which prepares students for secondary vocational education during a period of four years. Student ranked as intermediate are placed into HAVO which prepares students for higher education during a period of five years. Finally, high ability students are placed into VWO which prepares students for university during a period of six years. Within VMBO there are three separate tracks offered based on performance as well, however all students of these tracks offer students the possibility to enter secondary vocational education. For the remainder of this thesis therefore, these are all labeled as VMBO following (Borghans et al, 2012). Besides following mandatory core courses such as English and

Dutch, students in all tracks select a profile halfway their secondary that offers courses based on the preferred career path. For HAVO and VWO these profiles are similar although the curricula match their track-level. For example, HAVO & VWO students aiming for a career in beta sciences can pick the profile Nature and Technology. The same concept is applied to VMBO students, although their profiles are made to match their future secondary vocational education (Rijksoverheid, n.d.-b).

Another noticeable characteristic of the Dutch educational system is that it also allows students to extent tracking with one or two years. Throughout the Netherlands, secondary schools also offer so-called ‘brugklassen’ which are mixed classes of two levels instead of homogenous classes. These are available for all students among all levels. For example, students selected for either HAVO or VWO could apply for a HAVO / VWO class in which they are tracked one or two years later depending on the schools’ policy. In this case, a part of the tracking occurs later in the career of these students, compared to their peers in homogenous classes. Schools are free to decide whether they offer such mixed classes. Around 60% of the Dutch secondary schools offers such mixed classes of which half of these has a length of one year and another half has the length of two years. The remaining 40% of secondary schools are only offering homogenous classes. The mixed classes have a length of either one or two years. Altogether, around 54% of the first-year secondary pupils attends a mixed class (Bles & Van der Velden, 2024).

A visual representation of this educational system is shown in figure 1. The arrows indicate that while attending a mixed class individuals can move to two tracks after one or two years versus peers attending a homogenous class. In the figure it is assumed that the duration of the mixed classes that individuals attend is two years. This is not shown, but in that case forwarding to a homogenous class occurs after grade 9.

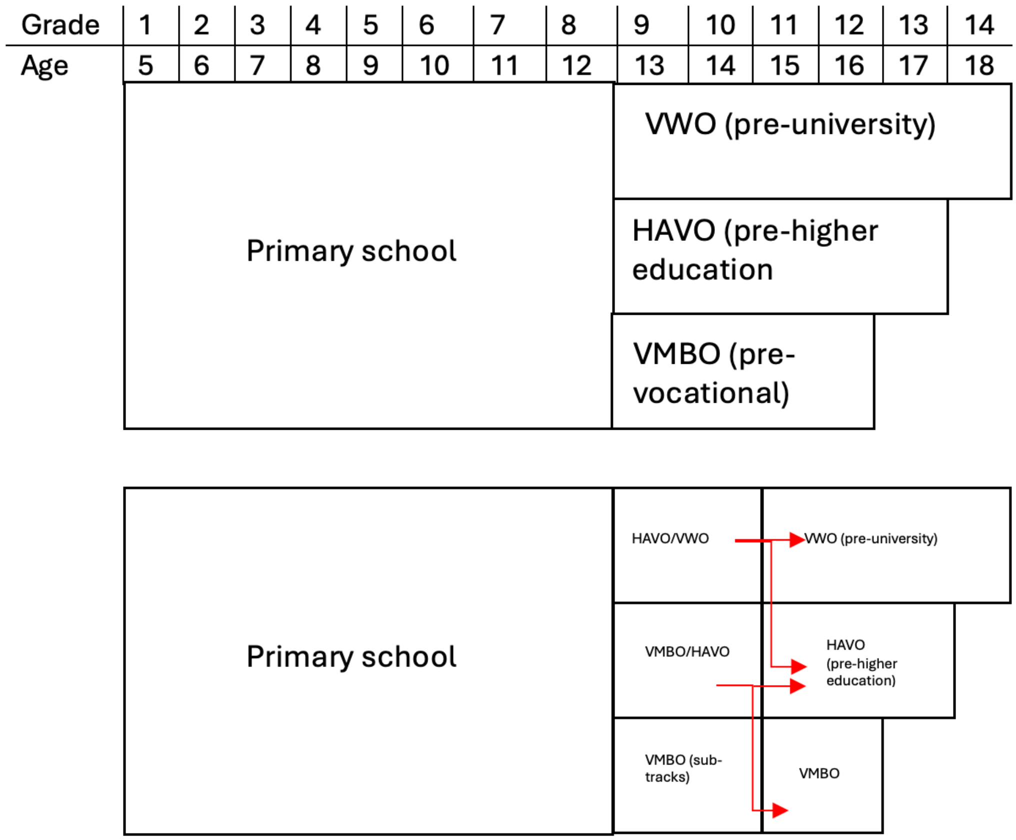

Figure 1: Possible trajectories in Dutch secondary education for individuals starting in a homogenous class (above) and those starting in a two-year mixed class.

2.3 The effect of ability tracking on average educational performance.

Different streams of literature have tried to contribute to the evaluation of ability tracking in terms of efficiency already. In the following section these are analyzed.

2.3.1 Comparative cross-country studies.

The first type of studies looking at the effect of ability tracking on educational outcomes mainly focus on comparisons between countries. Ariga & Brunello (2007) studied the effect of tracking on educational outcomes using data on multiple countries with different educational (tracking) systems. The findings from the IV estimation suggested that tracking increased individual's educational attainment. However, the authors constructed the LATE which may differ considerably from the ATE given that affected population is small. As a result, it remains unclear if these results are meaningful and into what extent.

Doubts about the validity of the found effects of tracking on educational performance are also present in other studies. Moreover, results are often contradictive and mixed. For example, Horn (2009) found that earlier tracking did not increase the average literacy of students. This is rejected by Rindermann (2007) who finds that early tracking increase average student performance. Lavrijsen & Nicaise (2014) question these findings explaining that the design used by Rindermann (2007) failed to control for heterogenous trends among countries.

It may not be surprising that these comparative studies find mixed results as it is very hard to estimate an empirical model controlling for all (un)observed factors that influence educational performance. On a country-level, countries have designed very different educational systems of which specific characteristics, such as tracking which are often non-random. Given that these countries differ greatly, a comparison among non-tracking and tracking countries will not yield a valid treatment effect but will be prone to heterogeneity and endogeneity. Besides that, countries also differ in which tracking system is applied. For example, the US applies a within-school tracking system where tracking systems differ among schools, while the Netherlands applies a between-school tracking system where tracking is applied on a national level. Overall, the findings of cross-country studies are therefore hard to generalize into a theory of average educational performance and ability tracking (Lassibille & Gómez, 2000; Betts, 2011).

Difference-in-difference designs

The use of difference-in-difference designs is an attempt to solve the endogeneity and heterogeneity issues part of comparative studies. As explained by Lavrijsen & Nicaise (2016) the design solves for differences in secondary education between countries assuming that these differences already existed in primary school and are constant over time. As a result, additional differences in outcomes between countries can be assigned to tracking which is assumed to be the only variable changing after primary school.

Hanushek & Wößmann (2006) and Lavrijsen & Nicaise (2016) both used this design. Both studies found that the net effect of ability tracking on average educational performance is negative, questioning the gains in effectiveness as argued by advocates of tracking.

However, several notes can be made regarding the findings of these studies. First, the magnitude of the effect is found to be mixed and dependent on the specification. Second, as explained above both designs rely on the parallel-trends assumption. If this assumption is met is questionable. There are numerous cofounders that might affect outcomes and possibly change after primary school. For example, countries use a different number of tracks and have different educational systems. Furthermore, the effect of family background may be reinforced after tracking. Finally, the rigidity of tracking also differs among countries. Even though the difference-in-difference designs partially solve for heterogenous trends, they fail to do so completely. As a result, the true effect of ability tracking on educational performance remains unclear (Hanushek & Wößmann 2006; Lavrijsen & Nicaise 2016, OECD & PISA 2020).

2.3.2. Single country studies

A possible way to work around differences between countries is by focusing on single countries and comparing students or schools from the same country, which are all exposed to the same educational system. This is done in multiple ways, as explained below.

Within-school tracking

Several studies have looked at the effects within-school tracking on average educational performance. For example, Betts and Shkolnik (2000) studied the effect of within-school stratification in American secondary schools on students' achievement and found that the effects on achievement are neutral. These findings are supported by similar studies also finding that the effects of within-school tracking on achievement are small or neutral (Fu & Mehta, 2018; Figlio & Page 2000).

Between-school tracking

A small number of studies looked at the effects of between-school tracking on average educational performance, suggesting that this type of tracking negatively impacts average student achievement.

Between-school tracking differs from within-school tracking given that students are tracked to different school levels based on ability. This is different compared to within-school tracking in which students are sorted within schools e.g. sorted into different into different classes but following the same curriculum.

Matthewes (2021) studied the effect of between-school ability tracking in Germany using a difference-in-differences framework. It was found that being part of comprehensive education for two more years increased student achievement.

Closely related to this thesis is the study done by Borghans et al. (2012) who studied the effect of early tracking on educational attainment and labor market outcomes. To do so, they used the supply of early tracking schools in municipalities as an instrument for early tracking. As explained before, the Dutch educational system offers mixed classes which partially extent the moment of tracking with either one or two years. Borghans et al. (2012) found that early tracking (hence attending a homogenous class) negatively impacts both labor market outcomes and educational attainment, therefore suggesting that mixed classes have a positive effect on average educational performance.

Borghans et al. (2012) make the strong assumption that the relative attendance of mixed classes on municipal-level is not influenced by demand, although they do acknowledge that on individual-level attending a mixed class is driven by endogenous factors. Arguing that the relative attendance of mixed classes on municipal level will be exogenous is therefore a very strong assumption which cannot be met with complete certainty. If mixed classes attendance is non-random and related to individual characteristics, the same argument can be made on municipal level to students' relative attendance of mixed classes. Let's sketch a possible but hypothetical scenario in which parental income is related to educational performance of a student, which is not part of the controls used by Borghans et al. (2012). In this case, this omitted variable is therefore part of the error term. Furthermore, it is believed that rich parents often send their students to mixed schools. In this case the instrument relative attendance is related to parental income which is part of the error term. Hence, this instrument is in this case endogenous. Although this is a hypothetical scenario of which existence cannot be claimed, it can also not be claimed with certainty that such scenarios are impossible. This leaves room for further research studying the Dutch mixed classes system, comparing new results with those of Borghans et al (2012).

National policy reforms

Another stream of literature focuses on policy reforms that changed the strictness of tracking. For example, Guyon et al. (2012) used a reform in Northern Ireland which allowed lower performing students to enter pre-academic, which would have entered vocational tracks before. Although the academic track and non-academic track follow the same national-examinations and curriculum, students are tracked to different schools offering different learning environments and hence part of a within-school tracking system. Jacobs et al. (2021) used a similar type of reform which by accident allowed lower performing students to enter higher secondary tracks in the Netherlands which are part of between-school tracking. Whereas Guyon et al. (2012) found that the reform increased overall educational attainment, Jacobs et al. (2021) found that there was no effect due to correcting mechanisms in the Dutch education system caused by mixed classes. Due to mixed classes lowering the entrance bar of higher tracks was found to be neutral, given that most students attend mixed classes. Low ability students that are placed too high can therefore still move down tracks, and of course this mechanism works the other way around too. Therefore, it seems that once again the effects of tracking on average educational performance are dependent on the educational system they are applied to.

However, these single country studies still fail to work around endogeneity and heterogeneity. Even though the use of national studies corrects for heterogeneity between countries e.g. differences in tracking application, the problem remains within schools. While studying the effect of tracking on performance, one needs to be sure that this found effect is solely due to tracking and not driven by other mechanisms. Unfortunately, this is still a concern given that existing literature shows that e.g. resources and teacher/student ratios differ between tracking and non-tracking schools (Betts 2011).

Whereas a lot of research has been done on the net effect of ability tracking and average educational performance, this effect seems to be dependent on the specific educational system it is applied to. Schools and countries use different tracking applications, which does not facilitate comparisons between tracked and non-tracked systems. In fact, the found effects in existing literature may be driven by heterogenous trends and endogeneity. This is resulting from the exclusion of (un)observable covariates influencing average educational performance in models trying to explain the effect of tracking on average educational performance. An example of this could be differences in resources between treated and untreated on school or national level. This still leaves room for further research in educational systems that have not been studied into much detail yet, of which the Netherlands is an example.

2.4 Heterogenous effects of tracking for different socioeconomic groups

A smaller stream of literature explains the potential heterogenous effects of tracking. These studies aim to explain that the effects of ability tracking on average educational performance are different among groups, as introduced by Betts (2011). Whereas most of the net effects of tracking are found to be neutral or slightly negative as discussed before, it is argued that low ability students are the one significantly harmed in tracking systems. Whereas high ability students profit from being forwarded into higher tracks, lower ability students may obtain a great loss. The educational performance of low ability students is believed to decrease compared to a comprehensive system, while the performance of high ability students is improved. This is supported by the idea that students are heavily dependent on their peer environment. Learning with better peers is believed to improve individual performance. Besides that, schooling resources are also different among the two groups, in favor of the performance of high ability students. Summarizing, the effect of ability tracking on average educational performance may differ among individuals. These heterogenous effects are not limited to groups with different ability but are also considered to be occurring among groups from different socioeconomic backgrounds. For example, individuals with lower socioeconomic statuses tend to be classified into lower tracks while individuals with higher socioeconomic statuses tend to be classified into higher tracks. (Betts & Shkolnik, 2000).

Heterogeneous effects of ability tracking among different socioeconomic groups have been studied a couple times. Although the number of studies is moderate, their findings are persuasive enough to suggest that the effects of ability tracking on educational performance could be heterogeneous. Scheeren & Bol (2022) studied the effect of the ability tracking age on differences in gender educational achievement and attainment. Their results indicate that depending on the educational system of a country, there are differences in the effects of the tracking age on attainment and performance among genders. Brunello & Checchi (2007) who found that the effect of tracking on average educational performance differs between individuals with different family backgrounds.

Van De Werfhorst (2019) extends this research using policy reforms among 21 European countries between 1925 and 1989. By using parents educational background and occupation, the author found that a later tracking age tended to reduce inequality in educational performance among groups with different parental backgrounds. Burger (2016) adds to these findings concluding that the earlier tracking occurs, the more intergenerational transmission of education occurs. While both studies come to similar conclusions, they face the same econometric issues as the country comparisons that are explained in previous sections.

A limited number of studies used policy reforms to find similar effects. These studies used reforms in which tracking was extended or replaced by a comprehensive schooling system in Finland and Sweden. Again, it was found that the change to a more comprehensive schooling system increased educational attainment of students from lower socioeconomic groups, while it decreased the earnings and attainment of students from higher socioeconomic groups (Pekkarinen et al., 2009; Meghir and Palme 2005). Heterogenous effects of tracking are also found by Matthewes (2020) concluding that the net effect of comprehensive secondary schooling Germany is positive, but more importantly beneficial to low ability students and neutral to high ability students.

The existing literature seems to suggest that the effects of ability tracking are possibly heterogenous among individuals and countries with different characteristics. However, the discussion of such potential heterogenous effects is rather limited and mostly focused on specific socioeconomic characteristics e.g. gender. Proponents of ability tracking concentrate on the net effect of ability tracking. However, that may not show the bigger picture especially because existing literature has shown that the effects of tracking may differ for and among individuals. From that perspective, it may for example be the case that tracking severely harms lower socioeconomic students, while the benefits for other individuals are small. It can then be questioned if ability tracking is really as efficient as believed by advocates or that tracking slightly benefits some while significantly hurting others. By also studying possible heterogenous effects, this study aims to provide academics and policymakers a detailed breakdown of the effects of tracking on educational performance among different socioeconomic groups.

2.5 Research question and hypothesis.

The research question to be answered in this thesis is:

What are the net effects of mixed classes in Dutch secondary education on average educational performance and to what extent are these effects heterogenous for low and high SES groups?

This research question will be answered by the following sub-questions:

1. *What is the overall effect of mixed classes on average educational performance in Dutch secondary education?*
2. *Do the effects of mixed classes on average educational performance in Dutch secondary education differ between low and high SES schools and municipalities? If so, to what extent?*

Existing literature has shown that the effects of ability tracking on educational performance are dependent on the educational system. To draw a clear hypothesis therefore, I focus on studies related to the Dutch educational system. The results from existing literature in the Dutch educational system are very limited. However, they do not suggest that the overall effect of mixed classes on average educational performance are strongly negative. This is in line with research done in other educational systems, although most of the literature is hardly comparable to the Dutch context. Therefore, the effect of mixed classes in the Dutch educational system on performance is expected to be neutral or slightly positive (Borghans et al., 2012; Jacobs et al. 2021).

As explained previously, existing literature indicates that the effects of ability tracking on performance are heterogenous among different specific SES groups. Whereas the effects for high SES groups are often neutral or positive, the effects for lower SES groups (e.g. females and poor family background) are often negative or neutral (Scheeren & Bol, 2022; Brunello & Checchi, 2007). In line with existing literature, therefore, possible heterogenous effects of mixed classes are expected to be positive or neutral among lower SES municipalities and schools. For higher SES municipalities and schools, the effects are expected to be neutral or negative

3. Empirical strategy.

3.1 Data collection and description.

3.1.1 School level dataset

For the school-level dataset I used multiple datasets from two data sources. From DUO I obtained the number of graduates, teacher/student ratio, number of students per profile, and if a school offers mixed classes. All variables are on school level. To this dataset the disadvantage score for each school from CBS is added. The data used is from year 2021. Although most of the data is available for multiple years, the mixed class supply of schools is only available for 2021 (DUO, 2022b; DUO, 2022c DUO, 2022d).

The resulting dataset is split by tracks. This is done due the reason that the curricula of the tracks differ, making the average grade at schools of different tracks not comparable. This results in three single datasets for VMBO, HAVO and VWO. This results in a dataset with 729 observations for VMBO, 447 for HAVO and 454 for VWO.

The dependent variable is the average exam grade of students from a school in year 2021, which in this sample represents the measure of average educational performance. This average exam grade is calculated by adding up the average grades of all students per profile plus per school and dividing that by the total number of students that attended the graduation exams.

The treatment variable is a mixed classes dummy which equals one whenever a school offers mixed classes and homogenous classes and if a school offers only mixed classes. When a school offers only homogenous classes, this dummy is zero.

Furthermore, the dataset contains multiple control variables of which it is expected that they might affect the dependent variable average exam grade. First, the teacher/student ratio is added, which is obtained from DUO. Secondly, the share of students per school that applied for each profile is added. As explained previously, students can pick a profile halfway their secondary education which next to the mandatory curriculum is given to prepare students for the post-secondary career and in which students also must do centralized exams. This means that students also have different exams besides common core subjects such as Dutch and English. E.g. students following a beta profile do an exam in physics, while students with economics profiles take an economics exam. Beta profiles are generally considered more difficult but might also be more often chosen by good performing students due to the general opinion that these difficult beta profiles are made for the smarter kids. To account for these possible differences in average ability and exams affecting possibly affecting average educational performance, I add the ratio of students per school that select each profile. This is calculated by the number of students following a profile at a school, divided by the total number of exam students at a school (Fleischeuer, 2022).

Finally, I added the disadvantage score for each school in 2021 to these datasets. This score is calculated based on the migration background, parental education and financial background of scholars. After the preliminary calculation, the scores are weighted. The score is used by the Dutch Ministry of Education, Culture and Science to identify which schools need additional funding to solve pre-educational inequality. A higher score indicates a bigger average educational disadvantage based on the socioeconomic backgrounds of students on this school. A higher score therefore relates to a lower SES-score. This score is standardized after addition to the datasets of tracks to ease interpretation. (Posthumus et al., 2021; CBS 2023a).

In tables 1-3 the summary statistics of these variables are shown. It can be seen that a majority of the schools offer mixed classes across all tracks. For HAVO this is the highest with 83%, 77% for VWO and 72% for VMBO. This shows that the treatment group of schools offering a mixed class is notably higher than the control group of schools that do not offer mixed classes. Furthermore, the average disadvantage scores for mixed class schools are 0.27SD (VMBO), 0.17 SD (HAVO) and 0.29SD (VWO) lower for every track, compared to homogeneous schools. This indicates that in this sample students from lower socioeconomic backgrounds are more likely attend mixed class schools in comparison to peers from higher socioeconomic backgrounds. For all other variables, the sample seems to be quite balanced. Furthermore, the profile ratios for different tracks are comparable among mixed class and homogenous schools, indicating that the selection of profiles does not seem to be related to mixed class or homogenous class attendance.

VMBO variables	Homogeneous			Mixed class		
	N	Mean	SD	N	Mean	SD
Mixed class	206	0.00	0.00	523	1.00	0.00
Teacher/student ratio	206	14.77	2.16	523	14.48	1.99
Disadvantage-score	206	-0.19	0.90	523	0.08	1.03
Exam students	206	83.94	47.64	523	93.75	57.09
Average grade	206	6.31	0.24	523	6.32	0.22
Economics profile ratio	206	0.33	0.21	523	0.36	0.21
Health & welfare ratio	206	0.37	0.21	523	0.34	0.18
Intersectoral ratio	206	0.09	0.22	523	0.11	0.23
Agriculture ratio	206	0.01	0.08	523	0.01	0.04
Technology	206	0.19	0.20	523	0.18	0.15

Table 1: Summary statistics VMBO sample on school level.

HAVO variables	Homogeneous			Mixed class		
	N	Mean	SD	N	Mean	SD
Mixed class	76	0.00	0.00	371	1.00	0.00
Teacher/student ratio	76	15.17	2.35	371	15.00	1.83
Disadvantage-score	76	-0.14	0.98	371	0.03	1.00
Exam students	76	91.89	46.32	371	94.92	42.17
Average grade	76	6.23	0.22	371	6.26	0.22
Culture & Society ratio	76	0.12	0.09	371	0.12	0.09
Economics & Society ratio	76	0.47	0.14	371	0.47	0.15
Nature & Health ratio	76	0.24	0.10	371	0.22	0.11
Nature & Technique ratio	76	0.03	0.05	371	0.04	0.05
Nature, Health and Technique ratio	76	0.11	0.09	371	0.12	0.09
Economics, Society & Culture ratio	76	0.02	0.04	371	0.04	0.10

Table 2: Summary statistics HAVO sample on school level.

VWO variables	Homogenous			Mixed class		
	N	Mean	SD	N	Mean	SD
Mixed class	106	0.00	0.00	348	1.00	0.00
Teacher/student ratio	106	15.82	2.26	348	15.05	1.84
Disadvantage-score	106	-0.41	0.91	348	0.12	0.99
Exam students	106	89.05	44.54	348	69.17	36.15
Average grade	106	6.47	0.28	348	6.37	0.24
Culture & Society ratio	106	0.07	0.07	348	0.05	0.06
Economics & Society ratio	106	0.16	0.14	348	0.22	0.14
Nature & Health ratio	106	0.17	0.10	348	0.19	0.12
Nature & Technique ratio	106	0.07	0.07	348	0.07	0.08
Nature, Health and Technique ratio	106	0.35	0.13	348	0.34	0.14
Economics, Society & Culture ratio	106	0.18	0.12	348	0.12	0.12

Table 3: Summary statistics VWO sample on school level.

3.1.1 Municipal-level dataset

The municipal level dataset is constructed in a similar way as the previous one, combining CBS and DUO data but with different variables. Three cohorts of secondary school students are used. A cohort is defined by the starting year of secondary education of students. The data available limits the range of cohorts to three cohorts, of which the starting and ending years for the HAVO and VWO tracks are show in table 4. Thus, cohort 1 of a municipality is given by students who start secondary education in 2015 and end their education with graduation in 2020 (HAVO) or 2021 (VWO).

Tracks & cohorts	**HAVO**	**VWO**
1	2015/2020	2015/2021
2	2016/2021	2016/2022
3	2017/2022	2017/2023

Table 4: Tracks & cohorts

The dependent variable is the graduation percentage on either HAVO/VWO or HAVO of a cohort. This is calculated by dividing the number of graduates in a municipality from a cohort, by the number of students in a municipality that started any type of secondary education at the first year of a cohort and multiplying the outcome by 100. The number of graduates is calculated from DUO (2022c). This dataset is on school-level and hence the number of graduates per school are aggregated to the municipal-level using the municipalities schools are part of. The number of graduates is measured five years after the starting time of a cohort in the case of HAVO, and six years after the starting time of a cohort in the case of VWO. This means that the percentage of graduates from cohort 1 for HAVO and VWO consist out of the number of HAVO graduates in 2020 plus the number of VWO graduates in 2021. This dependent variable is calculated in two ways. First by using the number of HAVO and VWO graduates, and secondly by only using the number of VWO graduates. This is done given that the idea behind mixed classes is to widen access to higher tracks (HAVO/VWO) for those that would not enter these tracks without mixed classes. By estimating VWO and HAVO/VWO outcomes it can be identified how the number of mixed classes effect average educational performance, but also through which ways. By comparing these two it can be studied if these mixed classes shift VMBO students over the VMBO/HAVO margin, and by comparing the VWO graduates per cohort it can be studied if these mixed classes shift HAVO students over the HAVO/VWO margin.

The number of first year students per cohort is calculated using DUO (2023) which shows the municipality of every school and how much students are in each year of secondary education from 2017 until 2019. I selected the number of first year students per school, summed these and aggregated this to the municipal level. Given that 2015 and 2016 are not available, the number of students who started secondary education in cohorts 1 and 2 are calculated by subtracting the average change of first year students in each municipality from the number of students in first year secondary education from 2017.

The treatment variable is the number of mixed classes per 100 students. This variable is calculated from the total number of mixed classes per municipality at the starting year of every cohort and divided by the number of first year students at the starting year of a cohort. To explain, in for example cohort 1 the number of mixed classes per municipality in 2015 is divided by the number of first year students in secondary education in 2015 and multiplied by 100 (DUO 2022a).

Finally, I add the SES-score on municipal level which is calculated by Statistics Netherlands. This score can be interpreted as how a municipality performs on socioeconomic variables, compared to other municipalities. The score of the average municipality is 0. A score above zero indicates that the citizens of this municipality have on average a higher socioeconomic status than citizens of the average Dutch Municipality. Below zero, the score tells the opposite story. The score is based on wealth, educational background and recent employment activity from citizens of each municipality. This score is standardized to ease interpretation and the score corresponding to each cohort is again the starting year of secondary education. Thus, for cohort 1 this is 2015. (Centraal Bureau voor de Statistiek, n.d.; CBS, 2023b).

In table 5 the summary statistics of the two datasets are shown. The first three columns refer to the resulting data set in which the dependent variable is calculated using both HAVO and VWO graduates. The final three columns refer to the resulting data set in which the dependent variable is calculated using only VWO graduates. What can be noticed is that around 18% of a cohort on municipal level graduates within the duration of the HAVO or VWO cycle, thus within 5 or 6 years. Around 7% of these are VWO students. Further, it is shown that on average 0.72 mixed classes are available per student. This number is comparable between the two samples, given that both are calculated in the same way. Small differences therefore occur when the number of VWO graduates from a municipality for a given cohort is unknown, while this is known for the HAVO graduates from this municipality. Therefore, the sample sizes differ between 177 and 189, but are comparable between the samples and corresponding cohorts.

It can be noted that the number of observations for each cohort are significantly smaller than the number of municipalities in the Netherlands. First, this is due to the reason that some educational data was missing for some municipalities. E.g. the number of graduates in a municipality was known, but there was no information available on the number of mixed classes in this municipality. Furthermore, a handful of observations was deleted. This due to the reason that the graduation percentage must be equal or greater than 0 and below 60. Other values were considered as outliers. Furthermore, the number of mixed classes per 100 students in a municipality cannot be negative.

HAVO/VWO SAMPLE	**Cohort 1 (N=189) HAVO/ VWO**		**Cohort 2 (N=191) HAVO/ VWO**		**Cohort 3 (N=192) HAVO/ VWO**		**Cohort 1 (N=183) VWO**		**Cohort 2 (N=183) VWO**		**Cohort 3 (N=183) VWO**	
Variables	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
% Graduates per cohort	19.50	8.48	17.76	7.74	17.34	7.58	7.12	4.29	6.93	5.49	6.49	4.41
Mixed classes per 100 students	0.76	0.58	0.76	0.53	0.77	0.58	0.71	0.39	0.73	0.42	0.73	0.40

Table 5: Summary statistics of municipal datasets.

3.2 Data analysis.

Identifying the effect of mixed classes on educational performance and its' relationship with inequality in this study comes with several challenges. At first it needs to be acknowledged that just like every other study focusing on ability tracking, the cross-sectional models may suffer from heterogenous trends and endogeneity. Especially because it is impossible to form a random treatment group all obtaining the same treatment, and a random counterfactual group isolated from the treated. For example, students apply by themselves for either a homogeneous class or mixed class due to which random selection is almost never possible. Besides that, the data available is limited in a sense that one could possibly think of other variables that need to be controlled for but are not available. From that perspective, this study faces the same limitations as previous studies explained by Betts (2011). By using cross-sectional, cohort fixed effects and TWFE models, I aim to carefully identify the effect of mixed classes on average educational performance and possible heterogenous effects among different SES groups.

3.2.1 The effect of mixed classes on average educational performance.

To study the effect of mixed classes on average educational performance, I start with the cross-sectional data on school level using the following models:

$$Y_s = \beta_0 + \beta_1 Mixed\ class + \beta_2 \frac{Teacher}{Student} ratio_s + \beta_3\ Disadvantage_s + \varepsilon_s \quad (1)$$

$$Y_s = \beta_0 + \beta_1 Mixed\ class + \beta_2 \frac{Teacher}{Student} ratio_s + Y\ Profile\ \%_s + \beta_3\ Disadvantage_s + \varepsilon_s \quad (2)$$

The regression is estimated for all tracks separately. This is done due the reason that the curricula of the tracks differ, making the average grade at schools of different tracks not comparable. The outcome variable is the average grade of the students at schools at their final year of secondary education. This variable presents the measure of average educational performance in this model. The coefficient of interest is the mixed class coefficient which implies the change in the average grade of exam students on school (s) when a school has heterogeneous classes, compared to schools having homogeneous classes. In this model, this represents the effect of mixed classes on average educational performance as is part of the first sub-question of my research question. Furthermore, the teacher/student ratio and disadvantage scores are added as control variables.

In model (2) the profile ratio is added. This is done after running regression (1) due to the reason that the selection of profiles done by students at a school might also be related with the availability of mixed classes, given that this occurs after attending mixed classes. This might therefore partially take away a treatment effect of mixed class supply. To control for this, the model is first estimated without and after with this control.

Closely related to this thesis is the study done by Borghans et al. (2012), who claim that differences in relative attendance of mixed classes between municipalities are exogenous. As a result, the effect of mixed class supply on educational outcomes can be identified. As explained before, this assumption cannot be taken with certainty, although it can still be true. To test this, I first use a similar but not the same model on municipal level to study the effect of mixed class supply on the average educational performance in a municipality. I use cohort fixed effects to ease interpretation, given that this is similar to using multiple cross-sectional models for each cohort. While using multiple cross-sectional models the coefficients would be averaged and compared, the cohort fixed effects model results in similar comparisons given that it controls for unobservables that change across cohorts. The averages of the coefficients cross-sectional models and coefficients of the cohort fixed effects model will therefore be close to equal.

On municipal level the following cohort fixed effects model will be estimated:

$$Y_{mc} = \beta_0 + \beta_1 Mixed\ classes_{mc} + \beta_2 SES_{mc} + \delta_c + \varepsilon_{mc} \qquad (3)$$

In this model Y_{mc} equals the share of HAVO and VWO graduates from a cohort, or the share the share of VWO graduates from a cohort. This regression is therefore estimated twice as explained in the previous data section.

The coefficient of SES_{mc} presents the effect of a one SD change in the socioeconomic score of a municipality during cohort (c) on the percentage of the HAVO and/or VWO graduates per cohort.

The coefficient of interest is $Mixed\ classes_{mc}$ which shows the increase in the percentage of graduates per cohort in municipal (m) if the number of mixed classes per 100 students would increase by 1. In other words, the coefficient will show what the effect of increasing the number of mixed classes is the on the number of HAVO and/or VWO graduates per cohort in a municipality. By doing so it contributes to answering what the effect of mixed classes is on average educational performance in this cohort fixed effects model. The cohort fixed effects are shown by δ_c and capture any common trend across cohorts.

After analyzing these cross-sectional and cohort FE models on school and municipal level, I will be addressing concerns of endogeneity in the cross-sectional models. To do so, I use the panel-data set on municipal-level and construct a two-way fixed effects model to show what effect of mixed classes supply on average educational performance remains after controlling for time-invariant differences between municipalities. By doing so, I will explain what I claim to be the effect of mixed classes on average educational performance and answer the first sub-question of my thesis. The fixed-effects model is similar to model (2) and again estimated twice, besides that cohort and municipal fixed-effects are added as shown by δ_c and θ_m. This results in the following regression equation:

$$Y_{mc} = \beta_0 + \beta_1 Mixed\ classes_{mc} + + \beta_2 SES_{mc} + \delta_c + \theta_m + \varepsilon_{mc} \quad (4)$$

Whereas it is impossible to control for covariates and endogeneity in a cross-sectional model besides by adding controls, a two-way fixed effects model allows to control for both municipal and cohort specific effects. Thus, if there is endogeneity in the cross-sectional models, the results from the fixed-effects model will be significantly different. By comparing the results of the models in this paper with each other, the aim is to check if these concerns of endogeneity and heterogeneity are indeed relevant while studying inequality, ability tracking and average educational performance besides studying the effect of mixed classes on average educational performance.

In the remaining part of this thesis, I attempt to explain if the effect of mixed classes on average educational performance is heterogeneous among schools and municipalities with different socioeconomic backgrounds. The methodology used is similar to the methodology used previously. I start with using a cross-sectional model on school level and two cohort fixed effects models on municipal level. I end with two two-way fixed effects model on municipal level. The difference between the remainder of the paper and the previous parts is that to study this possible heterogenous effect of mixed classes, I added an interaction term to all models. For the school-level model this interaction term explains how the effect of offering mixed classes on the average exam grade changes when the disadvantage score changes. Thus, this effect explains the possible differential effect of mixed classes between schools with different SES compositions. In the municipal models this interaction term presents of the mixed classes coefficient and the SES-term. The coefficient explains how the effect of the number of mixed classes per 100 students on the graduation percentage changes when the SES-score increases, ceteris paribus. If the effects of mixed classes are indeed heterogenous for different SES municipalities and schools, these effects should be statistically different from 0. If not, then no heterogenous effects are found in these models.

The resulting cross-sectional regression model on school-level is:

$$Y_s = \beta_0 + \beta_1 Mixed\ class_s + \beta_2 \frac{Teacher}{Student} ratio_s + \beta_3\ Disadvantage_s + \beta_5\ (Disadvantage_s * \ Mixed\ class_s) + \varepsilon_s \quad (5)$$

The resulting cross-sectional regression model on school level with the profile ratios added:

$$Y_s = \beta_0 + \beta_1 Mixed\ class_s + \beta_2 \frac{Teacher}{Student} ratio_s + Y\ Profile\ \%_s + \beta_3\ Disadvantage_s + \beta_5\ (Disadvantage_s * \ Mixedclass_s) + \varepsilon_s \quad (6)$$

Secondly, the resulting cohort fixed effects regression model on municipal level is:

$$Y_{mc} = \beta_0 + \beta_1 Mixed\ classes_{mc} + + \beta_2 SES_{mc} + \beta_3 (SES_{mc} * \ Mixed\ classes_{mc}) + \delta_c + \varepsilon_m \quad (7)$$

Finally, the resulting fixed-effects model on municipal level is:

$$Y_{mc} = \beta_0 + \beta_1 Mixed\ classes_{mc} + SES_{mc} + \beta_3\ (SES_{mc} * Mixed\ classes_{mc}) + \delta_c + \theta_m + \varepsilon_{mc} \quad (8)$$

3.3.3 Common trends assumption

Although the TWFE models used in this thesis are not a classical difference in differences in which the coefficient of interest is a binary treatment indicator, there is a treatment that is continuous. Therefore, adherence to the common trends assumption is necessary to claim unbiased results. The common trends assumption assumes that in absence of a treatment, differences in outcomes between control and treatment groups are constant over time. As a result, any change in this trend can be assigned to the treatment. To show adherence with this assumption, researchers commonly plot pre trends of treatment and control groups (Columbia University Mailman School of Public Health, n.d.).

This treatment is as explained before, the number of mixed classes per 100 students in a municipality and differs among cohorts and among municipalities. For example, a municipality could have an increase of 0.3 mixed classes per students during cohort 2 and have a decrease of 0.1 mixed classes per student during cohort 3. The continuity of the treatment and changing treatment exposure complicates testing the common trends assumption.

However as suggested by Hull (2024), testing pre-trends is not the only way in which adherence to the common trend assumption can be argued. This can also be done by adding time-variant controls and interaction-terms between cohort dummies and time-invariant controls. Although the latter is not possible given that my data does not contain time invariant controls, the first is done by adding the SES-control variable. Furthermore, Hull (2024) suggests allowing municipality linear trends by adding an interaction term between municipality dummies and a continuous cohort variable as a robustness check. This results in an even more restricted model compared to the previous TWFE model. The interaction term controls for unit-specific trends. A hypothetical example of such trend could be the move of a bunch of rich people from all-over the Netherlands into one specific municipality whereas an equal number of poorer families leave this municipality. What would be the outcome of such event? It could be the case that these rich kids do not attend these mixed classes but are all very good students, given that they can learn in the best conditions possible. The number of graduates per cohort at the higher levels (e.g. VWO) increase significantly as a result, but the number of mixed classes per student in this municipality remains equal. The outcome would be that an equal number of mixed classes in this municipality causes better average educational performance. However, this change in effect is not caused by a change in mixed classes supply, but by a change in the student composition. By using unit-specific trends such unit-specific shocks can be controlled for. Even though such a scenario is very unlikely, using this specification is a way to check if the found treatment effects are even robust to the inclusion of restrictive specifications and therefore support the internal validity of this study (Hull, 2024). The regression specifications used are as follows:

$$Y_{mc} = \beta_0 + \beta_1 Mixed\ classes_{mc} + \beta_2 SES_{mc} + \sum_{i=1}^{N} \delta\ Municipality_m + \beta_3 Cohort_c + \sum_{i=1}^{N} \theta\ (Cohort_c * Municipality_m) + \varepsilon_{mc} \qquad (9)$$

$$Y_{mc} = \beta_0 + \beta_1 Mixed\ classes_{mc} + \beta_2 SES_{mc} + \sum_{i=1}^{N} \delta\ Municipality_m + \beta_3 Cohort_c + \sum_{i=1}^{N} \theta\ (Cohort_c * Municipality_m) + \beta_4 (SES_{mc} * Mixed\ classes_{mc}) + \varepsilon_{mc} \qquad (10)$$

As can be seen, the regression specifications are again similar to the previous TWFE specifications, although the TWFE operators are replaced by a continuous cohort variable and a dummy for each municipality which are interacted at the end. Robustness of the results in this case is argued when controlling for unit specific trends does not change treatment effects significantly compared to the treatment effects found in the previous TWFE models. Thus, robustness is argued if the found effect of the number of mixed classes per 100 students and the interaction term with the SES variable on the outcome variable do not change significantly compared to the TWFE regression

4. Results.

In this section I will carefully discuss the effects of mixed classes on average educational performance and the different effects found among the different models. Besides that, I will also look at the possible heterogenous effects of mixed classes among different SES municipalities and schools. The latter will be done in the second part of this section, while the first will be done in the following part.

4.1 Cross-sectional models on school-level and average educational performance

Table 6 presents the regression output for the OLS models on school-level. The coefficients of the mixed class indicator for schools among all different tracks are found not to be significant. Therefore, in this model there is found not to be a differential effect on the average exam grade between schools offering or not offering mixed classes. The inclusion of profile controls does not change this conclusion. This indicates that from this model there is no effect of mixed classes on average educational performance found.

Table 6: OLS estimates of mixed classes and control variables on average exam grades on school level.

VARIABLES	VMBO Model 1	VMBO Model 2	HAVO Model 2	HAVO Model 1	VWO Model 1	VWO Model 2
Mixed class	0.02	0.02	0.04	0.04	0.04	0.03
	(0.019)	(0.019)	(0.027)	(0.027)	(0.027)	(0.027)
Disadvantage score (standardized)	-0.05***	-0.05***	-0.08***	-0.08***	-0.07***	-0.03
	(0.009)	(0.009)	(0.011)	(0.011)	(0.014)	(0.030)
Teacher/student ratio	-0.01**	-0.01	-0.01	-0.01	-0.01	-0.01*
	(0.004)	(0.004)	(0.005)	(0.005)	(0.006)	(0.006)
Number of final year students	0.00***	0.00**	0.00	0.00	0.00***	0.00**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Profile controls?	No	Yes	No	Yes	No	Yes
Constant	6.40***	6.38***	6.32***	6.27***	6.47***	6.26***
	(0.070)	(0.076)	(0.087)	(0.101)	(0.095)	(0.122)
Observations	729	729	447	447	454	454
F-test (p-value)	0.00	0.00	0.00	0.00	0.00	0.00
Average exam grade of school	6.32	6.32	6.39	6.39	6.39	6.39

Notes: All columns present OLS estimate for models (1) and (2) as shown by model 1 and model 2 in the table. Robust standard errors (school level) are reported between brackets. Significance levels: $* \ p < 0.10$, $** \ p < 0.05$, $*** \ p < 0.01$.

4.2 Cohort fixed effects model on municipal-level average educational performance

Similar effects to those in the previous section are also found in the cohort fixed effects model on municipal level. The regression outcomes in table 7 do not suggest that increasing the number of mixed classes per 100 students has an effect on the HAVO/VWO and VWO graduates percentage per cohort. The found effects are not significant. This is in sharp contrast with Borghans et al. (2012) who found that early tracking negatively effects average individual educational performance in their cross-sectional model. This does not directly indicate that the effect of mixed classes on average educational performance is non-existent, nor that I reject the findings of Borghans et al. (2012). As explained before, the use of cohort fixed effects models like these, and cross-sectional models that I used on school-level might be prone to endogeneity. This may be caused by unavailable omitted variables or year specific effects in the case of the school-level model. This possibly causes biased estimates and incorrect standard errors. However, it is not known to what extent. For now, it can therefore only be concluded that using a cohort fixed effects model on municipal level and a cross-sectional model on school level I did not find any effect of mixed classes on average educational performance.

Table 7: Cohort fixed effects estimates of the number of mixed classes per 100 students, control variables and the HAVO/VWO or VWO graduates percentage of a cohort on municipal level.

Cohort fixed effects VARIABLES	HAVO/ VWO (1)	VWO (2)
Mixed classes per 100 students	-1.54	2.24
	(1.512)	(1.383)
SES-score	1.74***	0.90***
	(0.551)	(0.316)
Constant	19.37***	5.23***
	(1.129)	(0.918)
Observations	572	550
F-test (p-value)	0.00	0.01
Average % of graduates per cohort	18.20	6.85

Notes: All columns present cohort fixed effects estimates for model (3) with the dependent variable % of HAVO plus VWO graduates per cohort in column 1 or with the dependent variable % of VWO graduates per cohort in column 2. Clustered standard errors (municipal level) are reported between brackets are reported between brackets. Significance levels: ∗ $p < 0.10$, ∗∗ $p < 0.05$, ∗∗∗ $p < 0.01$.

4.3 Two-way fixed effects model on average educational performance.

In table 8 the results of the two-way fixed effects regressions are shown. Whereas in the cohort fixed effects model the effect of the number of mixed classes per 100 students was found to be insignificant, the outcomes of this TWFE model show a different picture. The effect of an increases in the number of mixed classes per 100 students on the percentage of HAVO/VWO graduates from a cohort remains insignificant. However, the effect of a one unit increases in the number of mixed classes on the percentage of VWO graduates from a cohort becomes significant at the 10% level. A one unit increase in the number of mixed classes per 100 students increases the percentage of VWO graduates per cohort by 4.08 percentage points. The magnitude of this effect is strong, especially because the average percentage of VWO graduates from a cohort was 6.85%.

As explained before, the main difference between the previous cohort fixed effects model and the TWFE model used here is that it also controls for time invariant differences between municipalities. This partially controls for (un)observable confounders, reducing bias and resulting in a statistically significant coefficient in this case.

The idea behind mixed classes is often aims to keep opportunities for students who at later tracking moments might be able to enter higher tracks. In this municipal model, this is effect is only found among the students of the HAVO/VWO threshold, given that only these students can move forward to VWO in a mixed class. This idea stems from the regression outcomes which show that an increase in the number of mixed classes significantly increases the percentage of VWO graduates from a cohort. Therefore, I conclude that the net effect of mixed classes on average educational performance is positive in this model.

Table 8: TWFE estimates of the number of mixed classes per 100 students, control variables and the HAVO/VWO or VWO graduates percentage of a cohort on municipal level.

Fixed Effects VARIABLES	HAVO/ VWO (1)	VWO (2)
Mixed classes per 100 students	2.22	4.08*
	(1.378)	(2.236)
SES-score	-0.93	-2.14
	(2.979)	
Constant	16.42***	3.73**
	(1.053)	(1.653)
Observations	567	545
F-test (p-value)	0.91	0.15
Average % of graduates per cohort	18.20	6.85

Notes: All columns present TWFE estimates for model (4) with the dependent variable % of HAVO plus VWO graduates per cohort in column 1 or with the dependent variable % of VWO graduates per cohort in column 2. Clustered standard errors (municipal level) are reported between brackets are reported between brackets. Significance levels: $* \ p < 0.10$, $** \ p < 0.05$, $*** \ p < 0.01$.

4.4 Cross-sectional plus cohort fixed effects models and inequality

The results of the school-level cross-sectional models focused on possible heterogenous effects of mixed classes are shown in table 9. It was expected that the effect of offering a mixed classes on average exam grades at schools from all tracks would be heterogenous between schools with different socioeconomic scores. However, this is not shown by the regression outcomes across tracks. None of the coefficients from the interaction term are significant, indicating that the effects of mixed classes on average educational performance are not heterogeneous for schools with different socioeconomic compositions

The cohort fixed effects output shows that the effects of the number of mixed classes per 100 students on the percentage of VWO graduates from a cohort in a municipality might be heterogeneous across municipalities with different SES compositions. The coefficient is found to be significant and indicates that a 1SD increase in the SES-score of a municipality, increases the effect of an increase of a one unit increase of the number of mixed classes by 1.5 percentage point.

This is remarkable, because it was expected that if the effects of mixed classes were heterogeneous among different SES municipalities, that the effect on the graduates percentage of a cohort would be higher for lower SES groups. From table 10 however, it was found that the magnitudes of these effects are contrary.

This does not confirm that the effects of mixed classes on average educational performance are heterogeneous. As explained before both the cross-sectional and cohort fixed effects model may be prone to endogeneity. In the next section I therefore use TWFE once again to control for time-invariant differences between municipalities.

Table 9: OLS estimates of mixed classes and control variables on average exam grades on school level including interaction term.

VARIABLES	VMBO Model 1	VMBO Model 2	HAVO Model 1	HAVO Model 2	VWO Model 1	VWO Model 2
Mixed class	0.01	0.02	0.03	0.00	-0.04	-0.03
	(0.02)	(0.02)	(0.03)	(0.00)	(0.03)	(0.03)
Disadvantage score (standardized)	-0.03	-0.03	-0.05***	-0.05***	-0.09***	-0.08***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
Interaction term	-0.03	-0.02	-0.03	-0.03	0.03	0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
Teacher/student ratio	-0.01**	-0.01	-0.01	-0.01	-0.01	-0.01*
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Number of final year students	0.00***	0.00**	0.00	0.00	0.00***	0.00**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Profile controls?	No	Yes	No	Yes	No	Yes
Constant	6.40***	6.38***	6.32***	6.28***	6.46***	6.27***
	(0.07)	(0.08)	(0.09)	(0.10)	(0.10)	(0.12)
Observations	729	729	447	447	454	454
F-test (p-value)	0.00	0.00	0.00	0.00	0.00	0.00
Average exam grade of school	6.32	6.32	6.39	6.39	6.39	6.39

Notes: All columns present OLS estimates for models (5) and (6) as shown by model 1 and model 2 in the table. Robust standard errors (school level) are reported between brackets. Significance levels: $* \ p < 0.10$, $** \ p < 0.05$, $*** \ p < 0.01$.

Table 10: Cohort fixed effects estimates of the number of mixed classes per 100 students, control variables and the HAVO/VWO or VWO graduates percentage of a cohort on municipal level including interaction term.

Cohort fixed effects VARIABLES	HAVO/ VWO (1)	VWO (2)
Mixed classes per 100 students	-1.97	1.98
	(1.500)	(1.220)
SES-score	0.60	-0.43
	(1.242)	(0.700)
Interaction term	1.27	1.51*
	(1.484)	(0.776)
Constant	19.71***	5.46***
	(1.123)	(0.803)
Observations	572	550
F-test (p-value)	0.00	0.01
Average % of graduates per cohort	18.20	6.85

Notes: All columns present cohort fixed effects estimates for model (7) with the dependent variable % of HAVO plus VWO graduates per cohort in column 1 or with the dependent variable % of VWO graduates per cohort in column 2. Clustered standard errors (municipal level) are reported between brackets. Significance levels: ∗ $p < 0.10$, ∗∗ $p < 0.05$, ∗∗∗ $p < 0.01$.

4.5 Two-way fixed effects models and inequality

The outcomes of the TWFE models are shown in table 11 The coefficient of the interaction term for the percentage of VWO graduates per cohort becomes insignificant after adding municipal fixed effects. This suggest that the effect found previously was most likely driven by the exclusion of time invariant (un)observables, for which is controlled for in the TWFE model.

Summarizing, both the municipal-level and school-level models do not find any heterogenous effects of the number mixed classes on the average educational performance of students in a municipality. Although it was expected that such heterogenous effects existed, they are not found after adding TWFE.

Table 11: TWFE estimates of the number of mixed classes per 100 students, control variables and the HAVO/VWO or VWO graduates percentage of a cohort on municipal level including interaction term

Fixed Effects VARIABLES	HAVO/ VWO (1)	VWO (2)
Mixed classes per 100 students	2.20	4.28*
	(1.431)	(2.344)
SES-score	-1.53	-3.71
	(3.098)	(3.636)
Interaction term	0.65	1.60
	(1.424)	(2.262)
Constant	16.44***	3.73**
	(1.089)	(1.653)
Observations	567	545
F-test (p-value)	0.91	0.15
Average % of graduates per cohort	18.20	6.85

Notes: All columns present TWFE estimates for model (8) with the dependent variable % of HAVO plus VWO graduates per cohort in column 1 or with the dependent variable % of VWO graduates per cohort in column 2. Clustered standard errors (municipal level) are reported between brackets are reported between brackets. Significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5. Robustness checks

As suggested by Hull (2024) the robustness of the found effects on mixed classes, can tested by replacing the TWFE model with a unit specific trends model. As explained previously this model is more restrictive than the TWFE model. Instead of only controlling for common trends across cohorts, a unit-specific trends controls for municipal specific trends across instead of only common trends across cohorts. As a result, it does not only control for common trends across cohorts e.g. a pandemic but also for unit specific trends as explained in the example in the previous sections.

In table 11 the unit-specific trend model estimates are shown for the models that attempt to explain the effect of the number of mixed classes on average educational performance. What stands out is that both coefficients of the HAVO/VWO and VWO show strongly significance and a high magnitude. A one unit increases in the number of mixed classes per 100 students increases the graduates percentage of HAVO/VWO students from a cohort with 14.31 percentage point. For the VWO graduates percentage per cohort this is 8.43 percentage point. This robustness check therefore confirms the strong positive effect of the number of mixed classes on average educational performance as was found previously but denies that this effect is solely present among VWO graduates.

Similar conclusions can be drawn from the unit-specific trend model estimates resulting from the models that attempt to explain possible heterogenous effects of the number of mixed classes on average educational performance. Although no differential effects among different SES groups are found in the previous results, the unit specific trends outcomes suggest otherwise. This suggest that the absence of heterogeneous effects of mixed classes in the previous TWFE might have been driven by municipal specific trends and therefore I cannot conclude that my results are robust to the inclusion of unit specific trends.

Table 11: Unit-specific trend estimates of the number of mixed classes per 100 students, control variables and the HAVO/VWO or VWO graduates percentage of a cohort on municipal level.

Unit-specific trends VARIABLES	HAVO/ VWO (1)	VWO (2)
Mixed classes per 100 students	14.31*** (5.257)	8.43*** (0.952)
SES-score	2.99** (1.331)	1.20** (0.586)
Observations	567	545
F-test (p-value)	0.00	0.00
Average % of graduates per cohort	18.20	6.85

Notes: All columns present unit-specific trends estimates for model (9) with the dependent variable % of HAVO plus VWO graduates per cohort in column 1 or with the dependent variable % of VWO graduates per cohort in column 2. Clustered standard errors (municipal level) are reported between brackets are reported between brackets. Significance levels: ∗ $p < 0.10$, ∗∗ $p < 0.05$, ∗∗∗ $p < 0.01$.

Table 12: Unit-specific trend estimates of the number of mixed classes per 100 students, control variables and the HAVO/VWO or VWO graduates percentage of a cohort on municipal level including interaction term.

Unit specific trends VARIABLES	HAVO/ VWO (1)	VWO (2)
Mixed classes per 100 students	15.92*** (3.405)	8.42*** (0.940)
Interaction term	11.71** (4.859)	0.78 (1.630)
SES-score	-9.86* (5.487)	0.48 (1.946)
Observations	567	545
F-test (p-value)	0.00	0.00
Average % of graduates per cohort	18.20	6.85

Notes: All columns present unit- specific trends estimates for model (10) with the dependent variable % of HAVO plus VWO graduates per cohort in column 1 or with the dependent variable % of VWO graduates per cohort in column 2. Clustered standard errors (municipal level) are reported between brackets are reported between brackets. Significance levels: ∗ $p < 0.10$, ∗∗ $p < 0.05$, ∗∗∗ $p < 0.01$.

6. Conclusion & Discussion

For at least 10 years the average educational performance in secondary education has been decreasing across OECD countries. This is concerning, especially because education is believed to improve employment, social cohesion and economic growth. It is therefore in the interest of policymakers and academics to study what aspects of educational systems contributes to improving average educational performance. One of those aspects of educational systems is ability tracking. Ability tracking is defined as a system in which students are sorted based on their educational ability. This can be applied at different levels. For example, Dutch students are sorted into different tracks while US students are sorted in different classes while maintaining similar curricula. Whereas within-school tracking has been researched extensively in existing literature, between-school tracking has not been studied into detail yet. Moreover, a small stream of literature suggest that the effects of tracking are heterogenous between groups from different socioeconomic backgrounds. This thesis therefore aimed to contribute to this gap in the literature by examining the effect of ability tracking on average educational performance and breaking down possible heterogenous effects of tracking using SES variables.

The unique Dutch secondary schooling system offers the opportunity to so. It tracks students around the age of 12 and forwards them into different tracks based on their ability. Within this system, some schools offer the opportunity to partially postpone tracking by offering mixed classes. In these mixed classes students of two tracks are put together for at maximum two more years. Unlike schools that offer only homogeneous classes, schools offering mixed classes offer students the opportunity to postpone tracking partially. This system has not been studied frequently yet, except by Borghans et al. (2012) who found that early tracking (homogeneous classes) negatively impacts average educational performance in a cross-sectional model. I argue that the exogeneity of their instrument cannot be claimed, nor rejected. As a result, there is still room for research studying the effect of mixed classes on average educational performance in the Dutch educational system, besides studying possible heterogeneous effects of mixed classes among different SES groups.

I did this by using both school-level and municipal-level data in cross-sectional, cohort fixed effects and two-way fixed effects models. By comparing the effects found among these models I address concerns related to endogeneity in the cross-sectional models and explain what I claim to be the net and possible heterogenous effects of mixed classes on average educational performance in the Netherlands.

The effects of the cross-sectional and cohort fixed effects models on average educational performance for municipal and school level are found to be insignificant, but once municipal fixed effects are added this effect becomes significant. An increase in the number of mixed classes per 100 students in a municipality increases the percentage of VWO graduates per cohort by 4.8 percentage points, indicating that mostly students around the HAVO/VWO border are those who benefit, increasing average educational performance. This effect is not found among VMBO/HAVO students as the effect of increasing the number of mixed classes per 100 students in a municipality on the percentage of HAVO/VWO graduates is not found to be significant. All in all, I conclude that the net effect of mixed classes in Dutch secondary is positive. This is in line with existing literature who found that the effects of ability tracking on average educational performance are negative or neutral at maximum.

Possible heterogenous effects of mixed classes on average educational performance were found in the cohort fixed effects model, but disappeared when municipal fixed effects were added. This suggest that the found effect in this model was mostly driven by time-invariant differences between municipalities causing endogeneity. I conclude therefore that in these models, no heterogenous effects of mixed classes on average educational performance are found among different SES groups in Dutch secondary education. Athough the existing literature suggested that could have been the case, the outcomes do not support this.

To test the robustness of these results, I used municipal specific trends as suggested by Hull (2024). The regression outcomes suggests that after controlling for municipal specific trends the net effect of the number mixed classes on both percentages of HAVO/VWO and VWO graduates are strongly positive and significant. This in contrast to the findings from my TWFE model in which only the effect of the number mixed classes on the percentage of VWO graduates was found to be significant. I suggest therefore that more comprehensive schooling by the use of mixed classes in the interest of Dutch policymakers focused on the net effect of tracking. This due to the reason that the results and existing literature suggest that there is most likely no harm to be done by the use of mixed classes while it might even increase average educational performance.

Notably, the unit-specific effects model suggests that the effects of mixed classes on average educational performance are heterogeneous between different SES municipalities. This suggests that the results found in the previous TWFE models might be driven by endogeneity due to the lack of control on municipal-specific trends. Given that this part of results was found not to be robust, I cannot draw a valid conclusion about the existence of the heterogenous effects of mixed classes among different SES groups.

Expansion of these findings to a greater theory about between-school tracking, average educational performance and inequality must be done with caution for several reasons. First, existing literature has shown that the effects of ability tracking on average educational performance differ between educational systems. Given that educational systems are a complex web of different interrelated aspects that influence average educational performance, findings from a certain educational system cannot be one-to-one applied to another. This also counts for the relationship between average educational performance, inequality and ability tracking in this study. Furthermore, this study faces severe limitations regarding data availability. This results in the fact that the number of added control variables were limited as well as the range of data used.

The imperfections of this thesis therefore open room for further research. A great advantage of the Dutch mixed class system is that a treatment and control group occur which can be studied to identify the relationship between inequality, ability tracking and average educational performance. Future research could therefore focus on gathering new individual data and applying different methodologies such as. difference-in-differences in the Dutch educational system. By doing so it can be tested if effects found in this study can be applied to the greater understanding of ability tracking, average educational performance and (possible) heterogenous effects.

7. References

Ariga, K., & Brunello, G. (2007). Does Secondary School Tracking Affect Performance? Evidence from IALS. *IZA Discussion Paper No. 2643*. https://doi.org/10.2139/ssrn.970246

Betts, J. R. (2011). The Economics of tracking in Education. In *Handbook of the economics of education* (pp. 341–381). https://doi.org/10.1016/b978-0-444-53429-3.00007-7

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

Bles, P., & Van Der Velden, R. (2024). Het belang van gemengde brugklassen. In *Nationaal Cohortonderzoek Onderwijs*. Retrieved May 30, 2024, from https://www.nationaalcohortonderzoek.nl/sites/nco/files/media-files/nco_general_factsheet_flexibility.pdf

Borghans, L., Diris, R., Smits, W., & De Vries, J. (2012). The Impact of early tracking on later-life Outcomes: An Instrumental variable approach. In *https://www.sole-jole.org/assets/docs/13506.pdf*. Retrieved June 28, 2024, from https://www.sole-jole.org/assets/docs/13506.pdf

Blossfeld, H., Buchholz, S., Skopek, J., & Triventi, M. (2016). *Models of secondary education and social inequality: An International Comparison*. Edward Elgar Publishing.

Burger, K. (2016). Intergenerational transmission of education in Europe: Do more comprehensive education systems reduce social gradients in student achievement? Research in Social Stratification and Mobility, 44, 54–67. https://doi.org/10.1016/j.rssm.2016.02.002

Brunello, G., & Checchi, D. (2007). Does school tracking affect equality of opportunity? New international evidence. *Economic Policy*, *22*(52), 782–861. https://doi.org/10.1111/j.1468-0327.2007.00189.x

Columbia University Mailman School of Public Health. (n.d.). *Difference-in-Difference estimation*. Retrieved June 13, 2024, from https://www.publichealth.columbia.edu/research/population-health-methods/difference-difference-estimation#:~:text=Parallel%20Trend%20Assumption&text=It%20requires%20that%20in%20the,observations%20over%20many%20time%20points.

CBS (2023a) *Achterstandsscores per school in het voorgezet onderwijs, 2021*. [Data set]. CBS. https://www.cbs.nl/nl-nl/maatwerk/2023/41/achterstandsscores-per-school-in-het-voorgezet-onderwijs-2021#:~:text=Achterstandsscores%20(geaggregeerde%20onderwijsscores)%20per%20schoolvestiging,met%20een%20risico%20op%20onderwijsachterstanden.

CBS (2023b). *Sociaal-economische status; scores per wijk en buurt, regio-indeling.* [Data set]. CBS. https://opendata.cbs.nl/#/CBS/nl/dataset/85539NED/table

Centraal Bureau voor de Statistiek. (n.d.). *Hoe interpreteer je de SES-WOA-scores en hoe zijn deze bepaald?* Centraal Bureau Voor De Statistiek. https://www.cbs.nl/nl-nl/faq/infoservice/hoe-interpreteer-je-de-ses-woa-scores-en-hoe-zijn-deze-bepaald-

DUO (2022a). *Ontwikkeling brugklassen per gemeente 2014 - 2021.* [Data set]. DUO. https://duo.nl/open_onderwijsdata/voortgezet-onderwijs/onderwijsaanbod/aanbod-brugklassen.jsp

DUO (2022b). *Aanbod brugklassen 2021.* [Data set]. DUO. https://duo.nl/open_onderwijsdata/voortgezet-onderwijs/onderwijsaanbod/aanbod-brugklassen.jsp

DUO (2022c). *Examens vmbo, havo en vwo.* [Data set]. DUO. https://duo.nl/open_onderwijsdata/voortgezet-onderwijs/examens/examens-vmbo-havo-vwo.jsp

DUO (2022d). *Leerling-leraarratio's per instelling.* [Data set]. DUO. https://duo.nl/open_onderwijsdata/voortgezet-onderwijs/personeel/leerling-leraarratio-per-instelling.jsp

DUO (2023). *Leerlingen per vestiging naar postcode en leerjaar.* [Data set]. DUO. https://duo.nl/open_onderwijsdata/voortgezet-onderwijs/personeel/leerling-leraarratio-per-instelling.jsp

European Commission. (n.d.). *Education for jobs and growth*. European Education Area. Retrieved May 31, 2024, from https://education.ec.europa.eu/focus-topics/improving-quality/inclusive-education/education-for-jobs-and-growth

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(3), 497-514. https://doi.org/10.1006/juec.2001.2255

Fleischeuer, T. (2022, December 12). *Pretpakket: waarom bèta vakken meer aanzien hebben dan gamma vakken*. Onderwijs Van Morgen. https://www.onderwijsvanmorgen.nl/pretpakket-waarom-beta-vakken-meer-aanzien-hebben-dan-gamma-vakken/

Fu, C., & Mehta, N. (2018). Ability tracking, school and parental effort, and student achievement: a structural model and estimation. Journal of Labor Economics, 36(4), 923–979. https://doi.org/10.1086/697559

Guyon, N., Maurin, É., & McNally, S. (2012). The Effect of Tracking Students by Ability into Different Schools. *The Journal of Human Resources*, *47*(3), 684–721. https://doi.org/10.3368/jhr.47.3.684

Hanushek, E. A., & Woessmann, L. (2006). Does educational tracking affect performance and inequality? Differences- in-Differences Evidence across countries. *Economic Journal (London. Print)*, *116*(510), C63–C76. https://doi.org/10.1111/j.1468-0297.2006.01076.x

Horn, D. (2009). Age of selection counts: a cross-country analysis of educational institutions. Educational Research and Evaluation, 15(4), 343–366. https://doi.org/10.1080/13803610903087011

Hull, Peter. (2024). Chapter 3: Parallel Trends [PowerPoint slides]. Department of Economics, Brown University. https://www.dropbox.com/scl/fi/kft1q9opeox2sh2dm407l/Chapter3.pdf?rlkey=dql20451xrc0fuvovvab5992o&raw=1

Jacobs, M., Van Der Velden, R., & Van Vugt, L. (2021). Does lowering the bar help? Results from a natural experiment in high-stakes testing in Dutch primary education. *GSBE Research Memoranda*, *004*. https://doi.org/10.26481/umagsb.2021011

Lassibille, G., & Gómez, L. N. (2000). Organisation and Efficiency of Education Systems: Some empirical findings. Comparative Education, 36(1), 7–19. https://doi.org/10.1080/03050060027737

Lavrijsen, J., & Nicaise, I. (2016). Educational tracking, inequality and performance: New evidence from a differences-in-differences technique. *Research in Comparative and International Education*, *11*(3), 334–349. https://doi.org/10.1177/1745499916664818

Matthewes, S. H. (2021). Better Together? Heterogeneous effects of tracking on student achievement. *The Economic Journal*, *131*(635), 1269–1307. https://doi.org/10.1093/ej/ueaa106

Meghir, C., & Palme, M. (2005). Educational reform, ability, and family background. The American Economic Review, 95(1), 414–424. https://doi.org/10.1257/0002828053828671

OECD & PISA. (2020). PISA 2018 Results (Volume V). *OECD Publishing*. https://doi.org/10.1787/ca768d40-en

OECD. (2023). PISA 2022 results (Volume I): The state of learning and equity in education. OECD Publishing. https://doi.org/10.1787/53f23881-en

Pekkarinen, T., Uusitalo, R., & Kerr, S. P. (2009). School tracking and development of cognitive skills. Social Science Research Network. https://doi.org/10.2139/ssrn.1369804

Posthumus, H., Walhout, J., Rouw, B., Scholtus, S., & Zweerink, J. (2021, May 14). *Herziening onderwijsachterstandenindicator vo*. Centraal Bureau Voor De Statistiek. Retrieved May 11, 2024, from https://www.cbs.nl/nl-nl/longread/rapportages/2021/herziening-onderwijsachterstandenindicator-vo/4-samenvatting-en-conclusies

Rijksoverheid. (n.d.-a). *Schooladvies toelating voortgezet onderwijs*. Rijksoverheid.nl. Retrieved April 12, 2024, from https://www.rijksoverheid.nl/onderwerpen/schooladvies-en-doorstroomtoets-basisschool/toelating-voortgezet-onderwijs-gebaseerd-op-definitiefschooladvies#:~:text=Middelbare%20school%20neemt%20advies%20basisschool%20over&text=De%20school%20moet%20bijvoorbeeld%20een,deze%20klassen%20de%20leerling%20komt.

Rijksoverheid. (n.d.-b). *Welke soort scholen voor het voortgezet onderwijs zijn er?* Rijksoverheid.nl. Retrieved April 12, 2024, from https://www.rijksoverheid.nl/onderwerpen/voortgezet-onderwijs/vraag-en-antwoord/soorten-scholen-voortgezet-onderwijs#:~:text=In%20het%20voortgezet%20onderwijs%20zijn,Sommige%20scholen%20geven%20tweetalig%20onderwijs.

Rindermann, H. (2007). The g-factor of international cognitive ability comparisons: The homogeneity of results in PISA, TIMSS, PIRLS and IQ-tests across nations. European Journal of Personality: Published for the European Association of Personality Psychology, 21(5), 667-706.

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*, 100661. https://doi.org/10.1016/j.rssm.2021.100661

Van De Werfhorst, H. G. (2019). Early tracking and social inequality in educational attainment: educational reforms in 21 European countries. *American Journal of Education*, *126*(1), 65–99. https://doi.org/10.1086/705500

Waldinger, F. (2007). 'Does tracking affect the importance of family background on students' test scores?', Working Paper, London School of Economics.

Waalhout, J., Looijen, S., & Jochem Zweerink. (2022). Plausibiliteitsanalyses Onderwijsachterstandenindicator Primair Onderwijs 2018 - 2021. In *CBS*. Retrieved April 12, 2024, from https://www.cbs.nl/nl-nl/longread/aanvullende-statistische-diensten/2022/plausibiliteitsanalyses-onderwijsachterstandenindicator-primair-onderwijs

8. Appendix

Link to all files used in this study:

https://drive.google.com/drive/folders/1gOBWvlUU1sxSh7m4SRVb8YNBroSj08wj?usp=sharing