Additional Funding for Special Needs Students: Quasi-experimental Results From the Netherlands[1]

Spring 2019

Roel D. Freriks*
University of Groningen

Jochen O. Mierau
University of Groningen

Abstract

We compare the mental health and academic performance of Dutch special needs students and regular students before and after the introduction of a school inclusion policy in 2003. The school inclusion policy consisted of a personal budget for eligible special needs students to support progression in regular education. We use five waves of data between 2000 and 2013 from 1,112 students, aged 10 to 12 at baseline, of the TRacking Adolescents' Individual Lives Survey. We use validated instruments to measure mental health and academic performance. Following norms for eligbility of the Dutch Ministry of Education, Culture and Science, 295 of the 1,112 students were identified as special needs students. We use a Differences-In-Differences (DID) design to account for time trends unrelated to the policy, enabling direct identification of the effect. Individual-level covariates were included in the extended DID model to reduce the within-group variance. Estimates suggest that the school inclusion policy reduced the inequality in mental health and academic performance between special needs and regular students with 26.2 and 65.1 percent, respectively. Moreover, the achievement gap disappeared before adulthood, as special needs students did not differ in educational attainment at labour market entry from their regular peers. Ten-percent quantile DID estimates show that the policy was the most (least) effective at the lower (higher) end of the mental health distribution. Estimates and statistics are robust to informant bias and propensity score reweighting. Hence, this study shows the potential of supporting progression of special needs students in regular education to reduce inequality in health and human capital. Heterogeneity analyses present a cautionary note, as girls and children from an ethnic minority or a lower socio-economic environment benefited less from the policy.

Keywords: Mental Health, Human Capital, Achievement Gap, Education Policy
JEL Codes: I14, I24, I26, I28

[1]This paper uses data from the TRacking Adolescents' Individual Lives Survey (TRAILS). Our special thanks go out to Tineke Oldehinkel for useful feedback during the application procedure and Dennis Raven for managing and providing the data. Additionally, many thanks for invaluable comments from Gabriella Conti, Hans van Kippersluis, Martin Salm, and Beatriz Eugster. Conferences/Seminars: Tinbergen Institute Seminar (Rotterdam, Netherlands - March 19, 2019), Royal Economic Society Annual Conference (Coventry, United Kingdom - April 17, 2019). Finally, the findings and views reported in this paper are those of the authors and should not be attributed to individuals or organisations mentioned here.

1 Introduction

A large series of studies show that if children perform better in education this significantly increases their chances in life, as education is an important component of human capital and relates to a variety of outcomes throughout the life cycle. The literature demonstrates that more education leads to a higher probability of work (Oreopoulos, 2006b, 2006a, 2007), a higher income (Bhuller, Mogstad, & Salvanes, 2017; Devereux & Hart, 2010; Leigh & Ryan, 2008), healthier behavior (Brunello, Fort, Schneeweis, & Winter-Ebmer, 2016; Grimard & Parent, 2007; Jürges, Reinhold, & Salm, 2011), a smaller chance of becoming ill (Kemptner, Jürges, & Reinhold, 2011; Oreopoulos, 2006b, 2007), fewer teenage pregnancies (Cygan-Rehm & Maeder, 2013; Silles, 2011) and less crime (Cullen, Jacob, & Levitt, 2006; D. J. Deming, 2011; Amin, Flores, Flores-Lagunes, & Parisian, 2016).

To reduce educational disparities, effective policies have been developed and implemented.[1] In this study we focus on a Dutch school inclusion policy introduced in 2003, consisting of a personal budget for eligible special needs students to support progression in regular education.[2] To evaluate the policy effectiveness, we use a Differences-In-Differences (DID) design with mental health and academic performance as outcome variables. We hypothesize that the educational disparities are related to students' mental health. Specifically, the policy enhances the mental health of special needs students, enabling their progress in regular education. Subsequently, the progression in regular education reduces the inequality in academic performance with their regular peers.

Several studies show, indeed, a strong association between health and academic performance in the schooling period (Currie & Stabile, 2006; Currie, 2009; Aizer, Currie, Simon, & Vivier, 2018). For example, Aizer et al. (2018) linked preschool blood lead levels with third grade test scores for Rhode Island children born 1997-2005. They show that reductions of lead from even historically low levels have significant positive effects. Currie and Stabile (2006) focused on mental health, and examined US and Canadian children with symptoms of Attention Deficit Hyperactivity Disorder (ADHD). They found large negative effects on test

*Corresponding author. Faculty of Economics and Business, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands. Phone: +31(0)50 363 7018, E-mail: r.d.freriks@rug.nl.

[1] We refer the reader to Section 2 for an overview of the leading examples in the literature.
[2] We refer the reader to Section 3.2 for a detailed description of the policy.

scores and schooling attainment and suggest that mental health conditions are an even more important determinant of educational outcomes than physical health conditions.

Our DID estimates confirm our hypothesis and, hence, suggest that the Dutch school inclusion policy positively affected special needs students' mental health with a relative improvement of 26.2 percent to their regular peers. Our findings, indeed, show a strong relationship between mental health and education, as the educational disparities reduced with 65.1 percent during the same period. This led to no difference in educational attainment before adulthood. Hence, the school inclusion policy reduced the inequality in health and human capital within the school period, providing equal skills at labour market entry.

This paper is structured in the following way. Section 2 provides leading examples of school policies that delivered scientifically proven result. Subsequently, the Dutch education system and the school inclusion policy are described in Section 3. Section 4 provides a detailed description of the data. Subsequently, Section 5 describes the empirical strategy. Results from this strategy are provided in Section 6. Supplementary analyses of the results are given in Section 7. Section 8 contains a detailed discussion of the strengths and limiations of the study. Finally, the conclusion of this study can be found in Section 9.

2 School Policies

One of the most effective ways to prevent learning impairments is to put a good teacher in front of the class. In many OECD countries, this is evident from an overwhelming amount of literature (Aaronson, Barrow, & Sander, 2007; Chetty et al., 2011; Chetty, Friedman, & Rockoff, 2014a, 2014b; D. Deming, 2009; Hanushek & Rivkin, 2006; Hanushek, Piopiunik, & Wiederhold, 2014; Nye, Konstantopoulos, & Hedges, 2004; Papay & Kraft, 2015; Rivkin, Hanushek, & Kain, 2005; Staiger & Rockoff, 2010). This literature measures quality with the so-called Value-Added (VA) score, which indicates how much the student's learning achievement is driven by the quality of the teacher, measured as the difference between the best and the least quarter of the entire teacher population. Good teachers increase the students' language and math skills with 0.35 SD and 0.48 SD, respectively (Nye et al., 2004).[3] Students who received lessons from better teachers benefit their entire life from it. They achieve

[3] In this chapter, learning gain is expressed in the Standard Deviation (SD) of a test score. If students achieved 0.5 SD learning gain, they reached on average a level of education higher as a result of the policy.

a higher level of education and earn a correspondingly higher salary (Chetty et al., 2014b). Chetty et al. (2014b) also shows that students who belong to an ethnic minority or are from a poor family benefit less from a good teacher.

In addition, extra lessons yield higher learning results. Fitzpatrick, Grissmer, and Hastedt (2011) show that one year having regular lessons at primary school counts for 1.2 SD in math and 0.9 SD in language. It therefore seems reasonable to assume that extra lesson time would be beneficial for students' test scores. Similarly, Bellei (2009) finds that an extension of the Chilean school day from half a day to a 'whole school day' leads to 0.07 SD increased test scores on math and language. Carlsson, Dahl, Öckert, and Rooth (2015) found that ten extra school days in secondary education in Sweden lead to a 0.01 SD higher score on an intelligence test for eighteen-year-olds. Research by Agüero and Beleche (2013) shows that extra learning time resulted in much less extra learning performance among children in primary schools in disadvantaged neighborhoods than in rich schools. They attribute this to the fact that in Mexico schools in disadvantaged neighborhoods have significantly fewer educational resources at their disposal. To allow everybody more hours to go to school, therefore, led to an increase in inequality in Mexico.Sims (2008) also found that an extra week school time at rich schools in America led to 0.4 SD increase in math, schools in less rich districts, however, did not improve their students' math performance. The extra week had no effects on language performance in both types. Furthermore, a series of studies have recently been conducted into the 'double dose'-approach in the US. This concerned students from the last group of primary school and the first two years from high school. Students with a learning impairment in a certain subject received this course again half a year later. Additional support lessons in math allow children who count badly enough to be allowed to participate do 0.16-0.18 SD better, than children who just could not count well enough to participate (Taylor, 2014). Banerjee, Cole, Duflo, and Linden (2007) conducted a similair study. He compared participants with children from schools that not yet participated (but will soon join in). The program took place in primary schools in poor neighborhoods. Children with learning impairments received half of each school day separate lesson to reduce their learning gap. Banerjee et al. (2007) found that students were after a year 0.35 SD better in math and 0.19 SD in language than students who were the same at the beginning lagging behind, but in schools where this program was not yet offered. One year after the end of the program, the children that still followed the program were still 0.18 SD better in math and 0.08 SD better in language than children who had not followed the program.

Another way to reduce learning impairments is extra support in and around the class, for example by a class assistant, a teaching assistant, a teacher supporter or a second teacher in the classroom. The study by Andersen, Beuchert, Nielsen, and Thomsen (2018) shows that teaching assistants generate effect. They improved language skills of students in the last group of primary school with 0.13 SD with a teaching assistant at least 14.5 lessons per week for eight months. The language skills of children from highly educated parents did not change due to the use of teaching assistants, while children of lower educated parents increased 0.18 SD. Furthermore, in Serbia it has been investigated whether a supervisor for Roma students in primary education is effective for students with learning impairments and dropouts (Battaglia & Lebedinski, 2015). The Roma population is an ethnic minority who is often poor, poorly educated and unemployed. With the help of a supervisor who is Roma herself, schools can pay extra attention to teaching Roma students. Nearly 4,000 Roma students have been examined. The program proves effective only in improving language and math in schools where the supervisor works with fewer children. Since every school only has one Roma supervisor, the supervisor has more time per student. In schools with fewer than 43 Roma students, the language skills of the Roma students improved with 0.42 SD compared to schools with more Roma students. They increased 0.58 SD in math. The difference in gains between language and math explain to the researchers that the gain in language helps with the better understanding of math.

Another researched policy to stimulate learning performance is reducing class sizes, as demonstrated in a large amount of literature on class sizes in primary school. For example, Angrist and Lavy (1999) found that a class reduction of seven students led to 0.18 SD better learning achievements. Urquiola (2006) shows that a class reduction from 30 to 22 students in Bolivia led children to score 0.3 SD better on language and 0.22 SD better on math. However, the literature on class size was and still is dominated by the Student-Teacher Achievement Ratio (STAR) experiment from the mid-eighties of the last century. In the state of Tennessee 6,500 children from the lower classes of primary school were randomly divided into small classes (13 to 17 students) and ordinary classes (22 to 25 students). Class size had an effect on learning performance of the children during and immediately after the experiment

(Konstantopoulos & Chung, 2009). The effect on learning performance ebbed away slowly in the long-term from 0.21 SD during the smaller class period to 0.16 SD in the last group of primary education (Konstantopoulos & Chung, 2009; Ding & Lehrer, 2010). However, studies show that, even though the effect ebbeb away in primary education, the class reduction has had long-term effects. Adults who were assigned to a small class at baseline, appear after secondary education to take more often the entrance exam for tertiary education (Krueger, 1999), score higher on the entrance examination (Krueger & Whitmore, 2001), and obtain their degree from tertiary education more often (Dynarski, Hyman, & Schanzenbach, 2013).

In the case of very large learning impairments, schools can let children repeat a class. In this extra year, students can catch up and reach the required level, so that they are able to progress on that level in the subsequent school years. The idea behind this is that children who perform insufficiently and still move on to the next grade have insufficient basic knowledge to be able to handle the advanced teaching material. Social promotion, a system in which all students regardless of their learning performance can proceed after the summer holidays to the next school year, can cause that learning impairments of these children increase even further. Compared to social promotion repeat a class can, in theory, increase learning performance. Of course, it is not possible to randomly allow children to repeat a class, so the effect is measured by comparing the learning achievements of children who almost repeated a class with children who actual repeated a class. Schwerdt, West, and Winters (2017) demonstrated that students who repeated a class during primary education are 0.26 SD better in language and 0.31 SD in math after a year than their ex-classmates who had just passed.

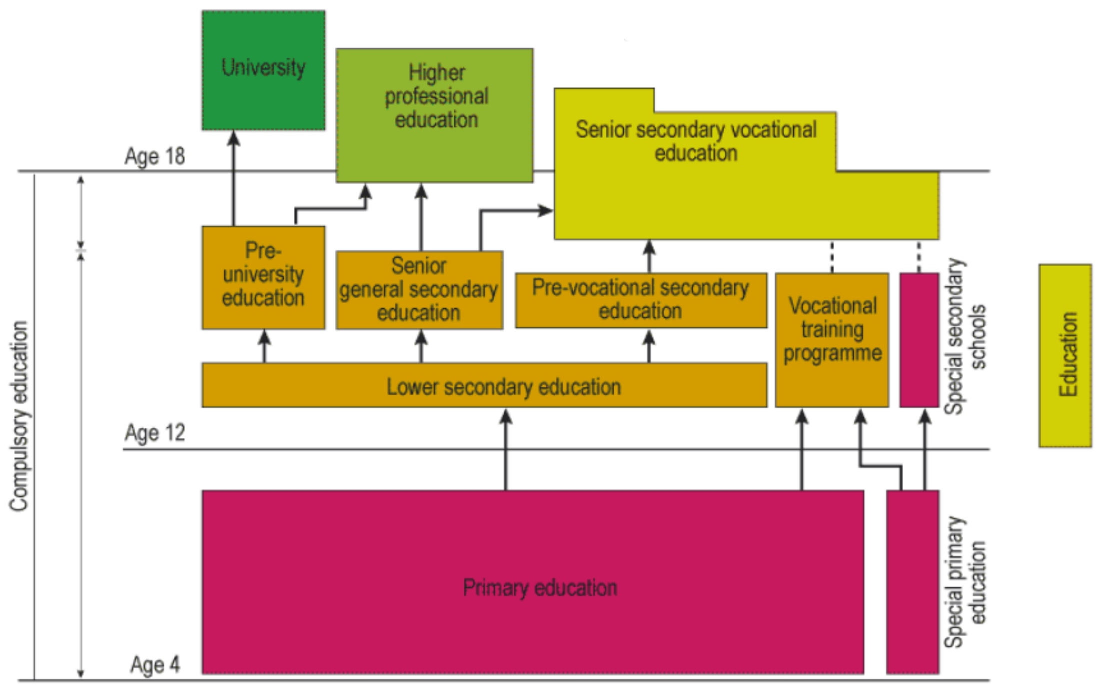
3 Dutch Context

3.1 Dutch Education System

A flowchart of the Dutch education system is presented in Figure 1. Most children start primary school at the age of four, although they are not required by law to attend school until the age of five.[4] At the end of primary school (at around 12 years) pupils receive an

[4] In this study we focus on the primary and secondary school period. However, children have the opportunity to follow early childhood education (Dutch: Voor- en Vroegschoolse Educatie (VVE)). Early childhood education is a form of education designed to optimize the development opportunities of children aged two to five. Early childhood education for pre-schoolers (two to four-year-olds) is provided by playgroups and day nurseries. Primary schools offer VVE programs for four and five-year-olds.

Figure 1: Dutch Education System



advice which type of educational programme would be the most appropriate, given the school results and capacities of the pupil. The advice is based on the judgment of the primary school teachers about the capacity of the pupil and on the results of an objective test. About 95% of the schools use the national test provided by Cito[5], the Dutch Testing and Assessment Company. Based on this advice parents and pupils can apply for the school they prefer and that provides the appropriate educational program.

After finishing primary education children can apply for secondary education. Secondary education consists of pre-vocational secondary education (Dutch: Voorbereidend Middelbaar BeroepsOnderwijs (VMBO)), general secondary education (Dutch: Hoger Algemeen Voortgezet Onderwijs (HAVO)) and pre-university education (Dutch: Voorbereidend Wetenschappelijk Onderwijs (VWO)). Schools are free in the composition of the classroom (homo- or heterogeneous) and in the manner they design education towards the attainment standards. Most schools in secondary education are combined schools, which makes it possible to make the first two years of secondary education as common as possible in so called bridge classes.

[5] We refer the reader to http://cito.com for more information.

This makes it easy during these years to switch from one type of secondary education as possible. Combined schools also give better opportunities for switches in the upper levels of secondary education. After finishing the first years of lower secondary education, students make their final decision on one of the three types of secondary education.

Pre-vocational secondary education (four years) is meant as a preparation for senior secondary vocational education (Dutch: Middelbaar BeroepsOnderwijs (MBO)). General secondary education (five years) is designed to prepare students for higher professional education (Dutch: Hoger BeroepsOnderwijs (HBO)). In practice, however, general secondary education school-leavers also go on to the upper years of pre-university education or to senior secondary vocational education. Finally, pre-university education (six years) is designed to prepare students for university (Dutch: Wetenschappelijk Onderwijs (WO)) and in most cases pre-university education certificate-holders go on to university. Some of them though enter higher professional education.

For children with disabilities, or such special needs that they cannnot attend regular education, special education is provided. Children can be placed into a special school based on an indication given by an indication committee. It is up to the parents to decide whether their child attends a special school or a regular school, given the indication. Is the latter the case the regular school will get extra funding through a personal budget which is provided to the parents, as further discussed in Section 3.2. There are special schools for the visually handicapped children, for the deaf or the ones with speaking problems, for the physically handicapped and for the children with severe behaviour problems. There are special primary schools for children of primary education age (Dutch: Speciaal Onderwijs (SO)) and special secondary schools for children from twelve age onwards (Dutch: Voorgezet Speciaal Onderwijs (VSO)). For children in special secondary schools a new provision is developed, the so-called on the job training schools. Special secondary schools students are stimulated to combine school with work from age sixteen on. This provision is aimed for an adequate labour market perspective of these students.

3.2 School Inclusion policy

In the Netherlands Regional Expertise Centra (REC) are aimed at the indication of special needs students. Each school for special education participates of one of these 34 RECs. In the

August of 2003 the Dutch Ministry of Education, Culture and Science introduced a funding scheme for special needs students to support progression in regular education ('school inclusion policy'). Specifically, special needs students could - based on the decision of the Committee of Indication of Assessment of the REC - became funded through a personal budget. Parents had the opportunity to allocate this budget to their preferred regular school. The system on appropriate education (Dutch: Passend Onderwijs) for special needs students came under serious change, due to heavy criticism and because the budget which is necessary to support the system was exploding. Therefore, the school inclusion policy has been abolished in July 2014. In the new policy the Dutch Ministry aimed to reduce the impact of the indication system, while raising quality of education, professionalization of staff, strong involvement of parents, and budget management. This new policy involved a budget cut of 300 million euros. In study we look back on this decision and evaluate the school inclusion policy of 2003 with student-level data between 2000 and 2013.

4 Data

We use data of the TRacking Adolescents' Individual Lives Survey (TRAILS), a prospective cohort study of Dutch children, who have been measured biennially at least until they were 24 years old. The key objective of TRAILS is to follow children into adulthood, both at the level of psychopathology and the levels of underlying vulnerability and environmental risk, to better understand disparities in mental health. Briefly, the TRAILS sampled ten-to-twelve-year-olds living in five municipalities in the North of the Netherlands, including both urban and rural areas to construct a representative sample. A detailed description of the sampling procedure and methods is described in the corresponding cohort papers (Huisman et al., 2008; Oldehinkel et al., 2015).

4.1 Cohort Description

The population cohort includes participants born between 1 October 1989 and 30 September 1991, who lived in one of the five municipalities the North of the Netherlands at the time of the baseline assessment in 2000. In total 2,230 children were included in the population cohort, corresponding to a response rate of 76%. The mean age was 11.1 and 51% of the sample were girls (Huisman et al., 2008). Subsequent data collection waves took place bi- or triennially, and all waves had good retention rates with 71% of the original cohort participating in all five data collections (Oldehinkel et al., 2015).

Few years after the start of the clinical cohort, in 2004, TRAILS started with the design of a comparable clinical cohort. The clinical cohort consists of individuals who have been referred to a child psychiatric outpatient clinic in the Northern Netherlands any time before the age of 11. In total 543 children were included in the clinical cohort, corresponding to a response rate of 43%. The mean age was 11.1 and 34% of the sample were girls (Huisman et al., 2008). Boys were overrepresented in the clinical cohort, because of their dominance in the most prevalent diagnostic groups in the outpatient clinical (ADHD, disruptive behavior and autism-spectrum disorders). Similairly to the population cohort, data was collected in follow-up waves at intervals of 2-3 years. 78% of the original cohort participated in all four collections (Oldehinkel et al., 2015).

The sample of those who missed one ore more follow-up waves in the population or clinical cohort contains more boys, are from a lower socio-economic environment, and had more externalizing problems at baseline (Oldehinkel et al., 2015). Extensive recruitment efforts were made to increase the representativeness of the cohort to (partially) prevent a non-response bias in estimates based on the mental health of the included participants (Huisman et al., 2008).

4.2 Outcome Variables

Mental health: We use the Child Behavior Checklist (CBCL) to measure students' mental health between 2000 and 2006. The CBCL is a parent questionnaire of 113 items, validated by the Achenbach System of Empirically Based Assessment (ASEBA) (Achenbach & Rescorla, 2013). All items are scored by the caregiver with a 0 ('not applicable at all'), 1 ('a little or sometimes applicable') or 2 ('clear or often applicable'). We use the average score between 0 and 2 with higher values for worse mental health. The questions about behaviour are the problem scales (withdrawn/depressed, physical complaints, anxious/depressed, social problems, thinking problems, attention problems, norm-deviant behaviour and aggressive behaviour). The first three problem scales together form the broad-band syndrome 'internalizing problems', the latter two the broad-band syndrome 'externalizing problems', and all problem scales together form the scale ‘total problems’ (Achenbach & Rescorla, 2013). For the main results of this study we used the scale total problems of the CBCL. We split the broad-band syndromes in the supplementary analyses.

School indicators: We use a validated teacher questionnaire to measure students’ academic performance between 2000 and 2006. This questionnaire, developed by TRAILS, consists of five items, i.e. student’s work pace, effort, achievement relative to their own capabilities, test scores on language, and test scores on math. The teacher assessed all items with a score ranging from 1 (‘strongly disagree’) to 5 (‘strongly agree’). We use the average score between 0 and 5 with higher values for better academic performance. Since the first three items are subjectively determined, we also use only the last two items and stratify by language and math peformance. Additionally, we use the ‘apprenticeship scores’ in 2003 and 2006 and ‘educational attainment’ in 2013 as school indicators. Apprenticeship score is a score between 1 and 7 in 2003 and 2 and 10 in 2006 and presents the relative (level of) progression in education. Finally, educational attainment is the highest obtained level of education in 2013, which is a categorical variable from 1 to 5 (1 = elementary education; 2 = lower tracks of secondary education; 3 = higher tracks of secondary education; 4 = senior vocational education; 5 = university).

Control variables: We use a dummy variable for ethnicity with non-Caucasian background as reference category for at least one foreign born parent. To define the family compostion, we use the number of children in the family. Educational attainment of the parents is similarly specified as for the children. Family income is a categorical variable with nine monthly household income classes, ranging from less than €680.67 (score 1), between €680.67 and €1,134.45 (score 2), ..., between €3,403.35 and €3,857.13 (score 8), to more than €3,857.13 per month (score 9). Futhermore, in the Netherlands, each municipality receives money for general purposes from the national “municipalities fund”. Many municipalities (especially the bigger ones) use this money for educational purposes. Children in the sample live in one of the following cities: Groningen, Leeuwarden, Assen, Winschoten, Dantumadeel, and

Grijpskerk. We discerned a dummy variable with large municipality as the reference category for individuals from the cities Groningen, Leeuwarden and Assen.[6]

4.3 Sample Selection

In this study we attempt to determine the effect of a Dutch school inclusion policy introduced in the August of 2003. For the empirical strategy, mental health and educational outcomes pre- and post-policy introduction are required. Since the policy has been introduced in the August of 2003, and is mainly focused on children (almost) eligible for regular education, effects are estimated on the population cohort. Data collection for children in the clinical cohort started in 2004, and these children are overrepresented in special education (Oldehinkel et al., 2015).

As mentioned above, mental health and educational outcomes pre- and post-policy introduction are required for evaluation. The first two waves of TRAILS, the years 2000 and 2003, are defined as the pre-policy period and the third wave, the year 2006, as the post-policy period.[7] Furthermore, in the extended empirical strategy we use a set of control variables. Consequently, children with missing mental health or educational outcomes in the first three waves or in the set of control variables at baseline are dropped from the sample. The selected sample consists of 1,112 children, which means that we dropped 50 percent of the original cohort of 2,230 children. Baseline characteristics of the dropped and selected sample are provided in Table 1.

The selected sample contains slightly younger children, more girls and less children from a non-Caucasian background, at baseline. Furthermore, children are from a higher socio-economic environment and their average mental health at baseline is higher. Although on some covariates the dropped and selected samples differ statistically significantly, the difference on mental health at baseline is economically small.

4.4 Identification of Special Needs Students

As discussed, eligibility for the school inclusion policy was assessed by the Committee of Indication of Assessment of the Dutch Ministry of Education, Culture and Science. We iden-

[6] In the empirical strategy we use the full set of cities as control variable.

[7] Although for some children data has been collected after the second wave, implementation of the school inclusion policy started for these students after the second data collection.

Table 1: Baseline Characteristics Dropped and Selected Sample

	Dropped Sample $N = 1,118$	Selected Sample $N = 1,112$
Student		
Mental health (0-2)	0.27	0.23***
Age (10-12)	11.13	11.09*
Males	0.51	0.47*
Non-Caucasian background	0.16	0.06***
Family background		
Number of children (0-8)	2.50	2.57
Educational attainment mother (1-5)	2.64	3.13***
Educational attainment father (1-5)	2.85	3.29***
Family income (1-9)	3.67	5.05***
Demographic		
Large municipality	0.85	0.84

Note: */**/*** indicate significant differences at the 10%/5%/1% level based on the mean differences of the two groups, assessed with a t-test.

tified special needs students accordingly. Baseline CBCL scores were discerned into normal-borderline-clinical cutoffs based on the validated Dutch norms of the ASEBA (Achenbach & Rescorla, 2013).[8] Students with a CBCL score above the clinical cutoff ($N = 171$) were assumed eligible. Additionally, students with dyslexia ($N = 44$) or a hearing aid ($N = 80$) complemented the group of in total 295 special needs students. Other students are classified as regular students, not eligible for the school inclusion policy. Baseline characteristics and school indicators fo both groups are provided in Tables 2 and 3, respectively.

The sample of special needs students consists of more boys and children live more frequent in a larger municipality. Furthermore, we observe that special needs students are from a similar ethnic and socio-economic environment as their regular peers. Furthermore, the means of the school indicators suggest that the achievement gap between special needs students and their regular peers disappeared after the policy. The means of all indicators were statistically significantly different before the policy ($p < 0.01$), while the differences were not statistically significant after the policy ($p > 0.10$).

[8] We multiplied the baseline CBCL score with 119. Subsequently, we followed the validated cutoffs for boys younger than twelve years old of CBCL scores below 39 ('normal'), between 39 and 48 (' borderline') and above 48 ('clinical'). The validated cutoffs for boys of twelve years or older are CBCL scores below 40 ('normal'), between 40 and 52 (' borderline') and above 52 ('clinical'). Similarly, we followed the validated cutoffs for girls younger than twelve years old of CBCL scores below 36 ('normal'), between 36 and 48 (' borderline') and above 48 ('clinical'). Finally, the validated cutoffs for girls of twelve years or older are CBCL scores below 36 ('normal'), between 36 and 44 (' borderline') and above 44 ('clinical').

Table 2: Baseline Characteristics of Special Needs and Regular Students

	Regular Students $N = 817$	Special Needs Students $N = 295$
Student		
Mental health (0-2)	0.18	0.37***
Age (10-12)	11.10	11.06
Males	0.45	0.54***
Non-Caucasian background	0.06	0.05
Family background		
Number of children (0-8)	2.56	2.58
Educational attainment mother (1-5)	3.12	3.15
Educational attainment father (1-5)	3.28	3.33
Family income (1-9)	5.06	5.02
Demographic		
Large municipality	0.82	0.89***

Note: */**/*** indicate significant differences at the 10%/5%/1% level based on the mean differences of the two groups, assessed with a t-test.

Table 3: School Indicators of Special Needs and Regular Students

	Regular Students	Special Needs Students
Before policy		
Academic performance (1-5)	3.88	3.63***
Language performance (1-5)	3.55	3.22***
Math performance (1-5)	3.49	3.26***
Apprenticeship score (1-7)	4.29	4.03***
After policy		
Academic performance (1-5)	3.25	3.17
Language performance (1-5)	3.11	3.05
Math performance (1-5)	2.89	2.83
Apprenticeship score (2-10)	7.25	7.08
Educational attainment (1-5)	3.49	3.41

Note: */**/*** indicate significant differences at the 10%/5%/1% level based on the mean differences of the two groups, assessed with a t-test.

5 Empirical Strategy

We use a Differences-In-Differences (DID) design to control for time-invariant unobserved heterogeneity, enabling direct identification of the school inclusion policy effect on the mental health of special needs students.[9] Four elements are specifc for the DID setting: the first one is the availability of a treated ('eligible') group and control ('non-eligible') group; the second is the existence of common trends ('parallel paths') in the pre-policy period; the third is the clear time cutoff when the policy starts, hence the existence of a pre- and post-policy period; and fourth is the assumption that, with-out the policy, the treated group would show a trend similar to that observed for the control group.

As mentioned above, the school inclusion policy was introduced in August of 2003. In the pre-policy period ($t = 0$), waves 1 and 2 of TRAILS, students i can not make use of the policy, since it has not been implemented yet. In the follow-up period ($t = 1$), wave 3 of TRAILS, students could have benefited from the policy. Hence, the clear time cutoff is between the second and third wave in the population cohort. As discussed, the policy is targeted on special needs students for whom it is difficult to proceed in regular education. Therefore, special needs students are assigned to the treated group ($Z_i = 1$) and regular students to the control group ($Z_i = 0$). Furthermore, we assume absence of any intervention in the baseline for either group ($D_{i,t=0} = 0|Z_i = 1, 0$) and the policy to have a positive effect on the mental health of special needs students in the follow-up ($D_{i,t=1}|Z_i = 1$).

For the outcome variable ($Y_{i,t}$) the population DID effect is then given by the mean difference in mental health for special needs students and regular students before and after the school inclusion policy. The corresponding DID setting is given by

$$\begin{aligned}\text{DID} = &\big\{E(Y_{i,t=1}|D_{i,t=1} = 1, Z_i = 1) - E(Y_{i,t=1}|D_{i,t=1} = 0, Z_i = 0)\big\} \\ &- \big\{E(Y_{i,t=0}|D_{i,t=0} = 1, Z_i = 1) - E(Y_{i,t=0}|D_{i,t=0} = 0, Z_i = 0)\big\}. \end{aligned} \tag{1}$$

[9] We use a similar DID specification to measure the policy effect on academic performance in Section 7.6 of the supplementary analyses.

In the DID extended model we add the individual-level control variables (X_i) age, gender, ethnicity, family composition, educational attainment mother, educational attainment father, family income, and municipality.[10] The corresponding extended DID setting is given by

$$\begin{aligned}\text{DID} = &\big\{E(Y_{i,t=1}|D_{i,t=1} = 1, Z_i = 1, X_i) - E(Y_{i,t=1}|D_{i,t=1} = 0, Z_i = 0, X_i)\big\}\\ &- \big\{E(Y_{i,t=0}|D_{i,t=0} = 1, Z_i = 1, X_i) - E(Y_{i,t=0}|D_{i,t=0} = 0, Z_i = 0, X_i)\big\}.\end{aligned} \tag{2}$$

The regression formulation of the population DID in (2) is given by

$$Y_{ist} = \alpha + \beta \cdot T_{st} \cdot (T = 2003) + \gamma_s + \delta_t + \boldsymbol{X}_{ist} \cdot \boldsymbol{\Gamma} + \varepsilon_{ist}, \tag{3}$$

where Y_{ist} denotes the mental health of student i for policy s in wave t, T_{st} presents the school inclusion policy indicator in 2003, γ_s the policy fixed effect, δ_t the time fixed effect, $\boldsymbol{X}_{ist}$ are the individual-level control variables, and ε_{ist} denotes the error term. The formulation in (3) enables to obtain the standard error and t-statistic of β.

6 Results

Table 4 displays the standardized estimates of β in (3) with and without additional control variables, respectively. The interpretation of the DID estimate is the reduced difference in percentage in mental health between special needs students and regular students before and after the policy.[11] Hence, we observe a reduction of inequality in mental health of 26.2 percent.

[10] Note that adding individual-level control variables only reduces the within-group variance. The within-group variation does not affect the identication of the policy effect but may reduce standard errors. Between-group variance becomes important when observed heterogeneity may confound the identification strategy. Given the features of our DID setting, observed covariates may play a role in the identification of the policy effect. Therefore, in Section 7.2 of the supplementary analyses we add non-experimental methods to the original DID setting to reduce the between-group variance. Specifically, we follow Heckman, Ichimura, and Todd (1997, 1998) and match the special needs students with their regular peers at baseline on the propensity of eligibility for the school inclusion policy conditional on the control variables in (2), X_i. In the second stage we add the kernel weights obtained in the first stage to the DID setting to estimate the exempt policy effect.

[11] Specifically, the absolute difference in mental health between the treated and control group, as measured with the CBCL, reduced from 0.172 before the policy to 0.127 after the policy. This corresponds to a population DID of -0.44 and a standardized DID estimate in regression formulation of 0.262 (0.054).

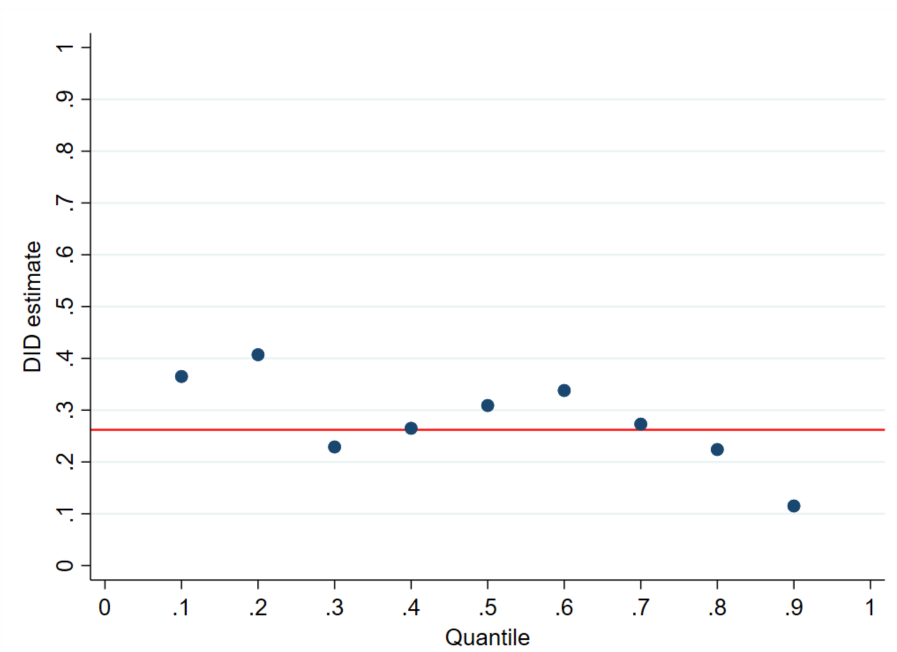
Table 4: DID Estimates for Mental Health

	(1)	(2)
DID	-0.262***	-0.262***
	(0.054)	(0.054)
Covariates	NO	YES
Number of ID	1,112	1,112
R-squared	0.23	0.26

Clustered standard errors in parentheses.
Inference: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

In Figure 2 we show the standardized estimates of β in (3) for ten-percent quantiles with the main result of Table 4 as reference line (red line).[12] There is no clear linear trend across the distribution. The school inclusion policy yields an effectiveness at the median of 30.9 percent. Furthermore, we can conclude that the policy is the most (least) effective at the second (ninth) quantile with a reduction of inequality in mental health of 40.7 (11.5) percent.

Figure 2: Ten-percent Quantile DID Estimates



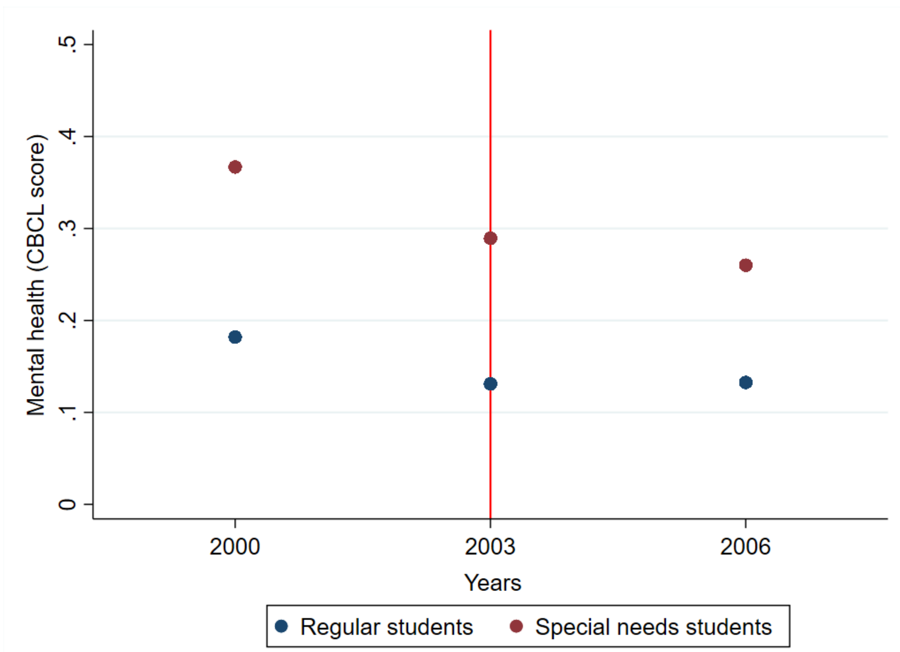
[12]The quantile DID estimates with t-statistic and standard error are presented in Table 14 in the Appendix.

7 Supplementary Analyses

7.1 Parallel Paths

First of all, we test the common trend assumption in the pre-policy period te ensure parallel paths before introduction of the school inclusion policy in 2003. The null hypothesis on pre-common dynamics in the control and treated group cannot be rejected ($p = 0.273$), also illustrated by Figure 3. Hence, this implies that differences in mental health between the regular and special needs students are time-invariant if untreated and that we do not capture placebo effects (Bertrand, Duflo, & Mullainathan, 2004).

Figure 3: Parallel Paths in Mental Health



7.2 Propensity Reweighting

In the third specification we follow Heckman et al. (1997, 1998) to reduce the between-group variance. We use the observed covariates X_i of (2) to estimate the propensity score (the likelihood of indication to the school inclusion policy) and to calculate kernel weights. Instead of only accounting for the within-group variance, this method matches the special needs students and regular students according to their propensity score. First, we obtain the propensity score (p_i) for both groups using probit estimation,

$$p_i = E(Z_i = 1|X_i). \tag{4}$$

Then according to Heckman et al. (1997), the kernel matching is given by the propensity score conditional on the covariates of (2), which leads to the calculation of the kernel weights,

$$w_i = \frac{K\left(\frac{p_i - p_k}{h_n}\right)}{\sum K\left(\frac{p_i - p_k}{h_n}\right)} \tag{5}$$

in which $K(\cdot)$ is the kernel function and h_n is the selected bandwidth, set equal to 0.05. Then we introduce the kernel weights into (1) to obtain a kernel propensity-score matching DID effect. The corresponding DID setting is given by

$$\begin{aligned} \text{DID} = &\left\{E(Y_{i,t=1}|D_{i,t=1} = 1, Z_i = 1) - w_i \times E(Y_{i,t=1}|D_{i,t=1} = 0, Z_i = 0)\right\} \\ &- \left\{E(Y_{i,t=0}|D_{i,t=0} = 1, Z_i = 1) - w_i \times E(Y_{i,t=0}|D_{i,t=0} = 0, Z_i = 0)\right\}. \end{aligned} \tag{6}$$

To increase the internal validity of the DID estimand, we restrict (6) to the common support of the propensity score for special needs students and regular students, following Rosenbaum and Rubin (1985). The common support is the overlap region of the propensity for both groups. This sample of i individuals is then restricted to the region defined as

$$\left(i : p_i \in [max\{min(p_i|Z_i = 1), min(p_i|Z_i = 0)\}, min\{max(p_i|Z_i = 1, min(p_i|Z_i = 0)\}]\right). \tag{7}$$

Then, we show that in absence of the policy, the outcome variable is orthogonal to the policy indicator given the set of covariates X_i. In other words, we test the balancing property at baseline,

$$Y_{i,t=0} \perp Z_i|X_i. \tag{8}$$

The kernel propensity-score matching DID estimate is given in column (3) in Table 5. According to this estimate, the inequality in mental health between special needs students and regular students before and after the policy reduced with 24.1 percent. Although this estimate is slightly lower than the estimate of our main result of 26.2 percent, estimates are not statistically significantly different ($p > 0.10$).

Table 5: Kernel Propensity DID Estimates

	(1)	(2)	(3)
DID	-0.262***	-0.262***	-0.241***
	(0.054)	(0.054)	(0.053)
Covariates	NO	YES	YES
Kernel weights	NO	NO	YES
Number of ID	1,112	1,112	1,112
R-squared	0.23	0.26	0.22

Standard errors in parentheses.
Inference: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

In Table 6 we show that in absence of the policy, the outcome variable is, indeed, orthogonal to the policy indicator, as on all included covariates the difference between the treated and control group is statistically insignificant ($p > 0.10$).

Table 6: Results Balancing Test with the Weighted Covariates at Baseline

Weighted variable	$\Pr(\|T\| > \|t\|)$	Mean Control	Mean Treated	Diff	$\|t\|$
Mental health	0.000***	0.159	0.328	0.170	16.41
Age	0.8126	10.549	10.539	-0.010	0.24
Males	0.8832	0.534	0.539	0.005	0.15
Non-caucasian background	0.6782	0.057	0.051	-0.006	0.42
Number of children	0.9131	2.569	2.567	0.008	0.11
Educational attainment mother	0.9525	3.157	3.153	-0.004	0.06
Educational attainment father	0.8840	3.317	3.329	0.011	0.15
Family income	0.8420	5.049	5.024	-0.025	0.20
Municipality	0.9954	2.142	2.142	0.000	0.01

Inference: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

7.3 Informant Bias

Because there is no golden standard for psychiatric disorders, and reports from different informants tend to correlate only moderately, using information from multiple informants seems the best strategy to chart mental health. Among other things, adherence to this first principle is expressed in the use of child (Youth Self-Report; YSR) and parent (Child Behavior Checklist; CBCL) questionnaires on child and adolescent mental health, which are part of the Achenbach System of Empirically Based Assessment (ASEBA) (Achenbach & Rescorla, 2013), and the use of a teacher-report (Teacher Checklist of Psychopathology; TCP), which was developed for TRAILS on the basis of the Achenbach Teachers Report Form. To understand the extent to what our DID estimate is affected by informant bias, we re-estimate (3) using the child (YSR) and teacher (TCP) questionnaires and compare the estimates with the main result in which we used the parent (CBCL) questionnaire. We follow similar steps of sample selection and identification. Results are provided in Table 7.

Table 7: DID Estimates for Informants Child, Teacher & Caregiver

	Child	Teacher	Caregiver
DID	-0.341***	-0.503***	-0.262***
	(0.043)	(0.087)	(0.054)
Covariates	YES	YES	YES
Number of ID	1,218	454	1,112
R-squared	0.28	0.26	0.26

Clustered standard errors in parentheses.
Inference: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

From a student and teacher perspective inequality in mental health between special needs students and regular students has reduced with 34.1 and 50.3 percent, respectively. Hence, effects of the school inclusion policy are larger through the lens of students and their teachers. For the main result we used the more conservative perspective of the caregivers, which yields an estimate of 26.2 percent.

7.4 Broad-band Syndromes

As discussed in Section 4.2 questions on the CBCL about behaviour are the problem scales (withdrawn/depressed, physical complaints, anxious/depressed, social problems, thinking problems, attention problems, norm-deviant behaviour and aggressive behaviour) can be split in the broad-band syndromes 'internalizing problems' and 'externalizing problems' (Achenbach & Rescorla, 2013), also referred to as 'internalizing socio-emotional skills' and 'externalizing socio-emotional skills' (Attanasio, Blundell, Conti, Mason, et al., 2018).

Table 8: DID Estimates for Broad-band Syndromes

	Internalizing	Externalizing
DID	-0.275***	-0.232***
	(0.067)	(0.068)
Covariates	YES	YES
Number of ID	1,112	1,112
R-squared	0.16	0.19

Clustered standard errors in parentheses.
Inference: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Results in Table 8 suggest a larger policy effect for internalizing socio-emotional skills with a reduced inequality of 27.5 percent to 23.2 percent for externalizing socio-emotional skills. Reducing inequalities in both socio-emotional skills in early years is highly important, as both skills strongly relate with the accumulation of health and human capital across the life course (Attanasio et al., 2018).

7.5 Effect Heterogeneity

In this section we determine the effect heterogeneity, as the main result in Table 4 only provides the mean DID effect for the entire sample. Results are provided in Tables 9, 10 and 11 for stratification by gender, socio-economic classes and ethnicity, respectively. We show estimates for both nested models and decomposed samples (Abadie, Chingos, & West, 2018).

Table 9: DID Estimates for Females and Males

	Full	Females	Males
DID	-0.205***	-0.205***	-0.274***
	(0.074)	(0.074)	(0.071)
DID × Males	-0.069***		
	(0.026)		
Covariates	YES	YES	YES
Number of ID	1,112	585	527
R-squared	0.26	0.24	0.28

Clustered standard errors in parentheses.
Inference: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 10: DID Estimates for Socio-economic Classes

	Full	Low SES	Middle SES	High SES
DID	-0.381***	-0.270***	-0.187***	-0.381***
	(0.133)	(0.086)	(0.066)	(0.133)
DID × Low SES	0.111***			
	(0.029)			
DID × Middle SES	0.194***			
	(0.040)			
Covariates	YES	YES	YES	YES
Number of ID	1,112	168	562	382
R-squared	0.26	0.38	0.28	0.14

Clustered standard errors in parentheses.
Inference: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

Table 11: DID Estimates for Caucasian and Non-Caucasian Students

	Full	Caucasian	Non-Caucasian
DID	-0.267***	-0.267***	-0.094
	(0.052)	(0.052)	(0.223)
DID × Non-Caucasian	0.173*		
	(0.092)		
Covariates	YES	YES	YES
Number of ID	1,112	1,050	62
R-squared	0.23	0.25	0.43

Clustered standard errors in parentheses.
Inference: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

The estimates in Table 9 suggest that males benefited more from the policy with a relative difference with girls of 6.9 percent. In Table 10 we use a validated categorical variable to define the student's Socio-Economic Status (SES): 25% low SES, 50% middle SES, and 25% high SES.[13] The DID estimates demonstrate that students from a high SES background capitalized to a greater extent from the policy than their lower SES peers. Specifically, the relative difference between high SES students and their low SES and middle SES peers is 11.1 and 19.4 percent, respectively. Finally, the estimates in Table 11 suggest that students with at least one parent foreign born benefited less from the policy with a relative difference with their peers from both parents born in the Netherlands of 17.3 percent.

7.6 School Indicators

Table 12 displays the standardized estimates of β in (3) with and without additional control variables, respectively. The interpretation of the DID estimate is the reduced difference in percentage in academic performance between special needs students and regular students before and after the policy.[14] Hence, we observe a reduction of inequality in academic performance of 65.1 percent.

[13] The SES variable is derived with a factor analysis on the control variables educational attainment mother, occupation mother, educational attainment father, occupation father, and family income.

[14] Specifically, the absolute difference in academic peformance between the treated and control group reduced from -0.214 before the policy to -0.060 after the policy. This corresponds to a population DID of 0.154 and a standardized DID estimate in regression formulation of 0.651 (0.319).

Table 12: DID Estimates for Academic Performance

	(1)	(2)
DID	-0.651**	-0.651**
	(0.319)	(0.319)
Covariates	NO	YES
Number of ID	1,112	1,112
R-squared	0.06	0.13

Clustered standard errors in parentheses.
Inference: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

In Table 13 we show the relative improvement in language and math performance separately. Estimates suggest that the achievement gap in language (math) between special needs students and regular students reduced with 80.5 (77.9) percent after the policy introduction.

Table 13: DID Estimates for Language and Math Performance

	Language	Math
DID	-0.805**	-0.779
	(0.354)	(0.574)
Covariates	YES	YES
Number of ID	1,112	1,112
R-squared	0.09	0.08

Clustered standard errors in parentheses.
Inference: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

8 Discussion

In study we evaluated a Dutch school inclusion policy of 2003 with individual-level data of 1,112 students between 2000 and 2013. The DID estimates suggest that supporting special needs students in their progression in regular education with a personal budget, allocated to the school, reduces the inequality in mental health and academic performance with 26.2 and

65.1 percent with their regular peers, respectively. Using a regression formulation in (3) we obtained the corresponding t-statistics and standard errors of the population DID in (2).

To reduce the probablity of placebo DID effects, we controlled for parallel paths before policy introduction. As illustrated in Figure 3, both groups have common trends before the introduction of the school inclusion policy in 2003. Additionally, we observe a structural break for both groups after 2003, the moment students enter higher tracks of secondary education. Furthermore, in the post-policy period the difference in mental health reduced with 26.2 percent, as measured in 2006. Although the DID setting proves consistent identification (Bertrand et al., 2004), the setting is limited by the lack of multiple post-policy periods. However, by using other school indicators in subsequent periods we were able to show that both groups did not differ in their educational attainment in 2013 with 'higher tracks of secondary education' as the attainment at the median for both groups.

Education can for disadvantaged students be an excellent opportunity to rise above their position. However, Chetty et al. (2014b) suggest that students with a high SES can capitalize on the long-term effects of education to a larger extent relative to their low SES peers. Our heterogeneity analyses confirm his cautionary note, as we found larger effects for students from more privileged backgrounds than for their less privileged peers. Specifically, we observe a difference between the lowest and highest SES class of 11.1 percent. Additionally, the difference between students with Dutch-born parents with peers with at least one foreign-born parent is 17.3 percent. Combining this observed effect heterogeneity with the sample selection differences regarding socio-economic environment, we should be careful with the generalizability of results to all societal classes in society.

In this study we extensively controlled for selection bias. Specifically, by using non-experimental methods as additional specification to the original DID setting in the supplementary analyses, we exempted the covariates age, gender, ethnicity, family composition, educational attainment mother, educational attainment father, family income, and municipality from the policy effect. We relate our choice for control variables to previous findings in the literature, as socio-economic environment of living (catched by our control variables educational attainment mother, educational attainment father, family income, and municipality) affects the lifespan of children (Chetty, Hendren, & Katz, 2016; Chetty & Hendren, 2018a, 2018b). Chetty et al. (2016) found that moving to a lower-poverty neighborhood when young increases college-attendance and earnings. This difference in opportunity is endogenous in our kernel propensity-score matching DID setting. Specifically, we show that in absence of the policy, the outcome variable mental health is orthogonal to the policy indicator given the region of living. The extended specification yields a reduction of the DID estimate of 2.1 percent, which indicates that the original DID setting is robust to selection bias.

9 Conclusion

The Dutch Ministry of Education, Culture and Science decided in 2014 to abolish the school inclusion policy, as the budget to support the system exploded and short-term evidence was not convincing. Therefore, they decided to cut expenditures in the subsequent policy with 300 million euros. This study looks back on this decision and shows the potential of high long-term cost/benefit ratios for policies that support progression of special needs students in regular education, taken into account the long-term return to education. We found large positive effects of the school inclusion policy on the mental health and academic performance with a reduced inequality of 26.2 and 65.1 percent, respectively. Moreover, the school inclusion policy combated initial learning impairments before adulthood of special needs students relative to their regular peers, as we observe no difference in educational attainment between both groups at labour market entry. Hence, by reducing inequality in health and human capital, the school inclusion policy equalises opportunities of both groups throughout the life cycle.

References

Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and student achievement in the Chicago public high schools. *Journal of Labor Economics*, *25*(1), 95–135.

Abadie, A., Chingos, M. M., & West, M. R. (2018). Endogenous stratification in randomized experiments. *Review of Economics and Statistics*, *100*(4), 567–580.

Achenbach, T., & Rescorla, L. (2013). Achenbach system of empirically based assessment. In *Encyclopedia of autism spectrum disorders* (pp. 31–39). Springer.

Agüero, J. M., & Beleche, T. (2013). Test-Mex: Estimating the effects of school year length on student performance in Mexico. *Journal of Development Economics*, *103*, 353–361.

Aizer, A., Currie, J., Simon, P., & Vivier, P. (2018). Do low levels of blood lead reduce children's future test scores? *American Economic Journal: Applied Economics*, *10*(1), 307–41.

Amin, V., Flores, C. A., Flores-Lagunes, A., & Parisian, D. J. (2016). The effect of degree attainment on arrests: Evidence from a randomized social experiment. *Economics of Education Review*, *54*, 259–273.

Andersen, S. C., Beuchert, L., Nielsen, H. S., & Thomsen, M. K. (2018). The effect of teacher's aides in the classroom: Evidence from a randomized trial.

Angrist, J. D., & Lavy, V. (1999). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics*, *114*(2), 533–575.

Attanasio, O., Blundell, R., Conti, G., Mason, G., et al. (2018). Inequality in socioemotional skills: a cross-cohort comparison.

Banerjee, A. V., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *The Quarterly Journal of Economics*, *122*(3), 1235–1264.

Battaglia, M., & Lebedinski, L. (2015). Equal access to education: An evaluation of the Roma Teaching Assistant program in Serbia. *World Development*, *76*, 62–81.

Bellei, C. (2009). Does lengthening the school day increase students' academic achievement? Results from a natural experiment in Chile. *Economics of Education Review*, *28*(5), 629–640.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, *119*(1), 249–275.

Bhuller, M., Mogstad, M., & Salvanes, K. G. (2017). Life-cycle earnings, education premiums, and internal rates of return. *Journal of Labor Economics*, *35*(4), 993–1030.

Brunello, G., Fort, M., Schneeweis, N., & Winter-Ebmer, R. (2016). The causal effect of education on health: What is the role of health behaviors? *Health Economics*, *25*(3), 314–336.

Carlsson, M., Dahl, G. B., Öckert, B., & Rooth, D.-O. (2015). The effect of schooling on cognitive skills. *Review of Economics and Statistics*, *97*(3), 533–547.

Chetty, R., Friedman, J. N., Hilger, N., Saez, E., Schanzenbach, D. W., & Yagan, D. (2011). How does your kindergarten classroom affect your earnings? Evidence from Project STAR. *The Quarterly Journal of Economics*, *126*(4), 1593–1660.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014a). Measuring the impacts of teachers I: Evaluating bias in teacher value-added estimates. *American Economic Review*, *104*(9), 2593–2632.

Chetty, R., Friedman, J. N., & Rockoff, J. E. (2014b). Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood. *American Economic Review*, *104*(9), 2633–79.

Chetty, R., & Hendren, N. (2018a). The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects. *The Quarterly Journal of Economics*, *133*(3), 1107–1162.

Chetty, R., & Hendren, N. (2018b). The impacts of neighborhoods on intergenerational mobility II: County-level estimates. *The Quarterly Journal of Economics*, *133*(3), 1163–1228.

Chetty, R., Hendren, N., & Katz, L. F. (2016). The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment. *American Economic Review*, *106*(4), 855–902.

Cullen, J. B., Jacob, B. A., & Levitt, S. (2006). The effect of school choice on participants: Evidence from randomized lotteries. *Econometrica*, *74*(5), 1191–1230.

Currie, J. (2009). Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood, and human capital development. *Journal of Economic Literature*, *47*(1), 87–122.

Currie, J., & Stabile, M. (2006). Child mental health and human capital accumulation: the case of ADHD. *Journal of Health Economics*, *25*(6), 1094–1118.

Cygan-Rehm, K., & Maeder, M. (2013). The effect of education on fertility: Evidence from a compulsory schooling reform. *Labour Economics*, *25*, 35–48.

Deming, D. (2009). Early childhood intervention and life-cycle skill development: Evidence from Head Start. *American Economic Journal: Applied Economics*, *1*(3), 111–34.

Deming, D. J. (2011). Better schools, less crime? *The Quarterly Journal of Economics*, *126*(4), 2063–2115.

Devereux, P. J., & Hart, R. A. (2010). Forced to be rich? Returns to compulsory schooling in Britain. *The Economic Journal*, *120*(549), 1345–1364.

Ding, W., & Lehrer, S. F. (2010). Estimating treatment effects from contaminated multiperiod education experiments: the dynamic impacts of class size reductions. *The Review of Economics and Statistics*, *92*(1), 31–42.

Dynarski, S., Hyman, J., & Schanzenbach, D. W. (2013). Experimental evidence on the effect of childhood investments on postsecondary attainment and degree completion. *Journal of Policy Analysis and Management*, *32*(4), 692–717.

Fitzpatrick, M. D., Grissmer, D., & Hastedt, S. (2011). What a difference a day makes: Estimating daily learning gains during kindergarten and first grade using a natural experiment. *Economics of Education Review*, *30*(2), 269–279.

Grimard, F., & Parent, D. (2007). Education and smoking: Were Vietnam war draft avoiders also more likely to avoid smoking? *Journal of Health Economics*, *26*(5), 896–926.

Hanushek, E. A., Piopiunik, M., & Wiederhold, S. (2014). *The value of smarter teachers: International evidence on teacher cognitive skills and student performance* (Tech. Rep.). National Bureau of Economic Research.

Hanushek, E. A., & Rivkin, S. G. (2006). Teacher quality. In *Handbook of the economics of education* (Vol. 2, pp. 1051–1078). Elsevier.

Heckman, J. J., Ichimura, H., & Todd, P. (1998). Matching as an econometric evaluation estimator. *The Review of Economic Studies*, *65*(2), 261–294.

Heckman, J. J., Ichimura, H., & Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies*, *64*(4), 605–654.

Huisman, M., Oldehinkel, A. J., de Winter, A., Minderaa, R. B., de Bildt, A., Huizink, A. C., ... Ormel, J. (2008). Cohort profile: The dutch "TRacking adolescents' individual lives' survey"; TRAILS. *International Journal of Epidemiology*, *37*(6), 1227–1235.

Jürges, H., Reinhold, S., & Salm, M. (2011). Does schooling affect health behavior? Evidence from the educational expansion in Western Germany. *Economics of Education Review*, *30*(5), 862–872.

Kemptner, D., Jürges, H., & Reinhold, S. (2011). Changes in compulsory schooling and the causal effect of education on health: Evidence from Germany. *Journal of Health Economics*, *30*(2), 340–354.

Konstantopoulos, S., & Chung, V. (2009). What are the long-term effects of small classes on the achievement gap? Evidence from the lasting benefits study. *American Journal of Education*, *116*(1), 125–154.

Krueger, A. B. (1999). Experimental estimates of education production functions. *The Quarterly Journal of Economics*, *114*(2), 497–532.

Krueger, A. B., & Whitmore, D. M. (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from Project STAR. *The Economic Journal*, *111*(468), 1–28.

Leigh, A., & Ryan, C. (2008). Estimating returns to education using different natural experiment techniques. *Economics of Education Review*, *27*(2), 149–160.

Nye, B., Konstantopoulos, S., & Hedges, L. V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, *26*(3), 237–257.

Oldehinkel, A. J., Rosmalen, J. G., Buitelaar, J. K., Hoek, H. W., Ormel, J., Raven, D., ... others (2015). Cohort profile update: the tracking adolescents' individual lives survey (TRAILS). *International Journal of Epidemiology*, *44*(1), 76–76n.

Oreopoulos, P. (2006a). The compelling effects of compulsory schooling: Evidence from Canada. *Canadian Journal of Economics*, *39*(1), 22–52.

Oreopoulos, P. (2006b). Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *American Economic Review*, *96*(1), 152–175.

Oreopoulos, P. (2007). Do dropouts drop out too soon? Wealth, health and happiness from compulsory schooling. *Journal of Public Economics*, *91*(11-12), 2213–2229.

Papay, J. P., & Kraft, M. A. (2015). Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement. *Journal of Public Economics*, *130*, 105–119.

Rivkin, S. G., Hanushek, E. A., & Kain, J. F. (2005). Teachers, schools, and academic achievement. *Econometrica*, *73*(2), 417–458.

Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, *39*(1), 33–38.

Schwerdt, G., West, M. R., & Winters, M. A. (2017). The effects of test-based retention on student outcomes over time: Regression discontinuity evidence from Florida. *Journal of Public Economics*, *152*, 154–169.

Silles, M. A. (2011). The effect of schooling on teenage childbearing: evidence using changes in compulsory education laws. *Journal of Population Economics*, *24*(2), 761–777.

Sims, D. P. (2008). Strategic responses to school accountability measures: It's all in the timing. *Economics of Education Review*, *27*(1), 58–68.

Staiger, D. O., & Rockoff, J. E. (2010). Searching for effective teachers with imperfect information. *Journal of Economic Perspectives*, *24*(3), 97–118.

Taylor, E. (2014). Spending more of the school day in math class: Evidence from a regression discontinuity in middle school. *Journal of Public Economics*, *117*, 162–181.

Urquiola, M. (2006). Using Maimonides' rule to estimate the effect of class size on scholastic achievement. *The Review of Economics and Statistics*, *88*(2), 171–177.

Appendix

A Additional information

Author names and affiliations:

Roel D. Freriks, MSc. (corresponding author)

Faculty of Economics and Business, University of Groningen

P.O. Box 800, 9700 AV Groningen, The Netherlands

Phone: +31(0)50 363 7018, E-mail: `r.d.freriks@rug.nl`.

Dr. Jochen O. Mierau

Faculty of Economics and Business, University of Groningen

P.O. Box 800, 9700 AV Groningen, The Netherlands

Phone: +31(0)50 363 3735, E-mail: `j.o.mierau@rug.nl`.

Declarations of interest: None.

Funding: This research is part of the TRacking Adolescents' Individual Lives Survey (TRAILS). Participating centers of TRAILS include the University Medical Center and University of Groningen, the University of Utrecht, the Radboud Medical Center Nijmegen, and the Parnassia Bavo group, all in the Netherlands. TRAILS has been financially supported by various grants from the Netherlands Organization for Scientific Research NWO (Medical Research Council program grant GB-MW 940-38-011; ZonMW Brainpower grant 100-001-004; ZonMw Risk Behavior and Dependence grants 60-60600-97-118; ZonMw Culture and Health grant 261-98-710; Social Sciences Council medium-sized investment grants GB-MaGW 480-01-006 and GB-MaGW 480-07-001; Social Sciences Council project grants GB-MaGW 452-04-314 and GB-MaGW 452-06-004; NWO large-sized investment grant 175.010.2003.005; NWO Longitudinal Survey and Panel Funding 481-08-013 and 481-11-001; NWO Vici 016.130.002 and 453-16-007/2735; NWO Gravitation 024.001.003), the Dutch Ministry of Justice (WODC), the European Science Foundation (EuroSTRESS project FP-006), the European Research Council (ERC-2017-STG-757364 en ERC-CoG-2015-681466), Biobanking and Biomolecular Resources Research Infrastructure BBMRI-NL (CP 32), the Gratama foundation, the Jan

Dekker foundation, the participating universities, and Accare Centre for Child and Adolescent Psychiatry. We are grateful to everyone who participated in this research or worked on this project to make it possible. The data that support the findings of this study are available, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available.

B Additional Figures and Tables

Table 14: Quantile DID Estimates for Mental Health

Quantile	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
DID	-0.365***	-0.407***	-0.229***	-0.265***	-0.309***	-0.338***	-0.273***	-0.225***	-0.115***
	(0.064)	(0.093)	(0.092)	(0.103)	(0.069)	(0.061)	(0.075)	(0.089)	(0.092)
Covariates	YES	YES	YES	YES	YES	YES	YES	YES	YES
Number of ID	1,112	1,112	1,112	1,112	1,112	1,112	1,112	1,112	1,112
R-squared	0.05	0.06	0.08	0.10	0.12	0.14	0.17	0.20	0.23

Standard errors in parentheses.
Inference: *** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.